KNOWLEDGE GRAPH AND SEMANTIC WEB MODEL FOR CROSS DOMAIN

SHITAL KAKAD¹, SUDHIR DHAGE²
¹Assistant Professor, Sardar Patel Institute of Technology, MMCOE, Pune
²Professor, Sardar Patel Institute of Technology
¹shitalkakad2604@gmail.com, ²sudhir_dhage@spit.ac.in

ABSTRACT

The industry and institutes systems are designed to develop a smart world. The heterogeneous data is available on the web. It is difficult to retrieve precise information from the current web. Semantic web is a solution for semantic interoperability. There are different existing ontology in the education domain. Nowadays, the Industry-Institute gap is increasing due to various reasons. These gaps can be bridged after finding gaps by planning and conducting different activities for students. It will automatically reduce the unemployment rate. Industry and Institute both are equally responsible to develop quality students. In this paper, cross domain (Industry domain and Institute domain) ontology based semantic models are developed to bridge the institute-industry gap using Protégé 5.5.0 editor. The classes and sub classes of Industry-Institute ontology are designed with the help of domain experts. Then, object property and data property are defined to enhance ontology. Next, the result of ontology is validated using HermiT reasoner, DL query and SPARQL query. The graphical representation of ontology is shown by using OntoGraf, OWLViz and VOWL plugin.

Keywords: Ontology, Semantic Web, protégé 5.5.0, Cross Domain, Data Annotation, Data Filtering

1. INTRODUCTION

Institute can provide everything to students as per industry requirement. The existing education domain ontology is mostly implemented for universities, courses, and libraries. In this paper, we have constructed cross domain ontology to bridge the gap between industry and institute. Ontology development takes a lot of effort and time. We surveyed a number of papers on the education domain. There is no ontology which bridges the industry-institute gap.

Why is the Industry-Institute gap increasing day by day?

The main reasons are as follows.
1. Students' interest is changing.
2. Students are not enjoying old teaching-learning methodologies.
3. Less interaction with industry.

There is a necessity to fill these increasing gaps, else it will increase the unemployment rate in the near future. Unemployment Rate is increasing day by day [1] [2]. It does not mean the jobs are not available. A number of jobs are available but students do not have deep knowledge to work on real time projects.

Actions taken to bridge gap:
1. Need to design more practical based curriculum.
2. Conduct more technical and non-technical activity.
3. Teaching and learning must be enjoyable and joyful.
4. Faculty skills development program - by organizing workshops and industry training.

The major contribution of paper are as follows:
1. Develop Institute Ontology
2. Develop Industry Ontology
3. Merged Institute and Industry ontology using Protégé 5.5.0
4. Execute DL queries and SPARQL Queries
5. Validate Ontology (OntoGraf, OWLViz, OntoViz)
We have discussed with the different institute and industry domain experts to construct a semantic model. The ontology is developed to bridge the gap by involving faculty, students, industry people and alumni. Faculty, students, alumni and industry people feedback are taken continuously to build ontology. Generally, Industry people feedback is collected after conducting technical and non-technical activity. Alumni feedback helps to reduce this gap because they know what is lagging in the institute, which area needs to improve more. The feedback of alumni can vary depending on their current job role. The faculty development programs are equally essential to develop faculty skills. The number of activities are conducted for faculty and students by collaborating with industry [3]. We have analysed alumni feedback, faculty feedback and industry person feedback to find industry institute gaps and take actions accordingly.

Industries can provide internships, research funds and sponsored projects to students as well as to faculty. Numbers of jobs are available but students are not capable [4] [5]. So, regular industry-institute interaction is essential to improve students as well as faculty’s technical skills. Faculty should interact with industry people to know the latest trends. Student development completely depends on faculty. So, faculty development programs like workshops, seminars, conferences, industry visits, and short term training programs [5] are necessary to organize especially for faculty. Because of regular interaction with industry faculty will be aware of the current industry requirement and the same knowledge will be delivered to students [17]. Industry people can help to design academic curriculum. Recently, curriculum needs to map and revise with industry requirements regularly. The quality of students can be improved using different teaching learning methods. The industry institute gap depends on Curriculum, faculty experience, student’s assessment process, Industry Institute interaction and many more factors [6].

