GENERATING STRUCTURED DATA FROM GOVERNMENT OPEN DATA USING RESOURCE DESCRIPTION FRAMEWORK FOR KNOWLEDGE

by Wilson Ambale Amutsa

Submission date: 19-Dec-2017 11:21AM (UTC+0300)
Submission ID: 897966771
File name: Research_for_Wilson_Ambale.doc (1.58M)
Word count: 18398
Character count: 103790

CHAPTER ONE: INTRODUCTION

1.1 Background

We are on the verge of an era in which everyone is moving data online and government entities aren't an exception to this. The government public administration and operation area is a large, heterogeneous and distributed environment where information, services and processes have previously been stored and produced in each specific department with no central control, no common models, holds minimal or no knowledge representation lacks the concept of domain knowledge sharing and thus has no service definitions. This lack has made it hard for exchange and cooperation of information between the different departments. Much of this push came in the wake of open data movement that sprung around the world with the earliest initiative launched in the US by President Obama in 2009 (Mutuku & Mahihu, 2014) key in advocating for linked data to prove the value of structured data on the web in standards such as RDF, OWL and SKOS. The experience from the UK and US government was that the structured data community was not quite ready for a major government to start creating a web of linked and structured government data (Nigel Shadbolt et al., 2012). According to (Anthopoulos, 2014) and with an example given of Digital Morocco, 2013, building an efficient e-Government for purposes for presenting information and electronic services (e-services) through web as portals to citizen and enterprises is the key to filling that gap within governments around the globe and Kenya is not an exception to this.

Semantic Web technologies have surfaced showing a potential solution to these issues(Álvarez Sabucedo, Anido Rifón, Corradini, Polzonetti, & Re, 2010; Apostolou, Stojanovic, Lobo, Mir?, & Papadakis, 2005; Xiao, Xiao, & Zhao, 2007) These tools as much as enhancing data and services description with additional semantic information do facilitate the building of common models which describe the information available and disintegrate it into domain knowledge with a purpose of achieving a given user task. Certainly, the information therein will be for both human consumption and interpretable to the machines. Developing the models requires deep knowledge of the domain; vision of the domain and experts. Moreover, suitable tools, standards, methodologies of semantic techniques are necessary ((Lamharhar, Chiadmi, & Benhlima, 2015). It's within this paper that such a model has been suggested to structure into semantically acceptable format the data available in the Kenyan open data portal. The paper presents a combination of an ontology building methodology and two state of-the-art Semantic Web platforms, namely Protégé and Java Jena ontology API, for semantic ontology development in e-government. Firstly, data from the Kenyan government open data portal is chosen to build a domain ontology describing the semantic content of a government service domain. Thereafter, Protégé and Jena API are employed to create the Web Ontology Language (OWL) and Resource Description Framework (RDF) representations of the domain ontology respectively to facilitate its computer processing. The aims of this study are;

- Show the importance of centralizing government data from several domains into one portal as e-government projects which can easily be accessible to the citizens for retrieval of knowledge and
- Strengthen the embracing of semantic technologies in e-government. The study would also be of interest to beginner Semantic Web developers who might use it as a beginning point for advanced investigations.

1.2 Problem Statement

The Kenya Open Data Initiative platform was launched in July 2011 with the intention of making Kenyan government data openly available through a single online portal, opendata.go.ke (Kwamboka, 2013). A new national Constitution, adopted in 2010, which mandated a new era of public participation in government and altered the way in which Kenya's counties communicated with central government is the reason behind platform's launch (Omolo & Rose, 2013). Right to information is enshrined in the Constitution's Bill of Rights under article 35.It states that an individual has a right of access to information held by the State and information held by another person required for the exercise or protection of any right or fundamental freedom. Therefore the state has an imposed duty of publishing and publicising any important information affecting the nation (Republic of Kenya, 2010). Yet, whilst there exists open data portal for information access in the Kenyan context - there is at the same time relatively little attention to formats and best practices. Data is published, granted, but formats vary from simple HTML tables to data in proprietary formats such as PDF, CSV and Excel, making it hard to combine, compare and reuse information on a large scale. Additionally, even though different government datasets have relation between each other and other related data under their different domains, the datasets are not linked together. Vocabularies and data formats are unfamiliar and inconsistent, especially when crossing from one government department to the other. Finding, assembling, and normalizing these data sets is time consuming and prone to errors and, currently, no tools are implemented in government dataset to make intelligent queries or reasonable inferences across it. This makes accessed data not useful for reuse to most people i.e. programmers, researcher etc.

1.3. Research Questions

Organizing data has had a primary objective of making the World Wide Web not only useful for sharing and interlinking information, but also for sharing and retrieval of the data at very detailed levels to get a customized answer as pertaining to specific search. The reasoning is that these technologies could revolutionize global data sharing, integration and analysis, just

like the classic Web revolutionized information sharing and communication over the last two decades. Deployment of structured government data comes with a specific set of requirements and the degree in which the requirements are accomplished (i.e. data representation, query and visualization) will decide how useful the data itself is to the consumers. This whole aspect leads to the following research questions that are dealt with in the course of this thesis.

- Which techniques have been used in data structuring of decentralized government web data for knowledge sharing and retrieval?
- What are the main factors that hinder sharing and retrieval of structured data?
- How can ontologies be applied in the structuring of data?
- How effective is the ontology-driven approach for structuring data?

89

1.4 Objectives

1.4.1 Main Objective

- The research seeks to demonstrate the usefulness of structuring decentralized government data on the web so that it's readable and useful to both human and machine by proposing an architecture and prototype implementation.

1.4.2 Specific Objectives

- To describe the various techniques used in structuring decentralised data that will make it knowledgeable.
- To design an ontology-driven approach for sharing and retrieval of knowledge.
- To evaluate the ontology-driven approach for sharing and retrieval of knowledge.

1.5 Research potential impact

This research work will have the impact: "*Generating structured data from government open data using Resource Description Framework for Knowledge sharing and retrieval*"

This research presents not only speed up the process of identifying and classifying open government data from online media and government archives streams for reuse, and redistribution but develop a heterogeneous web based e-government systems of government departments and agencies that can interoperate and be easily integrated. This project will

have a substantial impact on the social, economic, scientific and technical domains with access to government information as knowledge with easy retrieval and for reuse.

16

Public sector information is a strategic resource, holding great potential for a number of beneficiaries including public sector agencies, private businesses, the academia, citizens and civic organisations. In many countries, the open government data community still appears to be uncoordinated. It includes IT professionals, both small and large companies, and entrepreneurs as well as developers, government employees, civil society organisations and individual citizens active on the national or international level. Their motivations, level of understanding of government processes and structures, scope and priorities in advocating for OGD differ(Bizer, Heath, & Berners-Lee, 2009).

Entrepreneurs stress clear conditions for re-use and reliable licensing. Programmers demand raw data. Transparency activists want access to internal government documents. Individual citizens might not be interested in data per se, but in secondary information-type products (e.g. new services and mobile apps). Civil society organisations are keen to have a number of datasets that, if combined, may help improve life and service delivery to certain segments of the population or to certain neighbourhoods(Alemu, Stevens, Ross, & Chandler, 2012; Bizer et al., 2009)

CHAPTER TWO: LITERATURE REVIEW

This section presents a review of past studies. This includes what has been done in the area of semantic techniques and a brief description of what structured data is. It also provides detailed background information on structured data, its benefits, together with its barriers.

2.0 Introduction

A number of open data movements in the past years have spiralled up around the world, with transparency and data reuse as two of the major aspires (Alexopoulos, Spiliotopoulou, & Charalabidis, 2013). Some of these movements include ; the Public Sector Information (PSI) Directive in 2003 in Europe,U.S. President Obama's open data initiative in 2009,the Open Government Partnership in 2011, and the G8 Open Data Charter 4 in 2013(Alexopoulos, Zuiderwijk, Charapabidis, Loukis, & Janssen, 2014; Gil-Garcia, Helbig, & Ojo, 2012). As a result of these movements a number of Open government data portals resulted such as data.gov.uk, data.gov.us, data.gov.sg, and opendata.go.ke. This enabled a way in which citizens and stakeholders were able to obtain government information about their locality or country in subject. Though, some governments which are not comfortable with holistic approach of transparency on the government happenings have not embraced the idea of open data.

A lot of issues were a motivation to the founding of the government portals; corruption was one of the main issues where citizens felt that all government happenings need to be in light. Affecting people's lives and regularly infringing fundamental human rights corruption has been a global issue that seriously harms the economy and society as a whole. The deep-rooted corruption affects the democracy of many countries around the world as well their economic development.

2.1 Open government data initiative in Kenya

In the past most government data was siloed and archived in files which are physical storages. Policies, ingrained practices of the colonial and corrupt networks in public institutions benefited greatly from this culture of monopolizing access to information, and used this power to advance their personal interest, usually at the expense of the citizens which made access to this information extremely difficult and in some cases impossible (Schwegmann 2013, Kwamboka 2013, Majeed 2012). At the turn of the century and with little democracy in the country civil societies advocated for data held in the ministries availed

to the public. In July 2011, with the support of World Bank, Kenya launched an open data portal [1], and this made her the first developing country to have a national government open data platform (T. Davies, 2012).Key government data was to be freely available to the public through a single online portal.

The key driver of open data initiative in Kenya is the political will. Government institutions/departments are charged with the exclusive duty of collecting and storing data that relates to their mandates. The 2009 census, national and regional expenditure, and information on key public services were some of the first datasets to be released. A user friendly interface on the front end was created for users of the data. It allowed them to interact for visualizations and downloads of the data and easy access for software developers. Indeed, tools and applications have already been built to take this data and make it more constructive than it firstly was (Oyatsi, 2015).

The data ecosystem is as shown in the picture below

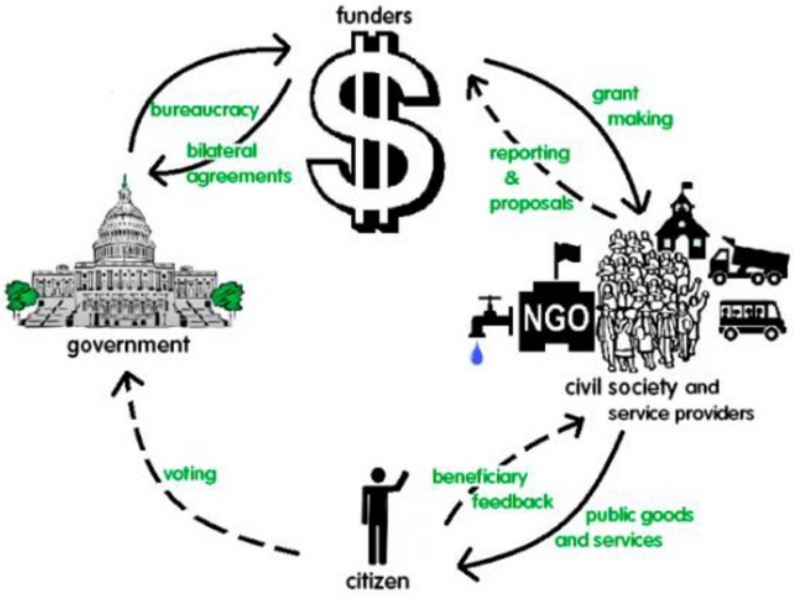

Figure 2.1: Kenyan Government Data Ecosystem – Comphrehensive Analysis Results, source (Oyatsi, 2015)

[1] Opendata.go.ke

Currently, there exists an open data portal for the Kenyan government hosted by the ICT authority board of Kenya. This is central data collected from government institutions /agencies.

2.2 Definition of Structured Data and knowledge

Structured data is defined as structured- information with a degree of organization that is readily searchable and quickly consolidate into facts. Examples: RDMBS, spreadsheet. The concept of structured data semantically labels to make it appropriately organized/stored for subsequent analysis. However, in the past such labels are not present when the data units are encoded in the returned result page after a web search. In that case human users can easily tell and understand the search results as opposed to machines. To make it machine readable huge human efforts was required to annotate the data units manually, which severely limits the scalability of such applications and re-usability of the data.(Lu, He, Zhao, Meng, & Yu, 2007).

Knowledge has heavily been mistaken to mean information or data. Knowledge is understood to come from information, and the information is from data. Data are sets of facts that contribute to the formation of information (Yao et al. 2009). Information, on the other hand, can be described as virtual or aural communication between people (Andreasian, 2013). For the complete cycle of communication both sender and a receiver who act as the sources of data are required. Information gives meaning to the data passed across depending on the perspectives of the receiver.

Knowledge is information that resides in an individual's mind (Yao et al. 2009). Knowledge can be classified as either explicit or tacit. Unlike explicit knowledge which is articulated, written down or published academic knowledge found in books, manuals, papers, e.t.c (Jer Yuen & Shaheen Majid, 2007), tacit knowledge is more dependent on its holder. It is attached to a person's mind and deeply grounded in an individual's action and experience (Panahi, Watson, & Partridge, 2013)

There are several ways in which knowledge can be created: socialization, externalization, internalization and combination (Andreasian, 2013). Socialization is the process of passing tacit knowledge from one person to the other. It can either occur in physical environments or virtual environments. Combination, on the other hand, is the conversion of explicit

knowledge into another explicit knowledge. The conversion of tacit knowledge into explicit knowledge is referred to as externalization while conversion of explicit knowledge into tacit form is internalization.

2.3 Knowledge Sharing

Knowledge sharing is about communicating knowledge within a group of people. These groups may be people in business organizations or learning institutions such as universities or citizens probing into government data to seek answers to how tax is spent. Similarly, knowledge sharing can be seen as the willingness act whereby knowledge is capable of being used again or repeatedly in the course of its transfer from one party to another (Fonou Dombeu & Huisman, 2011).The underlying purpose of all these is to utilize the available knowledge to improve the person's or group's performance. As a result, many governments have employed virtual government portals rich with government data in support of knowledge sharing (Dombeu & Huisman, n.d.; Fonou Dombeu & Huisman, 2011).

Knowledge sharing requires collaboration between the consumers and contributors of knowledge; namely the collaborators (Yang & Chen, 2008). Collaborators can be any virtual user who interacts to achieve the goals of resources discovery, access, and expert knowledge sharing (Yang & Chen, 2008).The web as one tool that encourages knowledge sharing is an excellent environment for these collaborations as government agencies or departments can interact and share knowledge despite their geographical location. Web technologies have enabled people to create, contribute and share expert knowledge with others around the world. These technologies cover a wide set of publishing tools and social networking (Dombeu & Huisman, n.d.) that enable government and its citizens share a variety of information. If e-Government systems are to achieve the anticipated goals of storing, sharing manipulating, and preserving knowledge, then these systems must incorporate mechanisms for domain-specific information.

