CHAPTER – 3

Literature Review

3.1. Survey of Existing Students’ Information Ontology Projects
3.2. Survey of Ontology Development Methodologies
3.3. Survey of Ontology Development Tools
3.4. Survey of Ontology Representation Languages
3.5. Conclusion
CHAPTER-3

Literature Review

In computer science, researchers assume that they can define the conceptual entities in ontologies mainly by formal means to specify the intended meaning of domain elements. In contrast, in information systems, researchers discussing ontologies are more concerned with understanding conceptual elements and their relationships, and often specify their ontologies using only informal means, such as UML class diagrams, entity-relationship models, semantic nets, or even natural language. In such contexts, a collection of named conceptual entities with a natural language definition—that is, a controlled vocabulary—would count as an ontology. In this Chapter, I have done a study of the tools and technologies, including existing projects related to the kind of work I will be carrying out. Also a brief study of the various area of work in the context of Semantic web is reviewed.

3.1 Survey of Existing Students’ Information Ontology Projects

Ontology creation in itself is an uphill task. Creation of a large Ontology such as the Students’ Information Ontology requires great efforts and time. The identification of which information to be included in the Ontology requires a study of various sources such as Faculty inputs, study of applicant Bio-data, study of various details of systems revolving around the students’ information and most importantly the disparate ontologies that has been created independently and available over the internet. As per the study done from the various sources, very few ontologies are available, agin which is representing very few details of a students life. Out of those present, majority focuses on storage, sharing and representation of the day-to-day information about a Students’
performance, Test results, e-Learning, Self-assessment, Faculty-Course-Teaching evaluation, Sharing learning resources, etc. None of the ontologies had complete focus on the kind of knowledge I was trying to gather. Hence it was difficult for me to correlate my work with anything existing anywhere near to my work.

Though there is no complete Ontology dedicated for the storing of students’ information, various projects have been undertaken to develop some aspect of the Ontology that is developed as a part of my research work. These projects have been studied with the intent of getting an insight into how the ontology will be created and the naming of the various elements. The following section discusses a few selected projects that are undertaken.

### 3.1.1 K-12 and NIEM

The ability for states to maintain statistics on students’ achievement from Kindergarten to Twelfth standard (K-12) has motivated policy makers to propose legislation to use these statistics to hold states accountable for complying with civil rights objectives. In November 2005 the US Department of Education awarded statewide grants to create longitudinal data standards for statewide student testing. A significant portion of these funds is being used to set up distinct statewide data dictionaries for data warehouse projects. The states of Minnesota, Wisconsin, and Michigan formed a collaborative agreement to develop standardized metadata based on federal standards [131]. The business requirements for this project were to build a shared metadata registry to be used by K-12 assessment data warehouse projects across these three states, as well as researchers at the Wisconsin Center for Education Research.

It contains four central abstract concepts: Activity, Person, Organization and Document. It contains classes like Activity, Address, Assessment,
Contact, Document, EducationalSite, Person, Student, Organization, etc. It also contains various Datatype properties like AddressCityName, AddressLine1Text, AddressStateCode, StudentHomeLanguageCode, StudentAverageDailyAttendancePercent, StudentAchievementLevelCode, OrganizationID, EnrollmentBeginDate, OrganizationMailingAddressText, EnrollmentEndDate, EnrollmentPercent, PersonRaceEthnicityCode, StudentGradeLevelCode, StudentMemberDaysValue, PersonBirthDate, StudentLimitedEnglishProficiencyIndicator, PersonFamilyName, etc.

3.1.2 SWRC Ontology

The SWRC (Semantic Web for Research Communities) [198] is an Ontology for modeling entities of research communities such as Persons, Organisations, Publications (bibliographic metadata) and their relationships [187]. There are classes like Publication, Seminar, Workshop, Conference, Meeting, Book, Project, etc. It contains Object properties and Datatype properties included from the Dublin Core project. It also contains other Object properties like worksAtProject, publication, isAbout, carriedOutBy, financedBy, etc. It has other Datatype properties like lastName, email, homepage, address, firstName, fax, phone, status, startDate, endDate, etc.