**protégé 5.5.0**

protégé 5.5.0 is a Java based application to construct ontology [26]. The tabs of software are classes, object properties and data properties, Individual classes, (Decision Logic) DL query, SPARQL Query, Semantic Web rule language (SWRL) tab and onto Graph etc. Class hierarchy is created using the class tab. In this, object property and data property as per domain and scope of the project can be added. The property characteristics are functional, InverseFunctional, transitive and symmetric. You can create your first ontology by studying and implementing pizza ontology https://protege.stanford.edu/ontologies/pizza/pizza. owl.

Generate java code of ontology: Tools->Generate Java code

Merge different ontology: Refactor-> merge ontology

Ontology represents the relationship between data/concepts. Ontology is a backbone of semantic web. The different ontology formats [10] are RDF(Resource Description Framework)/XML(Extensible Markup Language), OWL(Ontology Web Language)/XML, OWL Functional syntax etc. First step, select domain and scope of ontology. We can create different domains ontology such as Healthcare, transport, weather, education. OWLViz is used to visualise ontology. Resource Description Framework Schema - RDF Schema, Web Ontology Language. It is critical to retrieve heterogeneous Information from cross domain applications. Semantic web technologies are used to link web data. Semantic web is nothing but a web of data. It helps to represent relationships of web data. Ontology languages are used such as Resource Description Framework (RDF), Web Ontology Language (OWL), and XML to construct ontology for different domains.

The organization of this research work is represented as follows: Section 2 shows the background and related work on the papers education ontology. Section 3 represents the Industry-Institute Ontology Model and merged Industry and Institute ontology. Section 4 describes evaluation and analysis. Section 5 shows validation ontology (VOWL, OntoGraph, OntoViz). Finally, section 6 shows the conclusion of the paper.

2. BACKGROUND AND RELATED WORK

In the future, it will be very difficult to extract actual information from a huge database. Various ontologies are proposed and developed for the education domain. Institutes have academic pressure of completing syllabus, conducting exams and evaluating assignments. In academics, students and faculty mainly focus on how to increase score and grade instead of practical knowledge. The ontology is constructed for intelligent tutoring [6] using different pedagogical rules and data mining techniques. Fuzzy domain ontology extraction algorithm [7] and context-sensitive text mining are
implemented to analyze students' performance. Students rely on e-learning to learn other than academics. Different recommendation based e-learning using ontology and without using ontology methodology are reviewed as well as analysed [8] [18]. Education ontology for courses [9] is implemented using Protégé 4.0. This ontology helps students to select appropriate courses as per students interest.

Semantic web reduces information retrieval time [10] [11]. Ontology plays a very important role in the world of semantic web. Ontology is constructed for cross domain customer reviews [13] using expectation maximization and semantic similarity algorithm. Higher education (engineering) domain ontology [12] [14] is implemented using protégé 3.4 Editor. Ontology verified by solving SPARQL queries. The ontology is developed for the education domain [15] using different methods of relationship extractor. In this paper Tanwar et al. [16] education ontology is developed using Protégé 5.2.0. This ontology consists of classes and subclasses and verified using DL queries.

The gap [17] between industry and academia tried to find out such as curriculum, industry exposure. Expectations of industries from faculty and students are explained in detail. Cross domain based ontology is constructed using hybrid whale optimization algorithm and rider optimization algorithm [19]. The data is pre-processed, filtered using jaccard similarity and clustered using optimization technique. Performance is measured through precision, recall, ontology construction time and execution time. University ontology [20] is constructed using Protégé. The process of ontology construction using Protégé is explained. FACT++ reasoner is used to check the working of ontology. Ontology is developed to automate higher education activities [21] such as guide allocation and subject allocation. Education system ontology is implemented using clustering [22]. The performance and clustering are improved compared to other models.