2.4 Knowledge Retrieval

Government institutions are considered knowledge intensive organizations that contain knowledge ranging from personal skills (tacit knowledge) to explicit or documented knowledge (in traditional archive form for the past several years). To make use of the knowledge, a base of knowledge is highly recommended. However, for efficient knowledge exchange, there has to be a mechanism by which the knowledge can be accessed and retrieved. Effective knowledge identification and transfer mechanisms are very paramount in any knowledge sharing environment. One often unseen benefit for this activity is the knowledge embedded within and between records and legacy record keeping systems;

however, lacking government-wide documentation management systems and issue precise resource description, timely retrieval of relevant records is impossible.

The process of knowledge exchange, therefore, requires two important steps. First, the shared knowledge must be stored in a repository or government silos. Secondly, the interested person should be in a position and willing to retrieve the knowledge. In the past and upto date most used data retrieval systems such as database management systems (DBMS) are suitable for storage and retrieval of structured information (data) while information retrieval systems like web search engines are suitable for finding relevant data. These systems have only one problem which is management of knowledge retrieval. From a group of the retrieved documents a user is forced to analyze them all and identify those relevant to his/her search. Knowledge retrieval systems, an improvement of the two, are used for supporting knowledge discovery, storage, organization, and retrieval.

Yao et al. (2009), studies reveals retrieval systems as a hierarchy of three different levels mentioned; the data level, information level, and knowledge level. The data level comprises of data retrieval systems that base their retrieval mechanisms on the structure of the queries send to the database. The queries should be well structured and the concepts well defined to get required information from the stored data. Data mining can be utilized to get meaningful data from large databases. The next level is the information level that comprises information retrieval systems that utilize keyword search. The queries, in this case, are semi-structured and the concepts not clearly defined. Such systems are only effective where little information is collected. If the results retrieved are huge, identifying the relevant content might be cumbersome. The information retrieval level shows that there is much more to be done in addition to simple search. Extracting specific knowledge from volumes of information using information retrieval mechanisms is not easy. This difficulty is as a result of the difference between human thinking and the knowledge representation mechanisms of information systems.

According to Yao et al. (2009), the focus of knowledge retrieval systems is to link between human thinking to that of machines by organizing and structuring the information stored in machines. Instead of retrieving information, then manually mine knowledge, a person ought to directly retrieve knowledge thus extraction of knowledge is moved to machine level and not human thinking level. Hence this will solve problems experienced in information retrieval

through knowledge retrieval. This extraction of knowledge is based at the knowledge level Knowledge retrieval is a two-way process which starts with the identification of the required information from the repository of knowledge and then the retrieval.

2.5 Knowledge Sharing Techniques

Knowledge is an important asset to everyone around the globe ranging from individuals to governments. The knowledge ranges from personal skills (tacit knowledge) to explicit knowledge that is documented. However, the effectiveness of the knowledge lies in the mechanisms through which it can be effectively shared between individuals. Consequently, organizations and institutions have employed various mechanisms through which they exchange knowledge. The mechanisms that have proved helpful are web 2.0 and web 3.0. However, web 3.0 (Semantic Web) is considered better than web 2.0 tools because it enhances the web by providing mechanisms by which computers can process, interpret and connect information to enhance access and retrieval. In particular, it makes it possible to develop intelligent applications that can facilitate knowledge discovery, sharing and exchange.

According to (El Harbi, Anderson, & Amamou, 2011) is a study carried out to examine the mechanisms and procedures used in sharing knowledge and information in small ICT firms in Tunisia. The study was based on four small ICT companies which included Offshore Box, Ciel Informatique, THY Data Center and Meca-Precis. The companies mainly exchanged knowledge through formal meetings, frequent trainings, electronic forums, accessing electronic resources such as books and discussions on knowledge sites. THY Data also had access to Microsoft Network's electronic books and websites while Meca-Precis collaborated with partners from other countries such as Germany. Ciel Informatics also shared knowledge with university students through supervising their projects. However, the results also indicated that there are no clear and efficient mechanisms for knowledge sharing between universities and the companies.

This and many other studies available are important as they can be used as survey studies for researchers who want to venture into identifying and proposing the most effective platforms for facilitating knowledge exchange in government portals and how to display the information. This is because from the survey, it is clear that knowledge sharing platforms should be interactive, easily accessed and should be ones that easily connect people despite

their geographical location. This section presents various knowledge sharing techniques that have been proposed, especially those that make use of web 2.0 and web 3.0 techniques. Web 2.0 techniques have been found to be useful in supporting collaboration in an interactive environment while web 3.0 techniques (Semantic Web) have been found useful in structuring of content for easy machine interpretation.

2.6 Web Evolution

The web has become an increasingly important resource in many aspects of life: education, employment, government, commerce, health care, recreation, and more. The web is an interlinked system, hypertext documents accessible to the user only via the Internet. User view web pages with the web browsers and these pages may contain text, images, videos, and others multimedia. Users navigate between them by use of hyperlinks. The web was created in 1989 by Sir Tim Berners-Lee, at that time working at CERN (The European Organization for Nuclear Research) in Geneva, Switzerland. Since then there has been a lot of evolving in the web which now is at web 3.0 referred to as the semantic web. Berners-Lee has played an active role in guiding the development of web standards (such as the mark up languages in which web pages are composed), He is the brain behind the advocacy vision of a Semantic web (Aghaei, 2012).

2.6.1 Web 1.0

It was created in 1989 by Tim Burners-Lee and the dream behind it was to create a common information space which could aid people communicate by sharing information (Aghaei, 2012). Web 1.0 was mainly a read-only web as it was static and to some extent mono - directional. It acted like a platform presenting information from which people could read and extract knowledge e.g business presenting production brochures then people could read and contact them for business. Websites included static HTML pages that updated rarely whose main goal was to publish the information for anyone at any time and establish an online presence. Therefore, they didn't envisage interactiveness and indeed were as display-ware.

2.6.2 Web 2.0 Techniques in Knowledge Sharing and Retrieval

Currently, we are in the flourishing stage of Web 2.0, or what can be referred to as the read-write web as Berners-Lee's describes it. Web 2.0 has changed the landscape of the web with its ability to contribute content and interact with other web users has dramatically changed the landscape of the web in a short time. Technologies such as weblogs (blogs), social

bookmarking, wikis, podcasts, RSS feeds (and other forms of many-to-many publishing), social software, web APIs, and online web services such as eBay and Gmail provide enhancements over read-only websites. Web 2.0 has been portrayed as a thought in people's heads rather than reality. The thought people reciprocity between the user and the provider is what's emphasized which in other words means, genuine interactivity, since there's easy upload as well as download for the concerned parties (Rubio, Mart?n, & Mor?n, 2010).

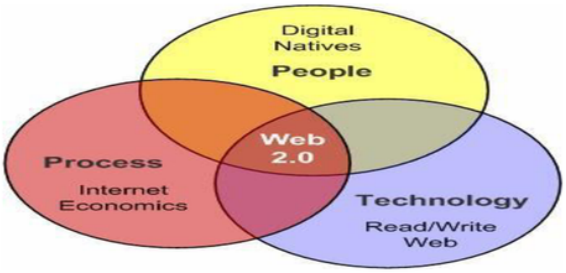

Figure 2.2: Web 2.0, (Rubio et al., 2010)

For there to be effective knowledge exchange collaboration between individuals is essential and for the collaboration to be successful there has to be willingness among or between these parties in sharing the knowledge. However, in virtual environments, effective collaboration depends on mechanisms put in place to facilitate interaction apart from just the willingness for collaborating parties. For instance, knowledge intensive organizations or institutions possess both explicit and tacit knowledge. Unlike explicit knowledge that can be easily exchanged, tacit knowledge is more of personal skills and therefore not easily exchanged. Though not clear on which mechanisms can facilitate interaction in virtual environments, some researchers believe web 2.0 techniques enhance interaction in collaborative environments and especially exchange of tacit knowledge (Andreasian, 2013). Since exchange of tacit knowledge involves direct interaction (socializing) between individuals, the only way to share it in virtual environments is through the interactive tools provided by web 2.0 (Andreasian, 2013).

However, there are still debates on whether web 2.0 tools facilitate exchange of tacit knowledge (Panahi et al., 2013). Some researchers especially those that conducted their study before introduction of Web 2.0 tools believe that tacit knowledge cannot be shared (Panahi et al., 2013). They believe that tacit knowledge is highly personal and inexpressible thus

impossible to share. That, it can only be shared through direct communication such as mentoring, observing, participation and storytelling among others. However, there are those that believe tacit knowledge is transferable through IT tools though not as efficient as in person interaction (Panahi et al., 2013). In references to their argument, IT supports tacit knowledge sharing through providing a platform where people share their experiences and express their personal ideas and arguments through dialog. Nevertheless, no empirical evidence supporting sharing of tacit knowledge through IT tools exists.

Regardless of the contest, web 2.0 techniques are known to be used in interactive virtual environments to enhance collaboration especially in learning institutions such as universities since there's frequent exchange of knowledge between individuals such as among students and between lecturers and students in the form of research hence there has to be mechanisms that can make the process simpler. People involved in the research could be situated in different geographical locations where interaction between them can only be made possible through the internet.

2.6.3 Web 3.0 (Semantic web) Techniques in Knowledge Sharing and Retrieval

The semantic web manifestation was to encourage web page designers and creators to provide some form of metadata (data about data) to the web pages so that content therein could be accessible and understood by machines as well as human beings (Jones, 2004). With this concept instead of hiding the useful content of web pages within text and pictures that was only easily readable by humans, people would create machine-readable metadata to allow machines to access their information. Inference engines would be used to make assertions and inferences from Meta data. RDF drafted by Hayes as the formal framework for this metadata is a simple language for creating assertions about propositions. RDF is of the idea of the "triple" whereby any statement can be composed into subject, predicate, and object (Jones, 2004). There exists another language as Web Ontology Language (OWL) which is more communicative than RDF. The Semantic Web archetype made one small but elementary change to the architecture of the Web: a resource (that is, anything that can be identified by a URI (Uniform Resource Identifier) can be about anything. This means that URIs, that were formerly used to denote mostly web-pages or anything from things whose physical existence is outside the Web to abstract concepts (Studer, Grimm, & Abecker, 2007).

In 2006, Tim Berners-Lee, James Hendler and Ora Lassila uncovered semantic web as a new vision of web pages which envisions a highly interconnected network of data that could be easily be accessed and understood by any desktop or handheld machine (**figure 2.3**). In their explanation to they painted a future of intelligent software agents that would head out on the WWW and without human intervention book flights, bring up to date our medical records and furnish us a single, made to order answer to a particular question without searching for information or navigating through results.

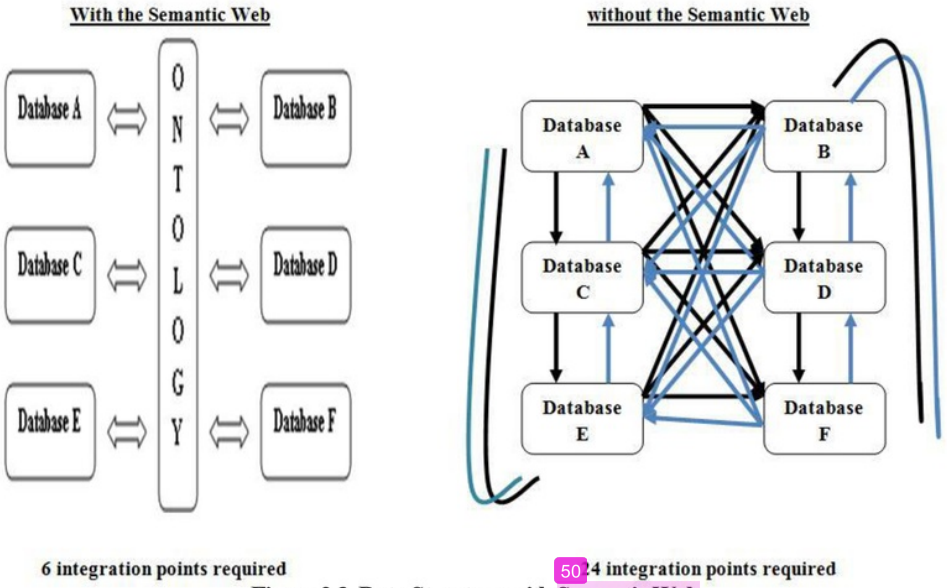

Figure 2.3. Data Structure with Semantic Web

The core belief of Semantic Web is built around the below:

- Data interchange formats (e.g. RDF/XML, N3, Turtle, N-Triples).
- Tools to query information described through relationships (e.g., SPARQL).
- Tools to have a better-quality and more comprehensive categorization and characterization of those relationships as well as the resources being characterized (e.g. RDF Schemas, OWL, SKOS).
- For complex cases, tools are existing to describe logical relationships among resources and their relationships (e.g. OWL, Rules).

- Tools to extort from, and to bind to traditional data sources to make sure their interchange with data from other sources. (e.g., GRDDL, RDFa).
- Semantic publishing on the Web or semantic web publishing refers to publishing information on the web as documents accompanied by semantic mark up. This kind of publication provides a way for computers to understand the structure and thus make meaning out of the published information, making information search and data integration more efficient (J. Davies, Fensel, & Van Harmelen, 2003; N. Shadbolt, Berners-Lee, & Hall, 2006).

2.7 Semantic web data technologies and tools

This section discusses the various web technologies and tools that support the semantic web:

2.7.1 Overview of Semantic Techniques

Semantic Web is an extension of the current World Wide Web and not a complete replacement whose aim is to add structure or meaning to what is on the Web thus allowing computers to process information, interpret, and connect it to enhance knowledge retrieval (Adebayo, Adio, & Adetokunbo, 2011). According to (Giri, 2011), the Semantic Web heavily relies on formal ontologies that structure underlying data for the purpose of comprehensive and transportable machine understanding. It is an XML-based ontological application that provides intelligent access to heterogeneous and distributed information (Adebayo et al., 2011). Therefore, the success of the Semantic Web depends strongly on the proliferation of the ontology.