3.1.3 OntoSem

The OntoSem Ontology is a formal model of the world that provides a meta-language for describing meaning derived from any source, be it language, intelligent agent perception, intelligent agent reasoning or simulation. The meta-language of description is unambiguous, permitting automatic reasoning about language and the world to be carried out without the interference of lexical and morpho-syntactic ambiguities [133]. It contains classes like student, graduate-student, university-student, document, printed-media, academic-activity, city, hobby-activity, article, work-activity, visual-media-artifact, contact-
information, email, etc. It also contains Object properties like has-industry, has-fax-number, height, created-by, has-street-address, has-address, has-nationality, state-address-of, etc. It consists of only classes and object properties linking these classes.

3.1.4 AKT Reference Ontology

The Advanced Knowledge Technologies (AKT) project [1] aims to develop and extend a range of technologies providing integrated methods and services for the capture, modeling, publishing, reuse and management of knowledge. The AKT Reference Ontology has been developed by the AKT partners to represent the knowledge used in the CS AKTive Portal testbed, and consists of several sub-Ontologies. It is the main Ontology, describing people, projects, publications, geographical data, etc. The Ontology consists of various Object properties and Datatype properties for the classes Person, Organization, Publication, Event, Research Areas, Projects, Locations and Technology Definitions.

3.1.5 RGB Ontology

The University of Maryland, Baltimore County’s Ebiquity Research Group maintains a number of Ontology, including the RGB (Research Group in a Box) Ontology used in publishing their website [56]. There are various OWL Ontologies like, Person Ontology, Event Ontology, Conference Ontology, Contact Ontology, Project Ontology, Publication Ontology and two Ontologies for the Assertion and Association of these Ontologies. All the Ontologies, except the Assertion Ontology and Association Ontology contains Datatype properties. The Assertion Ontology and Association Ontology contain Object properties for connecting the other ontologies.

3.1.6 National Student registry

The central management of students’ profiles at the ministry of education in Palestine is becoming an urgent need in the last few years. Many
students move from one university to another, and they need to transfer their academic records. Also, the ministry of higher education needs to certify the diplomas and mark-sheet of students. Moreover, there is a need to centrally manage and monitor students’ financial aids. Therefore, the Ministry of higher education decided to build a national student registry where, each semester Palestinian universities has to send the academic record of every student to the ministry. The ministry will then update and integrate the academic records according to the data combined from all universities into the national student registry [202]. The ministry wants to specify a shared ontology in OWL such that all universities can exchange their students’ profiles in RDF format and the data can be automatically validated and integrated after that. The main objects in the Academic Record are: University, Faculty, Program, Course, Semester, Student, and Enrollment. The Academic Record consists of Enrollment concept which represents the relation between the student, course, mark and semester. Each enrollment contains courses offered at semester and students study these courses and gets mark. The university contains Faculties and each faculty must offer programs such as master and bachelor, the program has its unique courses offered per semester. The properties used are hasName, hasUniversityName, offeredBY, hasCourseName, hasAcademicYear, hasSemesterNumber, etc. A total of 131 triples are available wherein 7 classes, 7 object properties and 16 datatype properties are used. It includes subClassOf relationship on Student class to foaf:Person.

3.1.7 The UK’s Metadata for Education Group

The UK’s Metadata for Education Group (MEG) [203] serves as an open forum for debating the description and provision of educational resources at all educational levels across the United Kingdom. This group seeks to reach consensus on appropriate means by which to describe discrete learning objects in a manner suitable for implementation in a range of
educational arenas. MEG has established itself as an authority in the
development of descriptive metadata to predominantly UK’s educational
resources. Its main project UK Learning Object Metadata (LOM) Core is an
application profile of the IEEE LOM that has been optimized for use
within the context of UK education. It has created 2 controlled
corpora: UK Educational Levels (UKEL) which provides a set of high-
level terms to name educational levels across all UK educational sectors
like UK Educational Level 1, UK Educational Level 2, and so on up to UK
Educational Level 12. UK Educational Contexts (UKEC) which provides a
set of terms for the environments within which learning and use of
learning objects is intended to take place like nursery education, primary
education, secondary education, and so on including community
education. UKEL can be used within IEEE LOM, IMS and qualified DC
metadata. UKEC can be used within IEEE LOM and IMS metadata. The
MEG registry will provide a mechanism for various learning initiatives to
manage interoperability between their metadata implementations by
publishing their schemas in a registry which is available to others.