The ontology is constructed for examination [23]. It is hard to set paper on each and every topic. This ontology helps to find the depth of examination papers. The ontology is developed to assign subject teachers to respective courses [24]. It checks the expertise and domain knowledge of the teacher before assigning a specific course. Ontology is constructed for course recommendation [25]. This ontology is useful to find a job on the basis of a student's course recommendation. This method is better than the filtering method. The researchers create ontology and publish it online. So, it helps other researchers to do their research work without starting research work from scratch. OWL language is used to construct different domain applications.

3. PROPOSED INDUSTRY-INSTITUTE ONTOLOGY MODEL

The Industry- Institute model is developed to find and bridge the industry-institute gap. Existing education system gap cannot be reduced automatically by just finding a gap. There is a need to take certain actions on identified gaps. Number of problems can occur while implementing an action plan to bridge this identified gap. We have considered a higher education program to build ontology. Next, the main problem, who will find out this gap? Who will take responsibility to work on it? To answer this, we have taken all the stakeholders that are directly or indirectly connected to the institute and industry. Faculty, alumni and students are directly connected to the institute. Industry person are indirectly connected stakeholders. All stakeholders play an equal role to bridge the gap. They know precisely where the problem is in the existing education system. As we know, industry and institute are dependent on each other. The quality of passed out students are concerned to develop a healthy and wealthy Nation.

Next step is collecting gaps from all the stakeholders whatever be their view point. Every stakeholder’s perspective might be different but the final goal is the same and that is to bridge the gap. Student development in all aspects is the prime goal of every institute. On the other side, industry analyses students' quality through several technical and non-technical rounds to hire them. It means Institute outcome is input to industry. Industry hires such students who can work from the first day. Nowadays industries are spending less time training fresher’s. Education is a primary part of student life to enhance their personal skills. The industries are continuously working on recent technology and trends. So, to incorporate all the recent skills within stipulated time in every student is a challenging task. Mostly, the University syllabus changes after 3-4 years. If we observe the syllabus changing duration and technology changing duration is not compatible. This gap is
wide. So there is a need to bridge the Institute-Industry gap by organizing various activities such as conferences, workshops, seminars, visiting industries, research projects, internships, expert lectures and online certification on the latest technology.

Next, we have designed competency questions to check whether implemented ontology works properly or not. We are designing the ontology to answer the following questions. It is one of the ways to validate ontology.

Some part of competency questions are as follows:
1. What are the different co-curricular activities conducted to bridge the gap?
2. Which are the different extracurricular activities conducted to bridge the gap?
3. Which are different activities planned to enhance faculty skills?
4. Which stakeholders are involved to design the curriculum?
5. What actions are taken to bridge the technical gap?

Steps to construct Industry- Institute Ontology are as follows:
1. Construct Institute Ontology
2. Construct Industry Ontology
3. Merged both ontology
4. Execute DL queries and SPARQL Queries
5. Validate Ontology (OntoGraph, VOWL, OntoViz)

3.1 Construct Institute Ontology
Cross domain i.e. Industry and institute domains are considered to build ontology. Finding a gap is a really difficult task. Industry faces the problem when students work on real time projects due to lack of knowledge and personal skills. Students mostly come from diversified regions with different skills. The gap may be applicable to some students but not to all. Some students might be technically sound and some non-technically.

Find and Collect the gap from stakeholders
Figure 1 shows flowchart of finding gaps and action taken. Initially, In this, all the stakeholders attend the meeting and give their valuable feedback. Stakeholders are alumni, faculty, industry experts and students. Meetings and discussions are conducted to find gaps. Gap can be one or many. After finding gaps, actions are taken to bridge this gap.

The gap is divided into two types:
1. Program gap
2. Course gap
Program gap means that particular course is not included in the respective program. This type of Gap Bridge using an add-on course. The duration of this kind of gap would be a minimum of 3 month. Second one is the course gap that represents the gap in the particular course syllabus. This type of gap can be fulfilled by conducting expert lectures, workshops and, hand on session etc.