Semantic Web has a number of components namely; eXtensible Mark-up language (XML), XML scheme, Resource Description Framework (RDF), RDF schema, Web Ontology language (OWL) (Shamsi & Khan, 2012).XML makes available a surface syntax for structured documents, XML schema is a language for restricting the structure and content

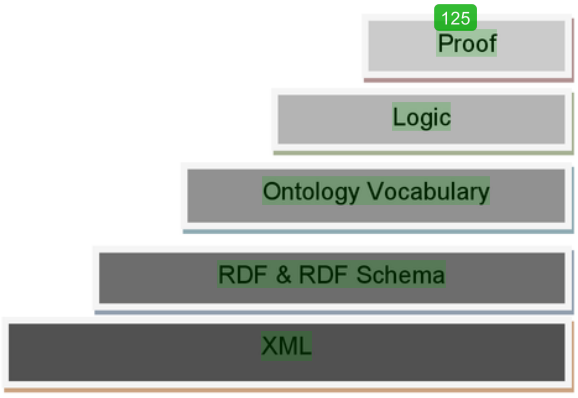

Figure 2.4 : Layers of the Semantic Web, source (El-Seoud & El-Sofany, 2010)

element of the XML document. Resource Description Framework (RDF) is a simple data model for referring objects and how they are related; RDF schema is a vocabulary for describing properties and classes of RDF resources. OWL adds more vocabulary for describing properties and classes. Since "Expressing meaning" is the main task of the Semantic Web, several layers of representational structure are needed; the basic ones being XML layer, RDF layer, ontology layer and Logic layer as shown in **figure 2.4**.

2.7.2 XML

XML is a very popular and effective way of exchanging information. It provides a surface syntax for structured documents in order to provide proficient solution to data sharing problem (Ghaleb et al. 2006). XML consists of three parts; the instance, schema and declaration, though the declaration and schema are not compulsory (malik, 2009). XML instance is a tagged document (see figure 2.5). It shows the hierarchy and boundaries of elements of which are either delimited by start tag and end tag or for empty element by empty tag (malik, 2009).An XML schema specifies the structure of an XML document while an XML declaration declares the version and encoding of XML being used (Ghaleb et al. 2006).

```
<?xml version="1.0"?>
<bookstore>
<book genre="science">
<title>Machine Learning</title>
<author>
<first-name>Tim</first-name>
<last-name>Mitchell</last-name>
</author>
<price>28.99</price>
</book>
<book genre="philosophy">
<title>The Gorgias</title>
<author>
<name>Plato</name>
</author>
<price>9.99</price>
</book>
</bookstore>
```

Figure 2.5: XML Instance[2]

[2] https://www.stylusstudio.com/docs/v2007/d_schema15.html

2.7.2 Ontologies or OWL

An ontology in information science is a vocabulary of terms used to model a domain of knowledge (Gruber, n.d.) . This vocabulary typically consists of terms for classes (or sets), properties, and relationships which are an accepted common vocabulary that provides semantics, making it achievable to apply concepts based on the regular properties of sets of data. Ontology terms usually exist in English or other languages and this provides convenience and mnemonic power, but itself not communicate meaning. The meaning is in the fact that statements have a common term to use. For example, an expression of your warehouse data as linked data would require the use OWL to explain in machine language what your metadata is and how it relates to other data in the Web of data.

2.7.3 RDF and URIs

RDF (Resource Description Framework) models' data relationships using subject–predicate–object statements. A Uniform Resource Identifier or URI represents an object or concept by holding a statement about it. Predicates are also represented by URIs. A statement therefore takes the form subject-uri predicate-uri object-uri, or subject-uri predicate-uri literal-text. RDF essentially offers an entity relationship model that presents statements about the things in our reality. RDF supports interoperability between claims that exchange machine-understandable information on the Web. RDF has three equivalent notations: RDF triples, RDF graphs, and RDF/XML. The RDF triples notation converts RDF statements directly into character strings.

Therefore, RDF results into a directed, labelled graph, which abstractly mathematically describes a graph of things. Since there is the act of connecting two resources that are identified through URIs together in a particular way there is a worthwhile claim about things. It basically presents a chance to discovering new resources in consistent way, whether these resource are stored locally or somewhere else.(Schultz, Bizer, & Isele, n.d.)

```
<rdf:RDF       xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl/org/dc/terms/">
<rdf:Description
rdf:about="http://www.example.com/books/LinkedDataHits">
<dc:title>Linked Data Greatest Hits</dc:title>
<dc:creator>Lisa Goddard</dc:creator>
```

```
<dc:contributor>Gillian Byrne</dc:contributor>
<dc:type>Text</dc:type>
<dc:publisher>Memorial University</dc:publisher>
<dcterms:issued>2010</dcterms:issued>
</rdf:Description>
</rdf:RDF>
```

Figure 2.6: An example of an RDF/XML snippet describing an imaginary book. This example combines definitions from the W3C RDF namespace, the Dublin Core Elements namespace, and the Dublin Core Terms namespace.

2.7.4 SPARQL

SPARQL is a protocol and a query language to retrieve and manipulate RDF data. SPARQL helps in querying both local and remote data sources, whether the data resides in RDF files or databases. The queries consist of graph patterns written in a fashion like Turtle (an RDF format), and allow modifiers for the patterns. (Schultz et al., n.d.)

SPARQL saves development time and cost by permitting client applications to work with only the data they're interested in. (Gür, 2012).Example: Find countries population, GDP, and HIV prevalence, ought to help find out if there is a relationship between population density and HIV prevalence. Without SPARQL, you might have to write several queries to get this information i.e. first query to pull information from population' pages on Wikipedia, a second query to retrieve GDP data from another source, and then another source to find the prevalence of HIV. A single SPARQL accomplishes all these tasks.

SPARQL language has four types of queries: SELECT, ASK, DESCRIBE and CONSTRUCT. The SELECT query is most similar to SQL in that you provide a query and you get back a response in tabular format. However, the similitude ends there. In SPARQL's SELECT query, you are asking to find resources. For example you have an example dataset in RDF as shown in **figure 2.7** below:

RDF about My Apartment

```
@prefix ex: <http://example.org/> .
@prefix taubz: <http://razor.occams.info/index.html#> .
taubz:me            ex:own       taubz:my_apartment .
taubz:me            ex:own       taubz:my_computer .
taubz:my_apartment  ex:contains  taubz:my_computer .
taubz:my_apartment  ex:contains  taubz:friends_junk .
taubz:my_apartment  ex:location  <http://example.org/Philadelphia> .
taubz:me            ex:own       taubz:my_desk .
taubz:my_desk       ex:contains  taubz:my_pens_and_pencils .
```

Figure 2.7: an example of an RDF about my apartment[3]

For instance, a query might ask for which resources "own" "my desk". That would be saying which resources x can be found in statements in the repository of the form x ex:own taubz:my_desk. The answer according to the example would be taubz:me. This differs from SQL, in that each row of results is a row from a table. Here, each row is a resource.

The reuse of existing software tooling and promotion of good interoperability with other software systems is possible since SPARQL builds on other standards including RDF, XML, HTTP, and SDL.

[3] https://www.w3.org/TR/sparql11-http-rdf-update/

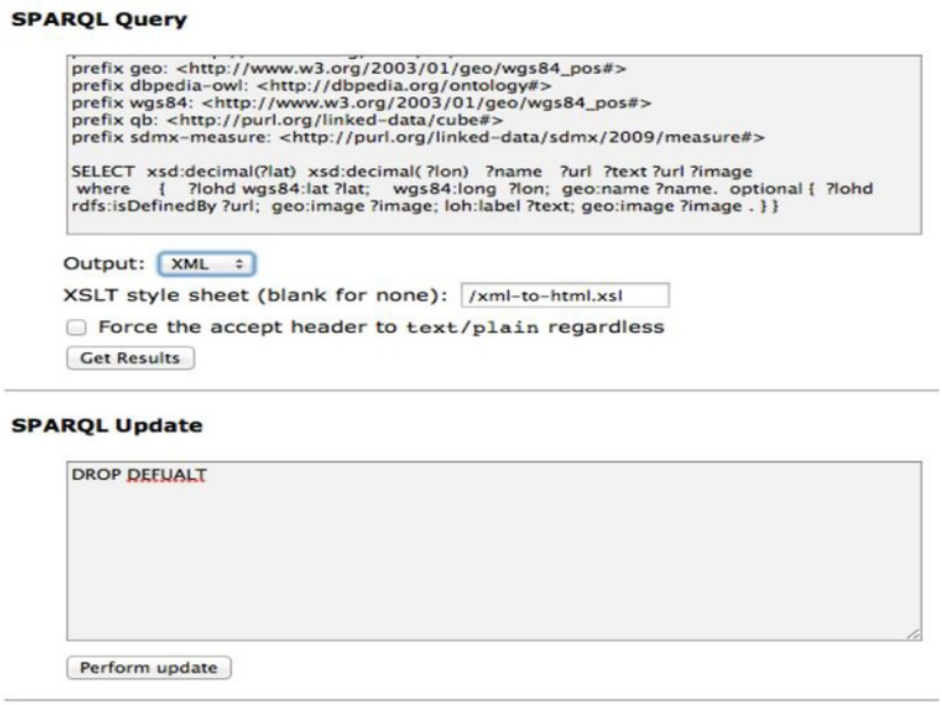

Figure 2.8: Representation of SPARQL query from FUSEKI Triple store Interface[4]

2.7.5 Graphical User Interface

A presentation layer of needed where user can interact with the data through various methods and interfaces. Linked data interfaces provide users with capability to RDF data accessible from SPARQL endpoint. From a broad perspective diverse RDF stores/tools support different querying and visualization tools. Users with knowledge and ability to program can write SPARQL queries and also choose a kind of visualization tool they need. Other end users who are not aware of SPARQL queries can use the linked data search engine to search for the information wanted just by giving a simple text like a search on Google.

2.8 Ontology vocabulary

The term ontology has been widely used in recent years in the field of Artificial Intelligence, computer and information science especially in domains such as, cooperative information systems, intelligent information integration, information retrieval and extraction, knowledge representation, and database management systems and this knowledge is in a form useful to

[4] https://www.w3.org/TR/sparql11-http-rdf-update/

both a human mind and machine (Guarino 2006). Ontology is a technique used to describe and represent an area of knowledge (Muhammad, 2013). Hence, ontology is domain specific; it is not there to represent all knowledge, but certain and specific area of knowledge. This domain specific subject area or sphere of knowledge could include medicine, education, etc. Taye (2010) describes Ontology as a key technique with which to annotate semantics and provide a common, comprehensible foundation for resources on the Semantic Web.

The typical Web ontology consists of both taxonomy and a set of inference rules whereby taxonomy defines all the classes (concepts) of objects and any relationships between them (Berners-Lee et al. 2001). Relationships between these classes can be expressed by using a hierarchical structure: super classes represent higher-level concepts and subclasses represent finer concepts, and the finer concepts have all the attributes and features that the higher concepts have. The classes, subclasses and relations allow developers to express large numbers of relations among different entities by assigning properties to classes and allowing subclasses to inherit these properties (Ghaleb et al. 2006). The inference rules allow an application to make decisions based on the classes supplied without needing to actually understand any of the information provided.

There are several formal ontology languages that are used in creating an ontology most of which are built on top of XML and RDF (Ghaleb et al. 2006).They include Ontology Inference Layer (OIL), DAML (Darpa Agent Markup Language) + OIL, and Web Ontology Language (OWL).The current version of OWL dubbed OWL 2 became a World Wide Web Consortium (W3C) standard in 2009. It is more expressive than its predecessor OWL 1 that was confirmed by W3C in 2004. OWL 2 has five sublanguages namely OWL 2 EL, OWL 2 RL, OWL 2 QL, OWL 2 DL and OWL 2 Full.

2.8.1 Logic

In this layer logics are added to the semantic content for intelligent reasoning with meaningful data.

2.8.2 Proof

After building a system that has some logic and semantic then this layer helps to prove deductions.

2.9.Conceptual Model

This work fosters a framework for open government data publishing that combines the three-stage process described by (Matasyoh, Okeyo, & Cheruiyot, 2016): Knowledge acquisition, knowledge representation, and semantic querying, with elements acquired from the Open Government Data Life-Cycle (Attard, Orlandi, Scerri, & Auer, 2015). Of specific interest in the knowledge acquisition phase, is the data selection, publishing, interlinking and exploration stages of the life-cycle. The adaptive framework employed in this research is shown in figure 2.9.

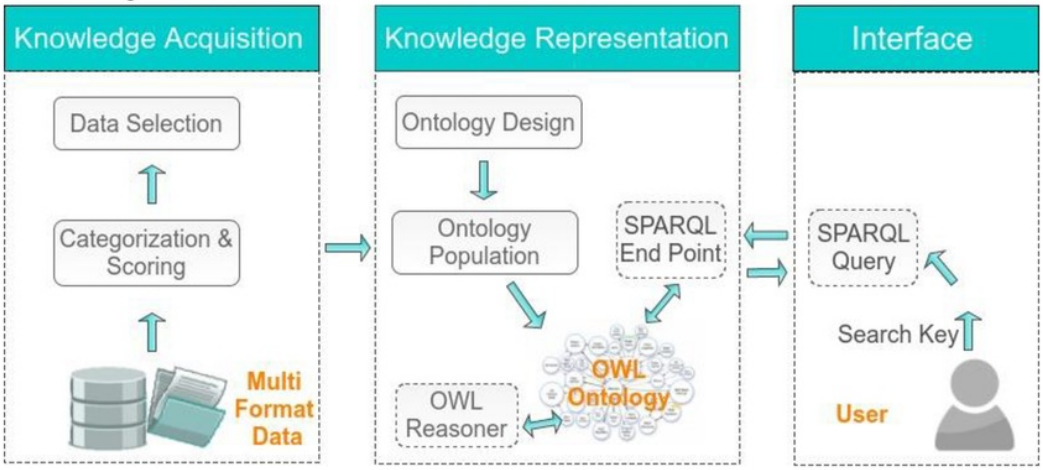

Figure 2.9: proposed conceptual framework

2.10 Practices of Ontology in E-government

Ontology is known to be employed by a number of countries employ in e-government projects (Adadi, Berrada, Chenouni, & Bounabat, 2015; Klischewski, 2003). The aim of OntoGov project is developing an ontology platform that eases the process of consistent configuration and re-configuration of e-government services as presented by (Apostolou et al., 2005). In the paper (Adadi et al., 2015), describes a methodology for building ontology in the social care domain within the context of e-government. (Gomez-Perez, Ortiz-Rodriguez, & Villazon-Terrazas, 2006)present an ontology-based model for efficient and fast retrieval of government documentation and further introduce a set of legal ontologies for the transaction domain in e-government (Gómez-Pérez, Ortiz-Rodríguez, & Villazón-Terrazas, 2006).