3.1.8 Schema.org
The site provides a collection of schemas that webmasters can use to
markup their pages in ways recognized by major search providers [170].
Search engines including Bing, Google, Yahoo! and Yandex rely on this
markup to improve the display of search results, making it easier for
people to find the right web pages. On-page markup enables search
engines to understand the information on the web. It consists of Classes
like Book, Movie, Event, Person, Place, EducationalOrganization, etc. It
also includes Properties like duration, endDate, offers, publisher,
pubishType, isbn, numberOfPages, activityDuration, startDate,
skills, email, faxNumber, telephone, streetAddress, postalCode,
birthDate, knows, memberOf, nationality, etc.
3.1.9 ARIES Laboratory

The University of Saskatchewan’s Advanced Research in Intelligent Educational Systems (ARIES) laboratory [4] has spent the past year using Semantic Web tools and e-learning specifications to develop a loosely coupled and reusable student modeling architecture. This architecture aggregates student data from multiple e-learning applications that have large amounts of use from real students. To support loosely coupled student modelling systems, developers are working with e-learning environments that conform to widely accepted e-learning specifications, such as those developed by the IMS Global Learning Consortium. Student modelling systems that are developed using techniques from the Semantic Web and e-learning specifications have the potential for greater relevance and reuse in real learning systems.

3.1.10 Education Ontology for C-programming

It represents the application ontology designed for the purpose of education and accumulates the authors’ experience of teaching several C-based programming courses [183]. Knowledge should be created in the reusable and sharable form, in a way that once developed it could be used by anyone as a whole or partially. Even greater need in making knowledge shareable and reusable is declared in the field of educational systems development. The KB of a modern computer-based educational system should support the import and export of the knowledge in a standard format using standard protocols. Even for the domains where knowledge is pretty stable, like C Programming, such a perspective lead to the exceptional opportunity of using different systems from different developers in a common framework. It identifies what is important in studying C and accumulates their experience of teaching C-related programming courses.
Chapter 3 Literature Review

Through the searches about ontologies focusing on students’ information and student detail done on popular semantic search engines, I observed very few results were found. The search performed on BING resulted into no search results about students’ information; rather it resulted into results which included e-learning projects, self-assessment projects, etc. Search on HAKIA [89] semantic search engine resulted in ontologies which were related to biological and biomedical applications. The most popular Semantic search engine SWOOGLE, used by many ontologists had a few relevant and repeating results. The keyword student details resulted into 35 results, student information had 162 results and student profile had 19 results, consisting of both RDF and OWL files out of which many were repeating many times. But nonetheless, no result consisted of ontology project focusing completely on the aspects that were covered by my work. My primary source of knowledge collection hence was interview with experienced academia. And my secondary source of knowledge gathering was from the numerous biodata that were received as a part of job applications in my organization.

3.2 Survey of Ontology Development Methodologies

The Ontology Development process refers to the activities that have to be performed when building ontologies. There are several Ontology engineering process proposals reported. In 1990, Lenat and Guha published the general steps [122] about the CYC project. In 1995, based on the experience of enterprise modelling in developing the Enterprise Ontology [206] and the TOVE (TOronto Virtual Enterprise) project ontology [85], the first guidelines were proposed and later refined by Uschold, et al. [205]. At the ECAI’96, Bernaras and colleagues [15] presented a method used to build an ontology in the domain of electrical networks as part of the Esprit KACTUS [171] project. The Methontology methodology [79] appeared during the same period and was extended in
papers by Fernández-López, et al. [66][67]. In 1997, a new method was proposed for building ontologies based on the SENSUS ontology [195]. Some years later, the on-to-knowledge methodology appeared as a result of the project with the same name [184].

All these methods and methodologies were proposed for building ontologies. Additionally, many methods have been proposed for specific tasks of the Ontology Development process, such as ontology reengineering [74], ontology learning [8][113], ontology evaluation [73][75][76][77][86][87][110][210], ontology merging [144][186], ontology evolution [116][142][145][146][185] and ontology alignment [32][57][58][72][134][140][153].

If we analyse the various approaches focusing on the specific aspects of the Ontology Development process, we can conclude [65]:

- None of the approaches covers all the phases involved in ontology building. Most of the methods and methodologies are focused on the development activities, specifically on the conceptualization and implementation, while keeping aside other important aspects related to management, learning, merge, integration, evolution and evaluation of ontologies.