There are different sources to collect gaps. Students: They are already aware of industry needs. So their involvement to find the gap is necessary. It is one of the biggest sources which can provide a gap for a better education system.
Industry expert: They submit the current needs of the industry. This need cannot be static. Obviously, it may be changed as per industry requirement.
Faculty: They go through the course syllabus and find the course gap and finalize what topic needs to be taught as a content beyond the syllabus. Content beyond the syllabus can be in the form of discussion on the latest journal or conference paper to know the latest research work done in the respective course.
Alumni: They also assist to find gap on the basis of current demand of industry. They can easily figure out what they missed while pursuing a degree. Alumni feedback really adds more value.
We started ontology construction by finalizing classes. In this, the class hierarchy of institute ontology is shown in Figure 2.

The main classes are Academic_Result, Extra_Curricular_Activity, Industry_Institute_Collaboration, Industry_Institute_Conducted, R&D_Center_Activity, Stakeholders, Training_and_Placement_Activity. Institute-Activity_Conducted has subclasses Non-Technical_Activity and Technical_Activity.

The description of classes are given below:

**Academic Result:**
Academic result is just a grade (some number on the scorecard). Assessment tools help to bridge the gap. Faculty can conduct activities such as poster presentation, topic presentation etc. Different teaching learning methodology improves students' leadership, confidence and presentation skills. The academic result improves if students study hard as well as it depends on the teaching method. Sometimes students might be technically strong but can not clear the exam. Academic result grade is the eligibility criteria for students to appear for interviews in each industry. Again, it varies from industry to industry. In this we analyse the students grade (Distinction, first class, second...
class). If a student has an excellent grade it does not mean he/she will be placed in the company. The academic grade is just eligibility criteria. Academic grade reflects the students' hard work, discipline etc. Some students pass without backlog and some with backlog in stipulated time (Program Duration).

**Action taken to bridge academic gap.**

a.Conduct more test: Give more test for practice  
b.Practical based assignment: Give more practical based assignment

**Industry Institute Collaboration**

Industry-Institute collaboration provides industry exposure to students and improves interaction. It provides research facilities to students and faculty both. Government runs different programs to present innovative ideas. These ideas can be useful to society.

**Curriculum Design**- Nowadays, the curriculum is quickly outdated due to rapid technical development. So there is a need to update the syllabus in a short period of time. Theoretical based syllabus does not help students to work in real time application. If the syllabus is designed more practically, It will surely add value to students' knowledge. So, there is a need to involve Industry persons to design academic curriculum as per industry expectations.

**Industrial Visit**- Students visit industry once or twice in a year to understand industry working. Industry visits improve students' development skills. It actually helps to bridge the Industry-Institute gap. Internship opportunities for students can be opened up by regular visits.

**Industry Sponsored Projects**: These projects are sponsored by industry. Faculty advisors and students work on real time projects to get real time knowledge.

**Internship**: Students work on live projects to get practical knowledge.

**Institute Activity Conducted:**

**Technical Activity:**  
Technical activities are conducted such as Add on Course, Conference, Expert Lecture, FDP, Online Certification, Seminar, Workshop.

**Add on courses** – The course which is not in the curriculum.

**Conferences** – Exchange of information in meetings for several days.

**Expert Lecture** – lecture on latest trends.

**Online Certification**- Students learn online certification courses as per their interest.

Seminar - Present and discuss an innovative idea  
Workshop- Mostly focus on practical knowledge of the subject.

**Non Technical Activity:**

**Foreign Language**: Students can complete foreign language courses such as German, Japanese. Those who want to pursue education in foreign country. More job opportunities will be available. It adds top up in resume.

**Soft Skills**- Personality development in terms of all aspects like communication, report writing, presentation and leadership etc.

**R&D Center Activity:**

**Publication**- Stakeholders publish the research work in a book, journal or conference.

**Research Projects** - work on real time research projects.

**Stakeholders**- They are directly and indirectly involved in the process to bridge the gap. Stakeholders are alumni, faculty, industry person and students.