An ontology-based fraud detection system for e-government is presented by (Alexopuolos, Kafentzis, Benetou, Tagaris, & GEorgolios, 2007).(Herborn and Wimmer,2004) present an ontology driven semantic for business registers built to help in facilitating business

transactions amongst companies across European Union countries. (Salhofer, Stadlhofer, & Tretter, 2009) shows to us an approach of using ontology for services integration in e-government; ontology is built and the result is information for domain modelling for generation of application services.

2.11. Application of Semantic Techniques

Semantic techniques have been fruitfully used in areas such as distance learning, information systems, exam systems and medical field for data extraction, integration, and sharing. Success to knowledge sharing and retrieval systems does not lie in simple search mechanisms and having it in a repository but in how specific knowledge can be extracted from large volumes of knowledge with easy. To access only applicable knowledge from wide available, organized and structured knowledge is of essence so that it's understood by machines. Ontologies which are the pillar of Semantic Web make this a reality (Kapoor & Sharma, n.d.). Below are some of the ways in which semantic techniques have been used in enhancing knowledge sharing.

Wiki-I (Lahoud, Monticolo, Hilaire, & Gomes, 2013) has used ontologies to organize and structure the knowledge shared by engineers during research activities. Wiki-I makes use of semantic tags to tag the keywords in the wiki pages. The tags provide links to the wiki pages associated with the tag names and the origins of the ideas associated with the tag name. The system has a keyword based interface where the users can request for innovative ideas. The user simply clicks a word or types a word that directs him or her to the particular wiki page or simply uses the tags to navigate to the wiki pages that are related to the tag names.

Learning institutions are also knowledge intensive institutions that involve continuous flow of information. The ontology based examination system framework (Adebayo et al., 2011) describes an examination ontology that is developed based on the methontology ontology development. It provides semantically rich question banks to avoid repetition of questions. It also provides a means for scheduling examinations, periods and the personnel in various examination venues. The core component of this framework is an exam ontology that provides a knowledge base for the semantic examination grid. The ontology was developed using Protégé_4.0_alpha which is based on OWL-DL. Inference of the ontology is done using FaCT++ reasoner. The ontology provides common vocabulary for web based examination administration and also allows sharing of information between web based examination applications. Though the ontology provides a means for integrating several electronic examination applications for easy access and retrieval of information and can also

be reused, there is no empirical evidence for the proposed ontology. The functionality of the ontology was not tested.

Summary

Various mechanisms have been implemented to enhance knowledge sharing in knowledge intensive. These mechanisms range from use of Web 2.0 tools to Semantic Web (Web 3.0). Though web 2.0 techniques have proved to enhance collaboration in virtual environments, they still have various limitations in terms of organizing and structuring content in a way that machines can understand. Structuring of the knowledge for easy retrieval is done using ontologies which are the core component of Semantic Web (Taye, 2010). Ontologies enable representation, processing and sharing of knowledge among applications in modern web based systems because they specify the conceptualization of the specific domain in terms of concepts, attributes and relationships (Taye, 2010).

Semantic techniques have been successfully applied in various areas such as distance learning, information systems, exam systems and medical field for data extraction, integration, and sharing, but they have not been applied in knowledge sharing among research students. This research thesis therefore seeks to address the challenge by applying ontologies in sharing and retrieval of knowledge on government development projects. In this case, ontology will be used to structure shared different government projects data available in the open portal and thus advance the efficiency with which they can be accessed and retrieved as organized data. Concepts such as representing knowledge with a Semantic Web language, ontology processing, reasoning and querying on ontologies will be implemented to realize a Semantic Web application.

CHAPTER 3.0 METHODOLOGY

Research methodology defines the research activities, the process, how to measure the research results by adopting a certain methodology, which can lead to achieving the author's objectives.

3.1 Research Design

The study was divided into three parts; the first part was literature review whereby an assessment of existing materials in the field under research was done. Its purpose was to form the background of the research, to gain insights into semantic web and its techniques and to identify the success attributes of using semantic web by highlighting the existing case studies all over the world.

Secondly, an exploratory research design as defined by Burns and Groove (2001:374) was conducted to gain new insights, and discover new ideas from literature and existing similar projects. This enhances understanding of tasks required and fosters appropriate selection of task implementation and evaluation tools. This survey and case study exploration enabled also the selection of reusable components such as existing vocabularies that could be referenced or extended in our scenario. Our findings in the exploratory stage thus include, a listing of tried and tested approaches, tools and frameworks, reusable vocabulary definitions, terms and complete Ontologies coupled with design methodologies used in the development of domain ontology for Government open data.

Subsequently, we adopted a descriptive research approach that would entail identification of representational attributes from our data. The attributes such as entities and their relationships that can be mapped into RDF triple representation are identified and described, other valuable annotations on these elements including labels and comments for metadata are also included in the description. The result of this description would be a complete ontology to be used as a prototype for the research. Lastly, an experimental evaluation of the functionality of the developed application is conducted.

3.2 Data Collection and Preparation

3.2.1 Data

There is already existent open data available to the public in the Kenya Open Data portal. This data, however, exists only in semi-structured or unstructured form making access and

utilization by programmers and analysts much complicated. For our work, we sampled a section of this data and use the same to demonstrate our approach. Our sampling technique is based on three aspects:

(i) To develop ontology, we first need data that can be described in form of concepts and their relationships as opposed to data which may not be represent able as a graph based Knowledge Base.

(ii) We filter out data which is only present as images and thus may not be easily interpreted into text or may require long span of time and computational resources.

(iii) We limit the scope of the remaining data by the size of the ontology we develop which at the current state development remains an initial ontology based on one domain (governance) of the available data.

The Kenyan open data portal contains various datasets of agriculture, education, environment, energy, finance, infrastructure, governance, government accounts, health, population, and census etc. The major format of the data is in spreadsheet format (Microsoft Excel) which is a semi-structured form of data that provides full comprehension to humans but does not lend itself to machine interpretation as opposed to the general principle that open government data should be in a form that is understandable both by humans and machines. Otherwise, to convert the data as is into some useful knowledge, a collection of several chunks of data is required then merged and thus will be hectic to the users.

Following our sampling, 114 files were selected from a pool of existing records providing a total of 821 records. These documents were inherently on the topic of governance which indicates all government projects held between the years 2013 to 2017. According to our review of the data this topic had the most sizable number of documents. In addition, these documents were either in semi-structured form or textual, meaning that this could easily be employed in our work to identify required Knowledge Base entities: Table 3.1 below shows the review of the data from the Kenya government portal and their qualities based on each topic:

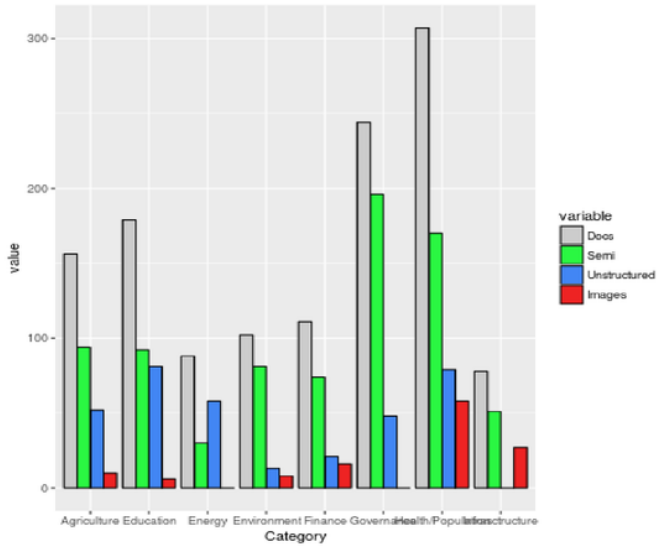

Figure 3.2.1.1 Graphical representation of data and its properties from Kenya government open portal

The approach is employed to build Ontology with specific features of interest from the governance dataset. Some of these features identified from the government development projects include information about project financiers, stakeholders, counties, supervisor, title, start and end date and addresses. This formed classes of ontology. These projects are coverage of several government departments with information about the location they are taking place at, its financiers, and beneficiaries (stakeholder) etc.

3.2.2 Software platforms for semantic Web Ontology Development

The vision of weaving open data to make it adhere to the requirements of linked which is making machine-readable content available on the Web, has seen several software platforms and application interfaces (APIs) developed to consent the automatic making and use of RDF(S) and OWL ontologies. Some of the existing RDF(S) and OWL editing platforms have been recently compared in (Kapoor & Sharma, n.d.; Khondoker & Mueller, 2010). A further in-depth list of these platforms can be found in (Kapoor & Sharma, n.d.; Knublauch, Fergerson, Noy, & Musen, 2004), (Calero, Ruiz, & Piattini, 2006); and they include Protégé, WebODE, OntoEdit, KAON1 etc.

Alongside the software platforms being used for the edition of RDF(S) and OWL ontologies, there exist APIs such as OWL API (Knublauch et al., 2004), Jena API (Wilkinson, Sayers,

Kuno, & Reynolds, 2003), Sesame (Watson, 2013), etc., that offer facilities for the persistence storage and query of RDF(S) and OWL ontologies. Protege and jena API are discussed below due to their interest in the implementation of this research.

3.2.2.1 Protégé

Protégé is an open-source platform developed at Stanford Medical Informatics. It provides an internal structure called model (Knublauch et al., 2004) for ontologies representation and an interface for the display and manipulation of the underlying model. The Protégé model is used to represent ontology elements as classes, properties or slots, property characteristics such as facets and constraints, and instances. The Protégé graphical user interface can be used to create classes and instances, and set class properties and restrictions on property facets. Additionally, Protégé has a library of various tabs for the access, graphical visualization, and query of ontologies. Protégé can be currently used to load, edit and save ontologies in different formats including XML, RDF, UML, and OWL (Knublauch et al., 2004).

3.2.2.2 Jena API

Jena is a Java ontology API which presents object classes for creating and manipulating RDF graphs called interfaces. A RDF graph is called a model and represented with the Model interface. The resources, properties and literals describing RDF statements are represented with the Resource, Property and Literal interfaces respectively. On the other hand, Jena also provides methods that allow saving and retrieving RDF graphs to and from files. The Jena platform supports various database management systems such as PostgreSQL, MySQL, Oracle, and so on; it also provides various tools including RDQL query language, a parser for RDF/XML, I/O modules for RDF/XML output, etc. (Wilkinson et al., 2003). The state of use of Semantic Web technologies in e-government is presented in the next section.

3.3 Ontology Development Process

There are five steps ontology development namely: ontology scope identification, ontology organization, building detailed ontology descriptions, ontology evaluation, and ontology commitment (Sheeba et al. 2012).

3.3.1 Ontology scope identification

This stage involves identification of the domain, purpose and users of the ontology. It brings to light if the ontology should be build from scratch or reuse the existing one.

3.3.2 Ontology organization

This stage entails design involving identification of concepts, properties and relations from the domain knowledge acquired in step one.

3.3.3 Building detailed ontology descriptions

At this stage, there is linking of properties to the different concepts and also identifying the relationships between the different concepts according to the needs of the knowledge domain that is being modeled is done. The concepts in are represented as classes in the ontology.

3.3.4 Ontology evaluation

Ontology evaluation helps ensure that it meets its prescribed and specified requirements. This process involves checking of inconsistencies in the ontology's syntax, logic, properties, semantic relationships and the overall functionality.

3.3.5 Ontology commitment

Lastly, the ontology is set out into the target environment.

3.4 OWL representation of the OntoDPM domain ontology

A framework adopted from (Uschold & King, 1995) ontology building methodology will be used alongside some knowledge found in (Dombeu & Huisman, n.d.) to build the ontology the ontology named OntoDPM in this research. The OntoDPM will indicate key concepts of the domain (projects, stakeholder, financier, supervisors etc.), the activities carried out in the domain and the relationships between the constituents of the domain.

The UML representation of the OntoDPM will be build by discovering and classifying classes and their instances in the OntoDPM and categorizing relationships between classes (Composition, association, inheritance).Classes, inheritance structure and properties/slots of the OntoDPM are used in the UML formalism for knowledge representation (Ceccaroni, n.d.). UML formalism allows modeling ontologies with instances/individuals, slots and classes, which are the terminologies of Protégé as well (Horridge, Knublauch, Rector, Stevens, & Wroe, 2004). In the formalism, a class is labelled "ontology class", a property/slot "slot relation" and an instance/individual "IndividualOf". Properties/slots represent the relationships between the concepts of the domain ontology. Each slot has a domain and a range which are labelled hasTempletateSlot and valueType, respectively. The inheritance relationship between classes is labelled "subclass-of". The resulting Protégé file

downloadable after the creation of the ontology was saved as an OWL file onto the disc. The following subsection, Java Jena ontology API will be engaged to generate the RDF formal representation of the OntoDPM.

3.5 RDF Representation of the OntoDPM Domain Ontology

Previously, we indicated that RDF is a Semantic Web language for representing resources on the Web with a resource being information put on the web that can be accessed using an URI. Three fundamental concepts in defining RDF include subject (S), predicate (P) and Object (O).

A subject is a resource being referred to on the web; a predicate is a property describing the resource and an object is a value of the property. Therefore, the triplet <S, P, O> forms an RDF graph or statement (Wilkinson et al., 2003).

An RDF graph or statement is essentially a graphical representation with two nodes and one arc; where, the arc is the predicate and the nodes at both sides of the arc, the resource and property's value respectively. Albeit, the RDF syntax represents each class of ontology as a resource which has properties with values. Thus, ontology will be represented in RDF with numerous statements in that each statement or a set of correlated statements forms an RDF sub-graph of the entire RDF graph of the projected ontology.

The RDF version of the OntoDPM will be created with the Java Jena Ontology API. Jena Model interface offers the union method needed to integrate different branches of a large RDF graph. Then, the main components (inheritance and instances) of the OntoDPM as individual Java files. In each of the Java files, the main () method will include generating and writing the corresponding RDF sub-graph into a text file. Thereafter, there will be coding of a small method to read each text file and construct its RDF sub-graphs with the Jena Model interface. Lastly, the individual RDF sub-graphs will be integrated with the union method of the Model interface, in a unique RDF graph of the OntoDPM.

3.6 Semantic Web Creation Framework Development

A rapid application development methodology will be used in the study. Rapid application development methodologies require minimal planning and in general tend to favour rapid prototyping. In most cases, planning is interleaved with the coding of the software. RAD methodologies allow for faster changes in requirement and improve the speed with which

software is written. Of the available RAD methodologies, the study will employ agile software development. Under agile methodology, the requirements and the solutions change via collaborative work. Testing of software is often conducted as it is being developed. Testing often involves two dimensions: the developers and acceptance testing focussing on the clients.