- Most of the approaches are focused on development activities, due to the fact that the ontological engineering field is relatively new. However, a low compliance with the criteria formerly established does not mean a low quality of the methodology or method. As de Hoog [44] states, a not very specified method can be very useful for an experienced group.

- Most of the approaches present some drawbacks in their use. Some of them have not been used by external groups and, in some cases they have been used in a single domain.
It is difficult to use multiple methodologies in a project, since the same phases within two methodologies differ in their technicalities.

Most of the approaches do not have a specific tool that gives them technology support. Besides, none of the available tools covers all the activities necessary in ontology building.

Since my work required development of the ontology, I have done a study of only the ontology development methodologies, as described in Chapter-2. But the process involved was not suiting my requirement. Hence I have developed my own ontology development methodology as described in Chapter-4, which suits my requirement for the development of the Students’ Information Ontology.

### 3.3 Survey of Ontology Development Tools

Researchers have spent their time to build new Ontology Development tools and evaluate their tools as the best one. As a consequence, large numbers of tools are available for Ontology Development. Until today, the number of semantic web tools is 188 [199].

In a survey by Michael Denny [50], Ontology Development tools are compared based on certain features such as modeling features/limitations, base language, web support and use, import/export format, graph view, consistency checks, multi-user support, merging, lexical support, and information extraction. User-experience has received least attention in comparing these tools.

Ontology tools appeared in the mid-1990s for supporting the development process, which may or maynot be following a specific method or a methodology. Based on their knowledge models, ontology tools can be classified in the following two groups:

- Tools whose knowledge model maps directly to an ontology language, hence developed as ontology editors for that specific
language. This groups includes: the Ontolingua Server [59], which supports ontology construction with Ontolingua and KIF; OntoSaurus [195] with Loom; WebOnto [52] with OCML; OilEd [13] with OIL first, later with DAML+OIL, and finally with OWL; and SWOOP [111] and Karlsruhe Ontology (KAON2) [100] with OWL.

- Integrated tool suites with an extensible architecture, with their knowledge model independent of ontology languages. These tools provide a core set of ontology related services and can be extended with other modules to provide more functions. This group includes Protégé, WebODE [5][6], OntoEdit [188], etc.

Ontology development is a complex procedure. Within last few years, researchers have developed a lot of tools for Ontology development, for example, Protégé, SWOOP [111], OilEd [13], Ontolingua [59], Top Braid Composer [201], OntoTrack [123], etc. But Protégé is most widely used by researchers, professionals, programmers, and others for developing ontologies in any domain. Protégé is an open source freely available ontology editor and KB framework that provides a suite of tools to develop domain models and knowledge-based applications with ontologies. Protégé’s knowledge-modeling structure supports the creation, visualization and manipulation of ontologies in various formats such as XML file, RDF file, OWL/RDF file and many more. The developers can create knowledge models by specifying Classes, hierarchy of classes, Object property, Datatype property, Annotate property and user can enter Individuals through GUI forms. Protégé can also be extended by plug-in architecture and a Java-based API for building knowledge-based tools and applications. The Protégé platform supports two main ways of modeling ontologies [162]:

The **Protege-Frames** editor is used to build ontologies which are frame-based with the OKBC protocol. In this model, ontology consists of a set of
classes organized into hierarchical structure, set of Datatype properties, relationships using Object properties, and instances for all classes.

Ontology developer can develop ontologies in OWL using **Protege-OWL editor**. OWL ontology also includes classes along with the hierarchical structure, list of properties and their instances. The major advantage of OWL Ontology is that, OWL formal semantic specifies how to derive its logical consequences. It supports machine readability because machine can derive new facts from existing one.

*Sir Jorge Cardoso* carried a survey on most widely used ontology editors and found that Protégé tool had a market share of 68.2% followed by SWOOP, OntoEdit, Texteditor, Altova SemanticWorks, and so on. Therefore, the Students’ Information ontology is constructed in OWL using Protégé. Figure 7 shows the graphical representation of the findings of Cardoso and Zaino [29][220].

![Figure 7: Ontology Editors used by respondents](image-url)
In an online survey done by *M. Rahamatullah Khondoker and Paul Mueller*, they asked questions to the Ontology developers regarding "Which ontology development tool did you try most?" They permitted only single response for this question [112]. Total 32 participants have participated in online survey as given in Table 2.