**Training and placement activity:**

**Interview** - Technical and HR Interview.

**Group Discussion**- check communication skills and leadership

**General Aptitude**: Test technical and logical thinking

**3.2 Construct Industry Ontology**

Figure 3 Shows class hierarchy of industry ontology. The classes are finalized by taking help of domain experts. In industry ontology, classes are **Company_Details**, **Industry_Person**, **Industry_ACTIVITY**, **Industry_R&D_Center**, **Industry_Selection_Process**.

The description of the classes as follows:

**Company Details**- shows the location, name of the company.
Industry Person: represents working employee of company ie. Technical or Non-technical.

Industry Activity: Represents activity conducted at industry like conference, Expert_Lecture, Online_Courses, Workshop

Industry R&D Center: Research and development in industry. Industry person publish novel work. It also represents completed and ongoing projects.

Interview Selection Process: The first criteria is student eligibility. Next round is General aptitude. Then, Group discussion and finally interview.

3.3 Merged Industry and Institute Ontology

In this paper, our main purpose is to build merged ontology. Initially, we finalized the domain and scope by reviewing the latest papers. We have taken help from a domain expert to gather ontology requirements. It is necessary to understand problems to build the relationship of data. Ontology is constructed for cross domain i.e. industry domain and institute domain. Both the domains are highly connected to each other.

Nowadays, just scoring a subject and getting a degree is not adequate to survive in real life. Students must be extraordinary to handle all real time situations. He/she must be capable of working in a real time environment. They should have good communication, designing, presentation, writing, problem solving, leadership and many more skills. These expectations from students by industry are increasing day by day. These skills can be merged in students with the help of technical and non-technical activity.

In this we will be merging industry ontology and institute ontology. Both ontology are described earlier. Figure 4 represents the process of Industry and Institute ontology merging. Figure 4(a) shows select the ontology which needs to merge then click on Figure 4 (b) Refactor -> Merge Ontology. Next, select merge type Fig (c) 1. merge into new ontology or 2.Merge into existing ontology. Finally, select the ontology to merge into. We have constructed a merged ontology and named Industry Institute ontology.
Figure 4: Process Of Industry And Institute Ontology Merging (A) Select The Ontology Which Need To Merge (B) Merge Ontology (C) Select Merge Type (D) Select Ontology To Merge Into

Figure 5 Shows class hierarchy of Industry-Institute ontology. It represents final classes after merging two different ontologies. We have not deleted or added a new class after merging. Class Hierarchy- Classes are collections of objects. Class hierarchy views show in two ways either asserted or inferred. Class hierarchy sections are EquivalentTo, SubClassOf, General Class Axioms, Instances, DisjointWith, Target for Key and, Disjoint Union Of. The Manchester OWL Syntax is used by the class expression editor. Rename class by clicking on refactor > rename entity.

Figure 5: Class Hierarchy Of Industry-Institute Ontology
Figure 6 shows object property hierarchy and data property hierarchy. Part of Object property are completed_internship, completed_projects, completed_research_projects_under_R&D, conducted_add_on_courses_by_institute, conducted_expert_lecture_by_institute, conducted_fdp_by_institute, conducted_general_aptitude_test, conducted_interview, conducted_online_certification_courses_by_institute etc. Tpart of data property hierarchy are conducted_online_certification_courses_by_institute, alumni_name, book_name, company_name, conference_name, designation, designed_curriculum, expert_lecture_name. Object property and data property help to build relations. Eg. faculty conducted workshops, conferences and seminars. Industry persons conduct expert lectures and workshops. Completely, relying on classes is not sufficient to design ontology. Object property shows relations within classes based on instances. So we have finalized object property and data property. Some of the individuals are Achal, Mahesh, Aditi, Jitendra etc.