Further, two state-of-the-art Semantic Web platforms for ontology development including Protégé and Java Jena API which will be employed in generating the machine processable version of the domain ontology in OWL and RDF has been discussed in details in sections 3.3.2, 3.3.3 and 3.3.4 respectively.

Jena API grants the RDQL language for RDF storage and query with various database management systems including PostgreSQL, MySQL, Oracle (Wilkinson et al., 2003), etc. However, the scope of this research will be to get data converted into RDF and not database storage and query of the generated ontology. Furthermore, Jena API offers parsing instruments that when exploited could help read OWL ontology developed with Protégé and generate an RDF graph; with this the development of real-world Semantic Web applications will be a reality. At the end, we will have ontology edition done with Protégé, while queries are handled with Jena interfaces.

3.7 Testing and Evaluation

Ability to evaluate and compare ideas within a discipline is key factor to making it scientific and the same holds also for Semantic Web research area while dealing with abstractions in the form of ontologies. The aim of evaluating an Open Data to LOD migration approach is to determine the success attributes of the stated approach and identify points of improvements and refinements. Users employ the results of such an evaluation process to decide on whether the approach best fits their requirements in comparison to other existing ones. Likewise, it help future implementers in result evaluation and be a guide to the users' in the construction process and any refinement steps.

In our proposing an approach to the evaluation of the migration of Open data into LOD we would be carry base it to the end product rather than the process itself which in our context refers to an RDF data store. By evaluating the performance and consistency of the resultant data structuring in the form of LOD, then we are able to conclude if our approach is suitable:

A lot of literature is available on ontology evaluation specifying the various classification of ontology evaluation approaches for the different existing categories namely: methods based on comparing the ontology to a "golden standard", evaluation by using the ontology in an application and evaluating the results, approaches that compare the ontology with selected collection of source documents regarding the domain to be covered and those that involve human expert evaluation (Brewster, Alani, Dasmahapatra, & Wilks, 2004; Lozano-Tello & Gomez-Perez, 2004; Maedche & Staab, 2002; Porzel & Malaka, 2004)

In our evaluation, we will employ three approaches. First, as mentioned by (Brank, Grobelnik, & Mladenić, 2005; Porzel & Malaka, 2004) we could build a test user interface to query the data so as to observe how the data store performs when queried from an application environment. This can also be achieved by utilizing Ontology query tools such as the Apache Jena platform to query the data and act as an application proxy and therewith, also observing query runtimes. Secondly, This research has emulated the use of protégé (Fahad, Qadir, & Shah, 2008; Singh & Malik, n.d.) software platform which incorporates a reasoner component. Reasoners are employed to check the validity of the ontology by establishing if the logic exists in the parameters used. The third and final stage of our evaluation will involve comparing the accessibility of our data in structured format as opposed to the data in its original state. Methods based on comparison with source of data will also provide a good opportunity since these can be used to check if the RDF is better compared to the data in its traditional form which is easy through querying. This will then have satisfied the essence of semantic web which is data available in machine readable format.

The data used in this research is not large-scale ontology but rather a test RDF store for the migration model, it would be unjustifiable in terms of cost, scope and time if we were to carry out a full evaluation using the methods for Ontology evaluation like comparison to a golden standard or employing an expert based evaluation to carry out a full evaluation using the methods for Ontology evaluation like comparison to a golden standard or employing an expert based evaluation.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1 Introduction

For successful utilization of knowledge in governments, there has to be ways through which the knowledge can be shared between interested parties. For this reason, most governments are adopting various mechanisms through which knowledge can be effectively transferred from one party to the other. The mechanisms that have been adopted make use of web 2.0 and web 3.0 (Semantic Web) techniques. However, web 3. 0 (Semantic Web) has proved to be the most effective since it provides mechanisms through which computers can process and interpret the shared knowledge for easy access and retrieval (Taye, 2010).

The purpose of the proposed ontology driven approach was to enable government data accessible in a format that is of importance to both human and machines. This would help each interested party who would want to question government ongoing and past projects to query and have the information and also those interested in reuse of the data for research and other objectives through efficient use of the knowledge resources they have access to. The ontology is used to structure selected government projects which serves as a prototype, so that machines can reason with the content and provide meaningful results to the interested persons.

4.2 Overview of the Approach

This approach makes use of ontology in structuring shared content for easy access and retrieval. It is mainly an application for accessing and querying information that is stored in the OWL ontology located on the web or local system. The content domain is data from government open portal shared from different government departments. The ontology is populated using data from one single domain.

The developed ontology not only structures the shared content but also allows easy access and retrieval of the content. All the information presented after a query through the application is extracted from the ontology. The ontology processing is done using Apache Jena which is a Semantic Web framework for Java (apache, n.d). The application retrieves information from the ontology using Jena API. Loading and creation of a model of the OWL ontology, content extraction and querying is done by the Jena framework.

Though the Jena framework can be used to create ontologies, the OWL ontology in this platform was created using an external editor then loaded into the application using Jena. This is because of availability of powerful editors. The interface of the application is friendly

and easy to use and allows users to navigate through and search the shared content and get results according to certain specifications.

4.3 Design of the Knowledge sharing approach

The ontology driven platform for sharing and retrieval of knowledge is made up of three main parts: the OWL ontology, query engine which is a SPARQL processor for Jena and the application interface for interacting with the platform as shown in figure 4.3.1. The main information resource for the application was the constructed ontology.

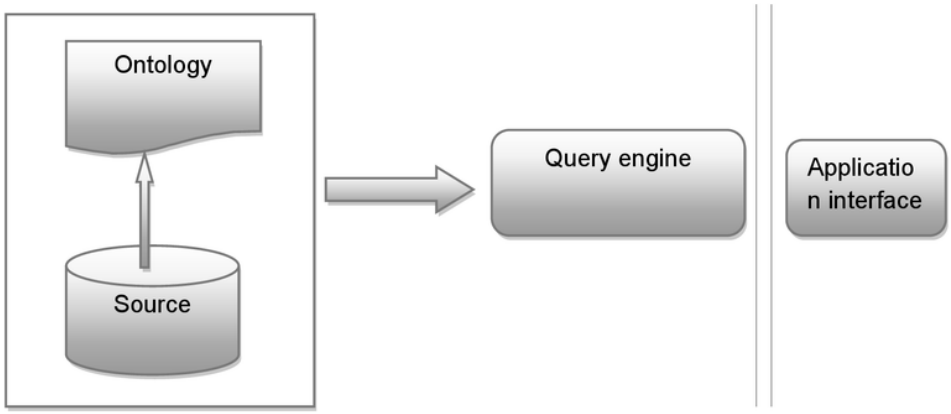

Figure 4.1: Knowledge sharing approach

The query engine, implemented using Jena, acts as a bridge between the ontology and the application interface. Through Jena framework, the OWL file is loaded into the application to enable execution of SPARQL queries. After loading the ontology, it is populated with individuals from the data source. The ontology could not be pre-populated in order to comply with the dynamic nature of the data source.

The interface, a java application, allows interaction between the user and the ontology. It accepts input from user, sends it to the query engine which processes the request and returns results that are displayed to the user through the interface. Whenever the contents of the data source change, the contents of the ontology automatically

4.4 Description Logics Background

Description Logics (DL) is the widely used knowledge representation language in ontology modelling since it provides one of the main supports to Web Ontology Language (The OWL working group, 2009). It helps model the relationships between entities in any domain of interest. DLs make use of concepts (classes), roles (properties or relationships) and

individuals to model a domain. Concepts are sets of individuals; roles are relationships between the individuals while individuals are single individuals in the specific domain (Steigmiller, 2015).

DL supports various languages that have been proposed and are distinguished based on the types of axioms and constructs supported. This means that the DL language of a knowledge base is characterized by the constructs it supports. DLs range from less expressive to more expressive ones that support more constructs. For instance, there are a variety of inexpressive DLs such as EL and DL-lite that support a few constructors. Such DLs are referred to as tractable and are best suited for developing very large knowledge bases. There also exists semi-expressive DLs such as Attributive Language with Complement (ALC) that allow constructs like disjunctions, negation, conjunctions, existential and universal quantifiers (Steigmiller, 2015). Finally, there are expressive DLs such as DL SROIQ that support many language constructs. Although the reasoning performance of such expressive languages is much, the expressiveness of helps in adequately describing a particular knowledge domain. DLs should therefore be both expressive and decidable. One such language that offers both aspects is DL SROIQ.

Since DL SROIQ is the logical underpinning for OWL 2 DL (Steigmiller, 2015) that is used in our ontology construction, it automatically becomes the DL used in our knowledge base. The knowledge base of description logic consists of sets of statements called axioms that partially describe a state of the world. These axioms are divided into terminological (TBOX) axioms, relational (RBOX) axioms and assertion (ABOX) axioms. An ontology can be basically a TBOX because it describes a domain in terms of concepts and relationships between the concepts. These relationships are created using roles

4.4.1 Assertion (ABOX) axioms

Axioms describe properties (facts) of named individuals (Krötzsch, Simančík, & Horrocks, 2012). They describe relationship between individuals and concepts to which individuals belong. The axioms mostly include:

4.4.1.1 Concept assertions:

They state the concepts to which individuals belong. For example, the assertion:

```
Project (project_001)
```

This assertion simply states that the individual project_001 is an instance of the concept project.

4.4.1.2Role assertions:

They describe the relationships between named individuals. For example, the assertion:

`hasSupervisor(supervisor001, proj001)`

The assertion states that the individual proj001 is in a relation with the individual supervisor001 through a relationship represented by hasSupervisor.

We should note that same names may refer to same individuals names unless are explicitly stated i.e DLs do not employ *Unique Name Assumptions.*This can be done through:

4.4.1.3 Individual inequality assertions:

States that two names refer to two different individuals. For example, the assertion:

person ≈ human

The assertion simply states that person and human are different individuals.

4.4.1.4 Individual Equality Assertions:

States that two names refer to the same individual. For example, the assertion:

Woman ≈ lady

The assertion states that the individuals named woman and lady are the same.

4.4.2 Relational (RBOX) axioms

The RBOX axioms describe relationships between roles (Krotzsch, Simancik, & Horrocks, 2014). DLs have an element that supports both *role inclusions* and *role equivalence* axioms. For example, the following role inclusion axiom:

`hasConclusion ⊑ hasSection meaning that hasConclusion is a subrole of hasSection` and therefore all the pairs of individuals related by hasConclusion are also related by hasSection.

Role equivalence on the other hand states the roles that are the same. For instance, the axiom:

hasAuthor ≡ hasWriter

The axiom states equality or sameness between the role hasAuthor and hasWriter, that is, any pair of concepts related by the role has Author can also be related by the role hasWriter.

RBOX axioms also include *disjoint roles* and *role characteristics* (Rudolph, 2011). With disjoint roles, two individuals cannot be related at the same the same time. Role

characteristics on the other hand, describe properties of roles such as reflexivity, transitivity, symmetry etc.

4.4.3 Terminological (TBOX) axioms

They are axioms that describe relationships between concepts in a domain (Krotzsch et al., 2014). The axioms are described in terms of general *concept inclusions* and *concept equivalences*. For instance, the fact that all projects are government projects is expressed by the *concept inclusion*:

projects ⊑government projects

The statement simply means that the concept project is subsumed by the concept government projects.

Concept equivalence on the other hand, shows that two concepts have the same individuals/instances. For instance, the axiom:

writer ≡ author

This axiom simply states that the concept writer and author have the same instances. The two concept expressions can be used with other concept constructors to define other complex axioms. These constructors include:

a) ***Boolean constructors***

These constructors provide basic Booleans operations that are related to union, intersection and complement of sets or conjunction, disjunction and negation of logical expressions (Krotzsch et al., 2014). For instance, through concept inclusions, a journal_article can be described as a paper that belongs to a journal. This statement is formed using the *intersection* (*conjunction*) constructor and the equivalence concept expression.

journal_article ≡ paper ⊓ belongsTo.journal

Meanwhile, *Union (disjunction)* is the dual of intersection. For example, the axiom:

paper ≡ research_material ⊓ belongsTo.(conference ⊔ journal)

The axiom gives the set individuals that are research materials and belong to either a conference or a journal.

Complement (negation) is used to describe the sets of individuals that do not belong to a certain concept. For instance describing papers that do not belong to a journal paper is stated as follows:

Conference_paper ≡ paper ⊓ belongsTo. ¬journal

Top concept, ⊤, is a special concept that comprises all the individuals (Rudolph, 2011).

Bottom concept, ⊥, is the dual of top concept and is a special concept with no individual (Rudolph, 2011).

b) ***Role restrictions***

Aside to concept constructors, DLs in addition provide means for linking concepts and roles to make complex axiom (Krotzsch et al., 2014). For example, the concepts data, date and the role hasDate can be linked by the fact that a proj001 has some supervisor supervisor001. This can be captured as follows:

Proj001 ≡ project⊓∃hasSupervisor.supervisor001

The *existential restriction*, ∃, used in the axiom shows that proj001 is a project that has some supervisor. *Universal restriction,* ∀,on the other hand, is used to show sets of individuals that have the same value (Rudolph, 2011).

Number restrictions allow restriction of the number of individuals that can be reached through a role (Rudolph, 2011). For instance, the axiom describes the *at-least restriction*:

Proj001 ≡ project ⊓≥1hasSupervisor.supervisor

The axiom states that a proj001 is a project with at least 1 supervisor.

The *at-most restriction* below states that a proj001 is a project with at most 2 supervisors:

Proj001 ≡ project ⊓≤1hasSupervisor.supervisor

The two can be combined to represent an axiom that states that a proj001 is a project with exactly two supervisors:

Proj001≡project⊓≥2hasSupervisor.supervisor ⊓≤2hasSupervisor.supervisor

c) ***Nominals***

Another way of defining concepts is by enumerating their instances. Enumeration is not natively supported by DLs, but nominals allow description of concepts by enumerating them. A nominal is a concept that has only one instance (Krotzsch et al., 2014) . For example, the concept **category** can be described by enumerating its instances: `proj001`, `proj002`, `proj003`, and `proj004`. `{ proj001}` in this case is a concept that has only one instance/individuals named `proj001`.

`project ≡ {proj001}` ⊔ `{ proj002}` ⊔ `{ proj003}` ⊔ `{ proj004}`

Using nominals concept assertions can be turned to concept inclusions. For instance the concept assertion **category (countyproject)** can be written as:

{countyproject}⊑ category

d) ***Role constructors***

DLs provide only few constructors that can support construction of complex roles and the largely essential constructor in this case is the *inverse role* constructor (Rudolph, 2011). For example, the relationship between hasSupervisor and supervised can be represented using *inverse role* as follows:

$\mathbf{hasSupervisor \equiv supervised^{-}}$

The complex role supervised^{-} represents the inverse of hasSupervisor.