**Table 2: Online survey participants grouped according to use of tools**

<table>
<thead>
<tr>
<th>Ontology Development Tools</th>
<th>No of Users (Out of 32 Participants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protégé</td>
<td>24</td>
</tr>
<tr>
<td>SWOOP</td>
<td>2</td>
</tr>
<tr>
<td>Top Braid Composer</td>
<td>2</td>
</tr>
<tr>
<td>Onto Track</td>
<td>1</td>
</tr>
<tr>
<td>Internet Business Logic</td>
<td>2</td>
</tr>
<tr>
<td>IHMC Cmap Ontology Editor</td>
<td>1</td>
</tr>
<tr>
<td>Total Participants</td>
<td>32</td>
</tr>
</tbody>
</table>

The observations [112] on the participants are listed as below:

- The most dominant and domain-independent tool is Protégé, which is used by 75% respondents. One reason of such enormous number of developers could be the mailing-list, large user community and help available online.
- 55.5% of Protégé users feel good to develop ontology with it.
- 58.8% users think that developing ontology using protégé is interesting.
- Protégé could be learnt in one month as responded by 64.7% developers.
- Around 11 Protégé users completed 100% task in average 4 minutes.
Majority of the polled participants (10 out of 14) showed positive attitudes toward satisfaction using protégé: very positive (5), simply positive (5). Some participants (3) showed negative attitudes: negative (2), very negative (1). 1 participant is neutral in this regard.

As per another study by Duineveld, et al. [55], on the tools Ontolingua, Webonto, ProtegeWin, OntoSaurus, ODE, The observation on the development of the first ontology was that ProtegeWin did not require much knowledge of the underlying representation language and was therefore aimed for naïve users. ProtegeWin was easier due to its straight-forward interface. ProtegeWin was found to be very useful during the conceptualization and formalization phase in ontology development.

Since the study showed that Protégé is an easy tool for ontology development, and is freely available with various third-party plugins for additional features, I have decided to use Protégé for the development of the Students’ Information ontology.

### 3.4 Survey of Ontology Representation Languages

Ontology languages appeared in the beginning of the 1990s, normally as an evolution of existing KR languages. Basically, the KR paradigms underlying such ontology languages were based on FOL (e.g. KIF [71]), on frames combined with FOL (e.g. Ontolingua [59][83], OCML [136] and F-Logic [114]), and on DL (e.g. LOOM [124]). In 1997, OKBC [34] was created as a unifying frame-based protocol to access ontologies implemented in different languages (Ontolingua, Loom and CycL, among others). However it was only used in a small number of applications.

The boom of the Internet led to the creation of ontology languages for exploiting the infrastructure of the Web. Such languages are usually called *Web-based Ontology Languages* or *Ontology Markup Languages*. Their syntax is based on existing markup languages such as HTML and XML [24], and whose purpose is not only ontology development but data
presentation and data exchange also. The most important examples of these markup languages are: RDF [121], RDFS [25], OIL [98], DAML+OIL [208], and OWL [46] out of which the ones that are being actively supported are RDF, RDFS and OWL. Also a new ontology language is being developed named Web Service Modeling Language (WSML) in the context of Semantic Web service based on Web Service Modeling Ontology (WSMO) framework.

There are many ORLs developed by which ontology can be represented. I have discussed some of the ORLs in Chapter-2. The Figure 8 shows users of most popular ORLs for the development of Ontology. As per the surveys conducted by Sir Jorge Cardoso and Zaino more than 75% of ontologists have selected these languages to develop their ontologies [127][220].

![Users of various Ontology Representation Languages](image)

**Figure 8: Users of various Ontology Representation Languages**

As per the study of surveys in this Chapter and the study of the ORLs in Chapter-2, I have identified that I will be using OWL as the language for the development of the Student’s Information ontology.
3.5 Conclusion

The study of various Literatures, Research publications and Projects has given an insight about the tools and technologies that will be used in the development of the Students’ information ontology. There are many tools and technologies available out of which, from the study I have identified to use the following for my work:

- My own methodology, as defined in Chapter-4.
- Protégé tool for ontology creation and testing.
- SPARQL query language for querying and testing the ontology.
- OWL as the language of the ontology.