DL queries can be added to ontology to extract more accurate information. Ontology is evaluated by an ontology matrix, and from a result of DL and SPARQL queries. Protégé 5.5.0 editor is used to build Industry Institute Ontology. Quality of the Industry-Institute Ontology is measured using HermiT (version 1.4.3.456) reasoner and executing DL and SPARQL queries in Protégé. Student and faculty information can be retrieved by executing DL and SPARQL queries. Some DL queries are shown below with results. Description of queries:

4. EVALUATION AND ANALYSIS

Execute DL queries and SPARQL Queries

4.1 DL Queries

The DL query [26] tab is available in protégé 5.5.0 software. You can save a DL query by selecting the Add to Ontology button. DL query tab is used to type queries. Run the reasoner before executing DL queries. Type OWL class expression in the DL query tab and click on the execute button to get the result. The result shows direct classes, subclasses, superclasses and instances.
List the Alumni who have conducted add on course or expert lecture or workshop
DL query Figure 7(b)
List the stakeholders who have designed the curriculum of subject cyber security.
DL query Figure 7(c)
Students who have completed projects, internships and also visited industry.
DL query Figure 7(d)
List the stakeholders Who have conducted machine learning workshops or designed cyber security curriculum or completed research projects or visited industry Infosys.
DL query Figure 7(e)
List the instances of conference
DL query Figure 7(f)
List all the subclasses and instances of technical activity conducted in the institute.
4.2 SPARQL Queries:

Figure 8 shows SPARQL Queries.

SPARQL Query Figure 8(a):
- Represents subject and object.

SPARQL Query Figure 8(b):
- List of stakeholders who have conducted workshops.

SPARQL Query Figure 8(c):
- List of stakeholders who have conducted expert lectures.

SPARQL Query Figure 8(d):
- List of stakeholders who have conducted workshops and expert lectures.
Figure 8: SPARQL Queries

Ontology Metrics
Ontology Metrics view shows in Figure 9 with Axioms, Logical axioms, Declaration axioms count, Class count, Object Property count, Data Property count, Individual count, DL Expressivity. Class Axioms are SubClassOf, EquivalentClasses, DisjointClasses, General Concept Inclusion (GCI) Count and Hidden GCI Count. Object Property axioms are SubObjectPropertyOf, EquivalentObjectProperties and so on.

Figure 9: Ontology Metrics

5. VALIDATE ONTOLOGY

5.1 Protégé VOWL
There are several protégé plugins to validate ontology. Protégé VOWL [26] is used to visualize graphical layout representation of ontology. First of all you have to download the JAR file and copy into the plugin folder. Next start the protégé and activate the plugin via Window → Tabs → VOWL. It is helpful to new researchers to understand Institute-Industry Ontology VOWL visualization is shown in Figure 10. It represents graphical overview of Institute-industry ontology.
5.2. OntoViz

Ontology is graphically visualized with the help of Graphviz software in OntoViz [26] Tab. Graphviz needs to be installed before working on the OntoViz tab. Graphical representation using OntoViz is shown in Figure 11.
Figure 11: Industry Institute Ontology OntoViz Visualization
5.3 OntoGraf
OntoGraf represents the structure of ontology. Figure 12 shows Industry institute ontology OntoGraf visualization. By Default OntoGraf views are 1. Class hierarchy, Class views -> Class hierarchy and 2. OntoGraf, Class views -> OntoGraf

![OntoGraf Visualization](image)

Figure 12: Industry Institute Ontology OntoGraf Visualization

6. CONCLUSION
In this paper, Industry-Institute ontology is developed to bridge Industry-Institute gaps. Analyzing and implementing all the institute-industry gaps are a critical task for both institute and industry. The faculty, industry person, alumni and students are involved to find the gap. First of all institute ontology is constructed then industry ontology is constructed using protégé 5.5.0 editor that aims to bridge the Industry-Institute gap. Then, Both ontologies merged using a built in merging plugin. The class hierarchy, object properties and data properties are designed by involving domain experts and reviewing papers. Industry expectations from students are efficiently considered to design this ontology. The Industry-Institute gaps are reduced by taking certain actions like conducting various technical and non-technical activities. Students and faculty personal improvement are recommended to construct ontology. Regular, industry person involvement in institute activities plays a vital role to bridge the gap. The ontology is represented graphically using VOWL, OntoGraph, OntoViz. The ontology is evaluated using DL queries and SPARQL queries. This ontology will be useful for new researchers and in academic development.

REFERENCES