4.5 Description Logics Reasoning

DLs are founded on Open World Assumptions (OWA) where there's an assumption information given in the knowledge base is not enough to determine the logical consequences of the knowledge base (Krotzsch et al., 2014). For instance, from the ABOX assertions hasAddress(county_addr001), project(proj001), and address(county_addr001), it cannot be assumed that proj001 has only one address (county_addr001) since there might be other addresses for proj001 that are not explicitly stated in the knowledge base.

Also, there's no Unique Name Assumptions (UNA) in DLs (Steigmiller, 2015) Such that individuals are not considered dissimilar just for the reason of them having different names. A computer system looks at entity names of a knowledge base as just sequences of characters used to identify the entity. For example in DLs, the names john and michael could be referring to the same person unless explicitly stated as different.

Description Logics have been designed to deal with incomplete information instead of making default assumptions to wholly describe one interpretation of a knowledge base, Instead DLs consider all the possible world situations (*interpretations).*In this case, the consequences that hold in all the likely interpretations that obey the limitations of the knowledge base are considered.

The only *interpretations* needed are those satisfying the restrictions of the knowledge base, that is, the ones that are compatible with the stated axioms. These interpretations are called *models* of a knowledge base and are used in automated reasoning to determine the logical consequences of a knowledge base. A logical consequence of a knowledge base is an axiom that is true in all interpretations that obey the restrictions of the knowledge base

(Krotzsch et al., 2014).A knowledge base with no model is termed as inconsistent or unsatisfiable (Krotzsch et al., 2014). DLs can also be termed as monotonic meaning that for cases where more axioms are added to the knowledge base, the existing logical consequences will still hold. Additional axioms only increase the knowledge base's logical consequences and not changing the existing ones.

As earlier on stated, inferencing in DLs is based on *interpretations* (how things appear possibly in real world states) that satisfy the limitations of a knowledge base. However some conditions that must hold for a particular axiom to be satisfied by an interpretation. An interpretation, denoted by I, that provides: Anon-empty set Δ^Ithat represents all the individuals in the domain of interest (Rudolph, 2011).

- An interpretation function, $\cdot^I$ used to connect the vocabularies (concepts, roles and individuals) of a knowledge base with Δ^I(Rudolph, 2011)by providing:
  - For each individual, $\mathbf{a} \in \mathbf{N}_I$, a corresponding $\mathbf{a}^I \in \Delta^I$ from the domain of interest
  - For each concept $\mathbf{A} \in \mathbf{N}_c$, a corresponding set $\mathbf{A}^I \subseteq \Delta^I$
  - For each role name$\mathbf{r} \in \mathbf{N_R}$, a corresponding set$\mathbf{r}^I \subseteq \Delta^I \times \Delta^I$of ordered pairs of elements from the domain.

Visibly, the semantics of different knowledge base vocabularies depend on interpretations. Therefore, the formal semantics of complex concepts and roles depends on the semantics of basic entities of the knowledge base. **Table 5.1**(Krötzsch et al. 2013) shows the semantics of compound expressions are derived from the semantics of basic entities. **Table 5.2** (Krötzsch et al. 2013), on the other, shows the semantics of SROIQ Axioms. From the semantics represented in the two tables, the semantics of complex axioms can be obtained.

Table 4.1 : syntax and semantics of SROIQ constructors

	syntax	**Semantics**
Individuals:		
Individual name	A	a^I
Roles:		
Atomic role	R	R^I
Inverse role	R^-	$\{\langle x, y\rangle \mid \langle y, x\rangle \in R^I\}$
Universal role	U	$\Delta^I \times \Delta^I$

Concepts:		
Atomic concept	A	A^I
Intersection	$C \sqcap D$	$C^I \cap D^I$
Union	$C \sqcup D$	$C^I \cup D^I$
Complement	$\neg C$	$\Delta^I \setminus C^I$
Top concept	$\top$	Δ^I
Bottom concept	$\bot$	$\emptyset$
Existential restriction	$\exists R.C$	$\{x \mid \text{some } R^I\text{-successor of } x \text{ is in } C^I\}$
Universal restriction	$\forall R.C$	$\{x \mid \text{all } R^I\text{-successors of } x \text{ are in } C^I\}$
At-least restriction	$\geq n\, R.C$	$\{x \mid \text{at least } n\ R^I\text{-successors of } x \text{ are in } C^I\}$
At-most restriction	$\leq n\, R.C$	$\{x \mid \text{at most } n\ R^I\text{-successors of } x \text{ are in } C^I\}$
Local reflexivity	$\exists R.Self$	$\{x \mid \langle x, x\rangle \in R^I\}$
Nominal	$\{a\}$	$\{a^I\}$

where $a, b \in N_I$ are individual names, $A \in N_C$ is a concept name, $C, D \in \mathbf{C}$, are concepts, $R \in \boldsymbol{R}$ *is a role*

According to **Table 4.1** the semantics of different entities of Knowledge base is given by their interpretations from the domain of interest. From the semantics of the different vocabularies, the semantics of complex concepts is derived. For instance, $\exists R.C$, is a complex concept that is linked with a role. The semantics ($\{x \mid \text{some } R^I\text{-successor of } x \text{ is in } C^I\}$) simply states that some individual, y(the successor of x), that constitutes the role interpretation (R^I)is an instance of C."R^I-successor of x" simply means any individual y such that $\langle x, y\rangle \in R^I$.

Table 4.2: syntax and semantics of SROIQ axioms

	Syntax	Semantics
ABox:		
Concept assertions	$C(a)$	$a^I \in C^I$
Role assertion	$R(a, b)$	$\langle a^I, b^I\rangle \in R^I$
Individual equality	$a \approx b$	$a^I = b^I$
Individual inequality	$a \not\approx b$	$a^I \neq b^I$

TBox:		
Concept inclusion	$C \sqsubseteq D$	$C^I \subseteq D^I$
Concept equivalence	$C \equiv D$	$C^I = D^I$
RBox:		
Role inclusion	$R \sqsubseteq S$	$R^I \subseteq S^I$
Role equivalence	$R \equiv S$	$R^I = S^I$
Complex role inclusion	$R_1 \circ R_2 \sqsubseteq S$	$R^I_1 \circ R^I_2 \subseteq S^I$
Disjoint roles	$Disjoint(R, S)$	$R^I \cap S^I = \emptyset$

Since interpretations describe the meaning of different entities, it is easy to know whether an axiom holds in an interpretation or not. An axiom α holds in an interpretation I or I satisfies α ($I \models \alpha$), if the condition in table is met.

Reasoning is a very vital in ensuring the quality of ontologies both in design and deployment phases of ontology development. During design, reasoning is to test if the concepts are non-contradictory and to derive the implied relations between concepts and roles. While in development, it is used to determine the consistency of axiom and inferring additional knowledge. Reasoning tasks provided by most ontology reasoners include:

- ***Knowledge base satisfiability:***

  This checks the consistency of the knowledge base. A satisfiable (consistent) knowledge base (KB) is a knowledge base that has atleast one model, that is, there is an interpretation,I, that satisfies KB (written as $I \models KB$) (Rudolph, 2011).A contradictory knowledge base can lead to modeling errors so consistency in a knowledge base is very satisfying and so should be ensured.

- ***Axiom entailment:***

  A knowledge base (KB) entails an axiom (α), written as $KB \models \alpha$, iff every model of the knowledge base is also a model of the axiom. Axiom entailment can be simply described as querying of a body of knowledge to check if a certain statement is true. This can also be used in deciding KB satisfiability through proof by contradiction (Rudolph, 2011). Proof by contradiction is proving that something holds by assuming the opposite and deriving the contradiction from the assumption. For instance if two axioms, α and β, are the opposite of each other, then every interpretation (model) of the KB is only one of the two axioms but not both. It is therefore easy to decide KB satisfiability using an axiom α as long as there is an axiom that is opposite of α.

- ***Concept satisfiability:***

  Concept satisfiability is checking whether a concept has an individual. A concept is satisfiable if there is an interpretation, I,of the KB that maps the concept to a non-emptyset in the domain(Rudolph, 2011). However, there are concepts that will always be unsatisfiable regardless of the knowledge provided in the knowledge base such as $\bot$ and $C \sqcap \neg C$.

- ***Instance checking:***

  This involves checking whether a certain individual belongs to a specific concept. Since knowledge base has several models and the domains keep changing, a specific individual maybe an instance of C in one model and not in the other. However, instance checking can be made successful by restricting the task to named individuals (Rudolph, 2011). This task can therefore be implemented as: given the knowledge base KB and a concept C, get all the individuals named $b \in N$where $b^I \in C^I$ in every model of the KB.

- ***Subsumption:***

  Determining subsumption is checking whether a concept C (subsume)always symbolizes a subset of the set denoted by D (subsumer)(Sengupta, & Hitzler, 2014).It is always written as$C \sqsubseteq D$.

- ***Classification:***

  Classification involves generation of subsumption relationships(Sengupta, & Hitzler, 2014). The concepts of a knowledge base can be put into an hierarchy basing on their subsumption relationships. Classification can be used as a preprocessing step that speeds up subsequent reasoning tasks on the knowledge base(Rudolph, 2011).

A reasoner is a software that derives logical consequences of a knowledge base using the explicitly stated axioms and facts (Sengupta, & Hitzler, 2014). Many reasoners use First Order Predicate Logic to perform reasoning whereby it's either backward chaining or forward chaining or both. Forward chaining derives valid inferences from known facts and it mainly infers additional knowledge to the knowledge base and answers queries. (Abburu, 2012). Backward chaining, on the other hand, begins from a query or a particular fact and tries to prove it.

Reasoners are classified according to the features they support and the methodology used in reasoning and they include; rule based reasoning, OWL API, ABOX reasoning, Jena API, and protégé among others. (Abburu, 2012). Methodology, on the other hand, describes the algorithm or technique used by a reasoner to solve basic reasoning problems (Abburu, 2012). These methodologies include automata based techniques, resolution based techniques, and tableau based techniques among others. However, the most widely used is the tableau based technique.

A semantic tableau is a tree based algorithm that is used to prove the satisfiability of a knowledge base. The nodes on the tree represent individuals which are labelled by concepts or complex concepts. The edges between the nodes are labelled using role names that link the corresponding individuals. Given a knowledgebase $K = (T, R, A)$, it is assumed that all the formulas in the knowledge base are in Negation Normal Form (NNF) (Steigmiller, 2015). This means that the negation is applied to the concepts only. NNF is achieved using De Morgan's laws and the duality between universal and existential quantifiers.The basic idea behind the algorithm is trying to prove the consistency of K by constructing a model of K. It does this by starting from the situation described in A then explicating addition constraints on the model implied by the axioms in$A$, R *and* T. Semantic tableaux is used to test whether:

1. A formula F is valid,
2. A formula G is a logical consequence of$F_1 \ldots \ldots F_K$
3. $F_1 \ldots \ldots F_K$ is satisfiable.

A path of a tableau is said to be closed if the path contains a conjugate pair of formulas, F *and* $\neg F$ (Steigmiller, 2015).

- A formula F is tested for validity by creating tableau that starts with the formula $\neg F$.
- A formula G is tested to be a logical consequence of $F_1 \ldots \ldots F_K$ by creating a tableau that starts with $F_1 \ldots \ldots F_K, \neg G$.
- $F_1 \ldots \ldots F_K$ is tested for satisfiability by forming a tableau that starts with $F_1 \ldots \ldots F_K$. if the tableau does not close then the $F_1 \ldots \ldots F_K$ is satisfiable.

4.6 Modelling an Ontology

In the above section there's information provided about use of DLs in ontology modelling, this is vital in the modelling of the information sharing ontology documented here. The axioms represented in this section only include TBOX axioms and RBOX axioms. The ABOX axioms are not included because the instances or individuals in the ontology keep on changing due to the dynamic nature of research content.

4.6.1 TBOX axioms

TBOX axioms comprise the relationships between concepts. Roles linked with concepts and role restrictions form complex axioms. The concepts in this research comprise of the below:

a) ***Govt project***

Project is the main concept that contains the many existing projects within the government of Kenya that are described through the relationships created between it or its sub-concepts and other concepts in the ontology. The concepts; address, county, date, financier, objective, stakeholder, supervisor, and title are linked with their respective roles and role restrictions to form a complex axiom that describes the concept project**.**

Apart from the concept project which is the main class, there are other concepts that contain individuals, relations and concepts that are used in this research ontology. The concepts include address, county, date, financier, objective, stakeholder, supervisor and title. They are described as in **figure 4.2**

```
address ⊑∃isAddressOf.(county ⊔ financier ⊔ stakeholder ⊔ supervisor )
county ⊑∃isCountyOf.project ⊓∃hasAddress.address
date ⊑∃isEndDateOf.project ⊔∃isStartDateOf.project
objective ⊑∃isObjectiveOf.project
stakeholder ⊑∃isStakeholderOf.project
supervisor ⊑∃isSupervisorOf.project ⊓∃hasAddress.address
title ⊑∃isTitleOf.project
```

Figure 4.2: description of other knowledge base concepts

4.6.2 RBOX Axioms

RBOX axioms describe relationships between roles and in this ontology roles are related using the two major expressions namely role equivalence and role inclusion and the inverse role constructor. They are described in **figure 4.3** .

```
hasFinancier ≡ financier ⁻
hasAddress ≡ isAddressOf ⁻
hasCounty ≡ isCountyOf ⁻
hasEndDate ≡ isEndDateOf ⁻
hasStartingDate ≡ isStartingDateOf ⁻
hasObjective ≡ isObjectiveOf ⁻
hasStakeholder ≡ isStakeholderOf ⁻
hasSupervisor ≡ isSupervisorOf ⁻
hasTitle ≡ isTitleOf ⁻
```

Figure 4.3 : description of roles

4.7 Ontology Development

This section presents implementation of the designed knowledge base. In the OWL ontology, concepts are represented as classes, roles as the different properties that describe the classes while individuals are the different instances of the classes.

4.7.1 OWL classes

The ontology built for this thesis is made up of several classes just as described in the design phase earlier on. We have a number of classes namely; address, county, date, financier, objective, project, stakeholder, supervisor, title. Relations between classes (concepts) is formed using object properties (roles). **Figure 4.3 is** an illustration of some of the classes in the ontology and their relationships. Solid lines indicate the class hierarchies while dotted lines represent relationships. Solid lines indicate the class hierarchies while dotted lines represent relationships.

4.7.2 Properties and relations

This section presents the properties and relations used in the ontology. As stated earlier, relations between different classes (concepts) are represented using object properties (roles). The object properties include; *hasInstitution, hasAddress, hasCategory, hasAuthor, hasTitle,* and *hasSupervisor* among others.Each of the object properties has an inverse property (inverse role). For instance, the relation between the classes proposal and author, is that a proposal has an author: represented by the *hasAuthor* object property (figure 5.4). Conversely, an author authors a proposal: represented by the inverse of *hasAuthor,* which is, *authored.* Datatype properties, on the other hand, describe the characteristics of the instances of a class. They include; *edition, year, volume, name,* and *url* among others. However, the

figure does not show all the relations in the ontology, it only dwells on relations between *proposal* class and some of the classes in the ontology.

4.8 Analysis and Discussion

Discussions around exchange of knowledge between different people indicate various difficulties experienced and government data isn't an isolation case to this. Citizens find getting useful government data for reuse an uphill task. Several barriers are revealed that hinder the sharing of knowledge or trying to access the available knowledge. Finding ways to manage the barriers could enhance sharing of knowledge among these groups of people and also enhance access to knowledge. This would in turn improve information sharing and awareness process.

4.8.1. Knowledge Sharing Habits

An important conclusion that can be drawn is that government data is useful to every citizen for reuse in different ways including researching. It's the only free and unlimited data with most knowledge covering all areas of the society. The difference is in the extent to which each user can share, the knowledge shared and the platform used. Most governments have moved their data online and therefore it's easy to get the materials from the internet. However, most of them rely on the internet as the main source. The materials from the internet are accessed through online search. Apart from the internet, government data can also be retrieved from archives from the files in different departments or institutions. Though the world with high speeds has made internet the main source of access to information. This has pushed most governments to put their data available online.

4.8.1.1. Importance of Knowledge Sharing

Knowledge sharing is practiced by all the governments because it is beneficial to the research process. The results unveil some of these benefits:

- Enhancing knowledge

Knowledge sharing enhances citizens' knowledge in various aspects. This is depicted as the greatest benefit of knowledge sharing. Through sharing knowledge, citizens gain more understanding on the areas that they have little or no knowledge about.

- Quickening access to information

Sharing knowledge saves on the time spent looking for knowledge content. Therefore, instead of spending a lot of time searching for materials, one could get knowledge from other citizens, which is quicker.

- Creating environment for interaction

Sharing knowledge creates a good environment for interaction amongst citizens. This interaction can be important in building trust among themselves. If people get to understand each other through interaction, then one of the challenges, that is, lack of trust, could be easily avoided.

4.8.1.2. Barriers to Knowledge Sharing

Despite the benefits of knowledge sharing, there are various barriers to knowledge sharing among the citizens. Though the barriers vary across various fields of study, there are some that are experienced across all the fields. The barriers include:

- Lack of proper platform for sharing knowledge

The lack of proper platform is a challenge that is experienced across all the fields. A convenient platform by the government is recommended because citizens live in different locations and should therefore have a platform that most of them can access. Another reason for a good platform is that these citizens require a lot of assistance from experts who cannot be easily located. Having this platform could make it easy for them to access the different experts. The platform can also reduce the time spent searching for knowledge in the vast knowledge from different experts that is found on the internet.

- Poor perception of knowledge sharing

This can be linked to lack of open minded sharing though in this case, some government departments/institutions may completely avoid sharing knowledge with others. This is probably because they feel knowledge is a personal property that should be owned by an individual. Some are entrenched into the traditional way of doing things where they believe that government data is secretive.

- Lack of trust

This can be attributed to governments fear that its data will find way into hands of wrong people or governments. These to some departments or institutions mean that it exposes government and its work to its enemies. Though knowledge sharing has proved to be very important during research, the mentioned barriers can easily prevent or hinder its effectiveness. Certain measures should therefore be taken to ensure adequate sharing of knowledge of government data.

4.8.2. Access and Retrieval of Knowledge from the Web

The use of electronic resources can be attributed to the fact that most of people are connected to the internet as a source of knowledge. Government resources are easily obtained from the

web in form of portals. However, the process of locating the materials on the web is difficult. The few who find it very easy are those who know what they want. That is they know the specific materials they are looking for. This problem can be solved by implementing a mechanism that simplifies search of materials on the web.

4.8.2.1. Barriers to Access and Retrieval of Knowledge

Though electronic resources have proved to be very helpful to citizens accessing government data, there are some challenges that hinder access to the resources. The dissatisfaction in the process of locating the resources. As mentioned above, the process of locating the resources difficult. The difficulties experienced are as a result of:

- ***Difficulty in locating resources***

Difficulty in locating resources has proved to be the greatest problem experienced by citizens while trying to get knowledge from the web. Most of them find it difficult to locate the required content from the vast knowledge that is found on the web. As much the web is rich in knowledge, obtaining the relevant content from the enormous search results has proved to be difficult. One is forced to go through a lot of information in order to identify the one that is relevant to his or her request.

- ***Lack of proper resources***

A part form difficulty in locating required resources, there's a feeling that there are no appropriate resources for certain domains of government. Though the web contains a lot of knowledge, one may not find the desired knowledge. At the end of the day, there's a possibility of not getting the knowledge that fully satisfies their area of interest.

- ***Insufficient resources***

This can be linked to lack of proper resources in the fields of less interest and therefore finding good and up to date knowledge to use in such fields is difficult.

To speed up access to government data, mechanisms should be implemented that try to solve the problems experienced during knowledge access and retrieval from the web. The mechanisms should enable easy access and retrieval of knowledge thus saving on time.

4.9 Ontology Presentation

This section presents implementation of the designed knowledge base. In the OWL ontology, concepts are represented as classes, roles as the different properties that describe the classes while individuals are the different instances of the classes.

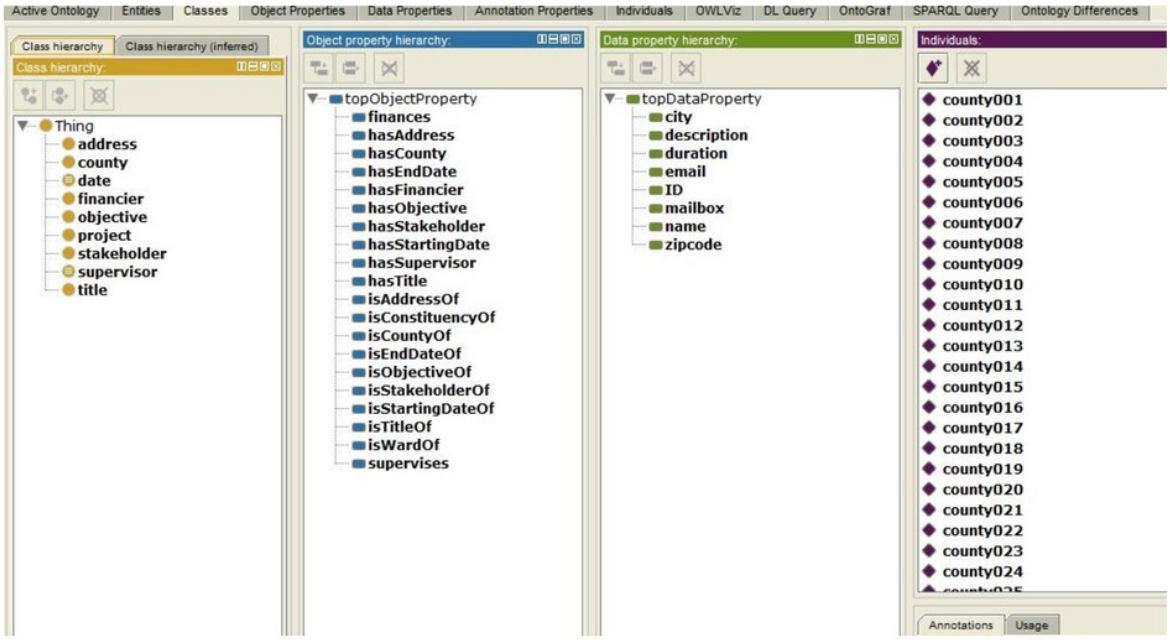

Figure 4.4: Protégé version of the ontology

4.9.1 OWL classes

This ontology is made up of several classes just as described in the design phase. Relations between classes (concepts) are formed using object properties (roles). Figure 4.6 is an illustration of some of the classes in the ontology and their relationships. For instance, the relationship between project and other classes like address and financier is that; a project has an address; it has a financier and also has a stakeholder. Some of this project properties are r inherited across the other classes. Solid lines indicate the class hierarchies while dotted lines represent relationships. Solid lines indicate the class hierarchies while dotted lines represent relationships. However, there are so many classes and many relationships between them that are not depicted in figure 4.6.

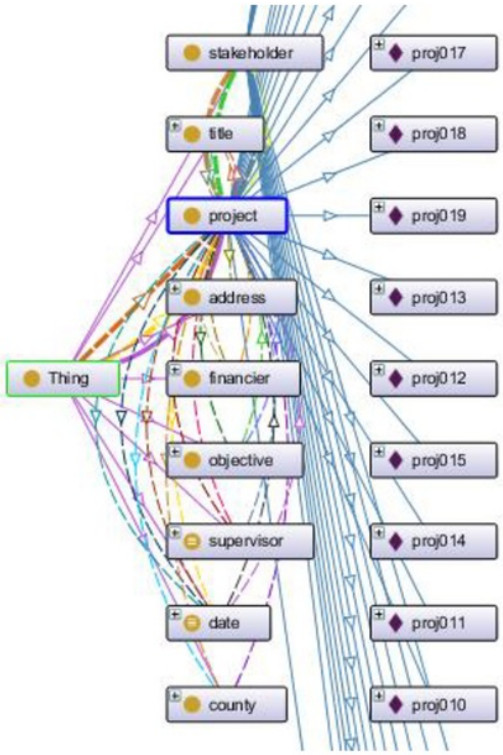

Figure 4.5: ontology classes

4.9.2 Properties and relations

This section presents the properties and relations used in the ontology. As stated earlier, relations between different classes (concepts) are represented using object properties (roles). The object properties include; *hasAddress, hasCounty, hasEndDate, hasStartingDate, hasFinancier, hasObjective, hasStakeholder, hasTitle,* and *hasSupervisor* among others. Each of the object properties has an inverse property (inverse role). For instance, the relation between the classes project and title, is that a project has a title: represented by the *hasTitle* object property (figure 4.7). Datatype properties, on the other hand, describe the characteristics of the instances of a class. They include; *city, email, mailbox, name,* and *zipcode* among others. However, the figure does not show all the relations in the ontology, it only dwells on relations between *proposal* class and some of the classes in the ontology.

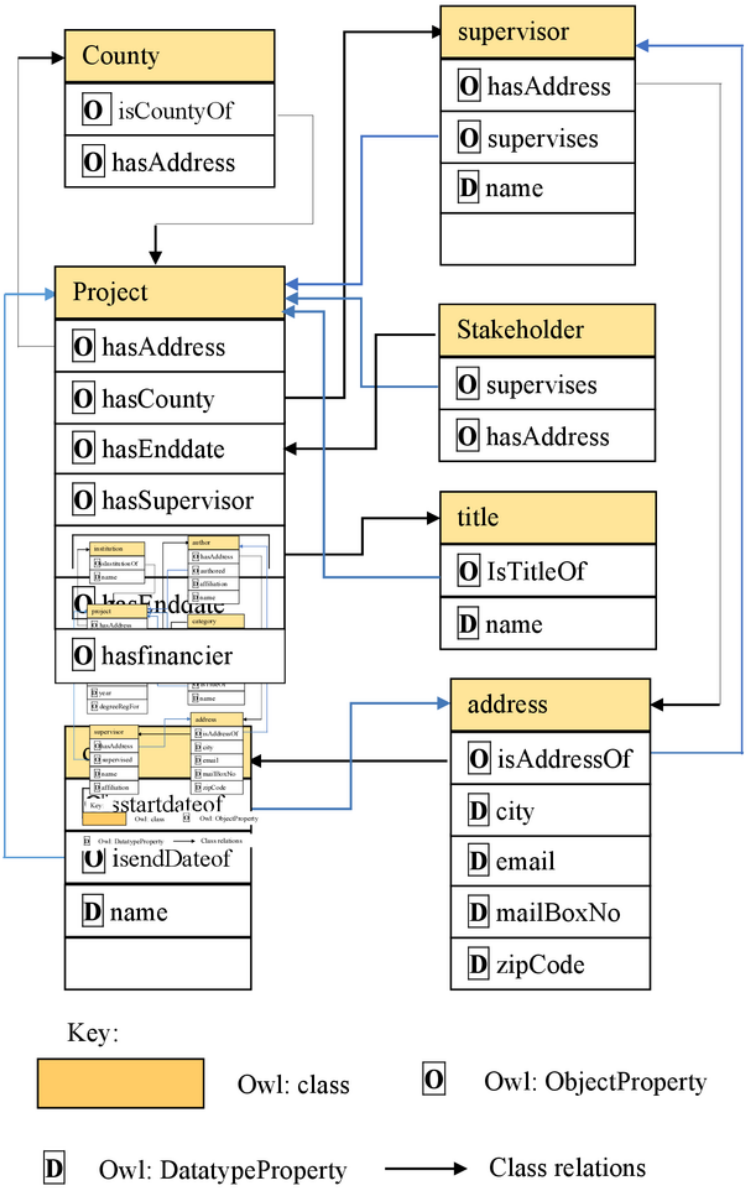

Figure 4.6: class relations

4.9.3 The Interface

This section gives a detailed discussion about the application interface and query processing. The interface is created using netbeans editor. The query engine used in this case is ARQ query processor for Apache Jena. Apache Jena is a Semantic Web framework for java. The created ontology is then loaded into the application and a model of it created using Apache Jena. The interface enables users/ clients to send requests to the ontology through the query engine. The query engine processes the query, makes necessary queries to the ontology and returns results to the user through the interface. The user sends request by clicking a particular menu item on the interface.

The different requests sent by a user are processed differently and responses returned depending on the type of query being processed. Therefore, the type of response to a query depends on the menu item chosen by a user. For example, if user desires to get all the projects in the various counties, the query that will be sent to the query engine for processing is as shown in figure 4.9.3.1.

```
//query statement
String queryString =
                "PREFIX:
<http://www.knowledge_sharing.com/ontologies/knowledge_sharing
.owl#>"+
        "SELECT ?project " +
        "WHERE { ?p a :project ; :title ?project ; :specField
'' }  ";
// execute the queryString to obtain results
Query query = QueryFactory.create(queryString);
QueryExecution    qe    =    QueryExecutionFactory.create(query,
model);
 org.apache.jena.query.ResultSet results =  qe.execSelect();
    // Output queryString results
 ResultSetFormatter.out(System.out, results, query);
    qe.close();
```

Figure 4.7: sample query

4.10.2 Experimental Setup

To evaluate the approach's performance, a set of queries were prepared as shown in the tables 4.3, 4.4, and 4.5. Table 4.3 shows general evaluation queries according to specification fields. Table 4.4 shows specific queries. They represent search of documents using topics or titles. Table 4.5, on the other hand, contains specific queries that represent search by author names. The F-score or F-measure is one of the most commonly used measures in Natural Language Processing, Information Retrieval and Machine Learning applications. F-measure is a weighted harmonic mean of Precision and Recall. Recall, also known as sensitivity or True Positive Rate (TPR), is the frequency by which relevant documents are retrieved by a system. Precision, also known as True Positive Accuracy (TPA) or Positive Predictive Value (PPV), is a form of accuracy that refers to the frequency by which the retrieved documents are relevant (Marijan &Leskovar, 2015). The f-measure combines the two into a single measure used to show the accuracy of a system. The three measures are calculated as follows:

$$\text{Presision} = \frac{\text{True Positives } (t_p)}{\text{True Positives } (t_p) + \text{False Positives } (f_p)}$$

$$\text{Recall} = \frac{\text{True Positives } (t_p)}{\text{True Positives } (t_p) + \text{False Negatives } (f_n)}$$

$$\text{F-measure} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{precision}}$$

A true positive is the number of relevant documents retrieved, false positive is the number of irrelevant documents retrieved while false negatives is the number of relevant documents not retrieved.

Table 4.3: General evaluation queries --- All projects in Nairobi County

	General queries	SPARQL queries
Q1	All projects in Nairobi county	SELECT DISTINCT ?title ?county WHERE { ?p a :project ; :hasTitle ?t ; :hasCounty ?y. ?t :name ?title . ?y :name ?county. FILTER regex (?county, '(?=.*?(nairobi))', 'i') }group by ?title ?county
Q2	Supervisors of Nairobi county	SELECT DISTINCT ?title (group_concat(?supervisor; separator = ',') as ?supervisors) ?county WHERE { ?p a :project ; :hasTitle ?t; :hasSupervisor ?a ; :hasCounty ?y. ?t :name ?title . ?a :name ?supervisor. ?y :name ?county. FILTER regex (?county, '(?=.*?(nairobi))', 'i') }group by ?title ?supervisors ?county
Q3	All project financiers and addresses	SELECT DISTINCT ?name ?mailbox ?city ?zip ?email WHERE { ?f a :financier ; :name ?name ; :hasAddress ?r. ?r :mailbox ?mailbox . ?r :city ?city . ?r :zipcode ?zip . ?r :email ?email }group by ?name ?mailbox ?city ?zip ?email

Table 4.3 Query Clarification

Q1 gives all the projects in the field of Nairobi County. Q2 gives all supervisors in charge of the said projects in Nairobi County while Q3 gives all project financiers and their addresses.

Table 4.4 : Specific evaluation queries--- Search by project title

	General queries	SPARQL queries
Q1	Get financier of a specific project title	*SELECT DISTINCT ?title ?financier WHERE { ?p a :project ; :hasTitle ?t ; :hasFinancier ?j . ?t :name ?title . ?j :name ?financier. FILTER regex (?title, '(?=.*?(Lokichogio))(?=.*?(market))(?=.*?(shed))', 'i') }group by ?title ?financier*
Q2	Get projects with construction as a title	*SELECT DISTINCT ?title WHERE { ?p a :project ; :hasTitle ?t . ?t :name ?title . FILTER regex (?title, '(?=.*?(construction))', 'i') }group by ?title*

Table 4.4 query clarification

Q1 gives financier of specific project queried with its title. Q2 gives projects with "Construction "as a key word in its title.

Table 4.5: Specific evaluation queries- Search by project supervisor names

	General queries	SPARQL queries
Q1	query by one name	*SELECT DISTINCT ?title ?supervisors WHERE { ?p a :project ; :hasTitle ?t; :hasSupervisor ?a . ?t :name ?title . ?a :name ?supervisors . FILTER regex (?supervisors, '(?=.*?(loki))', 'i') }group by ?title ?supervisors*
Q2	Query by both names	*SELECT DISTINCT ?title ?supervisors WHERE { ?p a :project ; :hasTitle ?t; :hasSupervisor ?a . ?t :name ?title . ?a :name ?supervisors . FILTER regex (?supervisors, '(?=.*?(ambira))(?=.*?(josphat))', 'i') }group by ?title ?supervisors*

Table 4.5 Query clarification

The table contains queries that are supposed to retrieve supervisor names as a single name of by both names. *Q1* represents search using one name, 'Loki' while *Q2* represents search using both names 'Ambira josphat'.

4.11. Presentation of Results

Tables 4.6, 4.7 and 4.8 show the results of the evaluation queries found in tables 5.1, 5.2 and 5.3respectively. The expected number of documents to be retrieved is identified for each of the queries. The results retrieved in each query in table 5.5 & 5.6 depend on whether the discipline or category (eg telecommunication engineering, agriculture, computer science, etc) of documents has been specified or not. In order to calculate precision and recall, relevant documents have to be identified from the total number of documents retrieved. According to table 5.4, all the expected documents were retrieved. Table 5.2 contains queries that were expected to retrieve specific research materials. The approach's ability to retrieve the materials when the category of materials is specified and when it is not specified is tested. Table 5.5 shows the results of Table 5.2 queries. According to the table, not specifying the

category of documents widens the search scope thus increased number of relevant documents retrieved. Table 5.3, on the other hand, contains queries used to retrieve documents written by a specific author. Table 5.6 shows the results of table 5.3. According to the results, querying using a single name widens the search scope thus increasing the chances of both relevant and irrelevant documents being retrieved.

	Expected Number of documents	Retrieved number of documents	
		Total retrieved	Relevant
Q1	4	4	4
Q2	4	4	4
Q3	24	24	17

Table 4.6: Results of Table 4.3 Queries

	Expected number of documents	Retrieved number of documents	
		Total retrieved	Relevant
Q1	4	4	4
Q2	10	8	8

Table 4.7: Results of Table 4.4 Queries

	Expected number of documents	Retrieved number of documents	
		Total retrieved	Relevant
Q1	8	8	7
Q2	7	7	7

Table 4.8: Results of Table 4.5 Queries

4.12 Analysis of the Results

The tables 4.6, 4.7 and 4.8 show the analysis of the results in tables 4.3, 4.4 and 4.5 respectively. The analysis is done by calculating precision, recall and f-measure using the given formulas. F-measure is used to measure the accuracy of the approach. Table 4.6 shows that being specific in search of documents increases the chances of retrieving all the relevant documents while decreasing the chances of retrieving irrelevant documents. According to the results of table 4.6, the approach is precise and very accurate. This can be seen in the precision, recall and f-measure of all the queries in the table. The high precision and recall can be attributed to the specificity of the queries sent to the ontology. The queries were to retrieve documents from specific categories.

Unlike table 4.6 that shows evaluation results of general queries based on specific categories, tables 4.7 and 4.8 show evaluation results of specific queries to the ontology. Table 4.7 shows evaluation results of queries that retrieve items based on specific financier of specific projects or project titles. According to Q1 of table 4.7, Search by a specific title increases the precision and recall of the approach thus making it more accurate. However, when searching for documents on a specific topic, specifying the category of documents required narrows down the search scope thus decreasing recall and increasing precision. This is evident in Q2 of table 4.7 where the recall when the category of documents is not specified is higher than when the category is specified.

Table 4.8, on the other hand, shows evaluation results of queries that retrieve documents based on a specific supervisor of a project. According to Q1 of table 4.8, using a single name widens the search scope thus decreasing precision and increasing recall. A wider search scope increases the chances of irrelevant documents being retrieved thus the low precision. The wider search scope also increases the chances of relevant documents being retrieved thus the high recall. Using full names of authors increases both precision and recall thus increased accuracy as shown in Q2 of table 4.8

Table 4.9: Evaluation **Results for table 4.6**

Queries	Precision (%)	Recall (%)	F-measure
Q1	100	100	2
Q2	100	100	2
Q3	73.91	100	0.84

Table 4.10: Evaluation Results for table 4.7

Queries	Precision (%)	Recall (%)	F-measure
Q1	100	100	2
Q2	80	80	0.8

Table4.11Evaluation Results for table 4.8

Queries	Precision (%)	Recall (%)	F-measure
Q1	100	87.5	0.94
Q2	100	100	2

4.13. Discussion

To evaluate our approach, we design sets of queries one according to the evaluation model of (Matasyoh et al., 2016).The results obtained in the tables 4.6, 4.7 and 4.8 demonstrate that utilization of semantic techniques is rewarding since the approach produces high rates of Precision and Recall. Regarding Precision, the results in all the tables show that semantic search presents a higher probability of the retrieved documents being relevant. However, precision decreased in Q1 of table 4.8 owing to a wider search scope that resulted from use of a single name during search. The results show that the approach is Precise since most of the documents retrieved through the queries are relevant.

Recall, on the other hand, decreased in some instances such as in Q2 of table 4.7. This can be attributed to the queries being specific. Firstly, specifying the category of documents required, in Q2, limited the search scope to financiers of specific projects. Secondly, limiting the search scope to documents that contained the words "construction" in the title locked out other documents on the same topics that used other words such as "constructing". In return the two limitations decreased the chances of all relevant documents being retrieved. Though specificity reduces the probability of all relevant documents being retrieved, it increases the probability of only relevant documents being retrieved. This is clearly shown in the high precision and low recall rates depicted in table 4.7.

The accuracy of this approach is evaluated using f-measure. High precision and recall evaluated to high F-measure. The approach, therefore, proves to be more accurate since the evaluation results recorded high rates of f-measure. Generally, the approach is precise and accurate. The more specific a query is, the higher the probability of only retrieving relevant documents.

In general, the approach is effective since it provides a better solution to the problem of identification and retrieval of knowledge shared through World Wide Web. There is a lot of content shared and found on the web through various mechanisms but the main problem is the ease with which the content can be identified and retrieved. Various researchers have come up with different ways through which the knowledge can be easily shared but most did not consider the fact that knowledge is not complete without being effectively utilized by interested parties i.e there should be mechanisms through which the knowledge can be easily

identified and retrieved. Though some researchers considered mechanisms that enable easy identification of content (Semantic Web techniques), most have not evaluated their work and this leaves a gap of identifying how effective their ontologies are in knowledge sharing and retrieval. Most researchers stop at the building of the ontology. The mechanisms only applied to small groups of people who are working on specific projects or are in the same discipline. From the results, it is clear that application of Semantic Web in knowledge sharing is effective since it allows easy identification of relevant content by machines in order to give appropriate results to the user.

CHAPTER FIVE: CONCLUSION AND FUTURE WORK

5.1. Conclusion

The main objective of this research thesis sought to demonstrate the usefulness of structuring decentralized government data on the web so that it's readable and useful to both human and machine by proposing an architecture and prototype implementation. This main idea was broken down into three objectives namely: description of the various techniques used in digital knowledge sharing and retrieval, to design an ontology-driven approach for sharing and retrieval of knowledge and to evaluate the ontology-driven approach.

The first objective, which is describing digital knowledge sharing and retrieval techniques, was realized through detailed literature review. A detailed review of existing literature was conducted in order to identify the techniques that have been used in digital knowledge sharing and retrieval. From the review, two main techniques, web 2.0 and web 3.0 (Semantic Web), were found to have been used or proposed by most researchers. However, the most effective technique was Semantic Web which proved to solve digital knowledge retrieval problem. Apart from identifying knowledge sharing and retrieval techniques, literature review also helped in gaining more insights into Semantic Web which is the main technique used to implement the ontology driven approach.

The second objective, which was design of an ontology-driven approach for sharing and retrieval of knowledge, was carried out basing on the insights on semantic web gathered during literature review and the kinds of knowledge shared from the existing Kenyan government open data portal of various projects that have taken place or are yet to. A specific domain was chosen to cater which detailed government projects touching on the title of the project, its objective, commencement and end date, beneficiaries, financiers etc. There were two main stages involved in constructing the ontology. First, the ontology was designed by describing the knowledge domain using description logics. The different concepts of the domain and their relationships were described using language constructs provided by description logics language. After design, the ontology was developed using protégé ontology editor then loaded into a java application using Jena Framework. The ontology language used is OWL 2 that is supported by SROIQ DL. Description Logics, is important in enabling proper reasoning in the ontology. The ontology is developed separately and loaded into a java application that provides the interface through which the ontology can be queried. The query

engine receives queries from the interface, sends them to the ontology, receives feedback and then sends it back to the application interface. The ontology is populated after it has been loaded into the application.

The third objective, that seeks to evaluate the ontology driven-approach, is realized through performing tests on the approach then performing some calculations using f-measure, precision and recall. Our evaluation results for the retrieval task show comparable performance to the work by (Matasyoh et al., 2016).To be able to calculate precision and recall, relevant documents have to be identified from the total number of documents retrieved as per the earlier mentioned queries. The F-score or F-measure is the commonly used measures in Natural Language Processing, Information Retrieval and Machine Learning applications. It is a weighted harmonic mean of Precision and Recall e.g. all questions return a precision of 1, meaning that the SPARQL queries were able to retrieve only relevant documents.

5.2. Future Work

Currently, we restrict querying of the database to selection boxes via the developed user interface. In essence, this limits the expressiveness of the resulting queries. Natural language processing approach can be an extension to this and linked to the ontology database to improve expressivity of queries. For instance, with NLP, a sentence such as "projects in Nairobi county with title construction can be used to construct a query. The example above would give results as instances of "projects" with specification field "construction" that are running within the county of "Nairobi". Again, semantic techniques combined with natural language processing will also help automate the search and make it free of human intervention.

GENERATING STRUCTURED DATA FROM GOVERNMENT OPEN DATA USING RESOURCE DESCRIPTION FRAMEWORK FOR KNOWLEDGE

ORIGINALITY REPORT

15%	10%	10%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	acikarsiv.atilim.edu.tr Internet Source	<1%
2	"Reasoning Web. Semantic Technologies for the Web of Data", Springer Nature, 2011 Publication	<1%
3	www.iiste.org Internet Source	<1%
4	Submitted to Higher Education Commission Pakistan Student Paper	<1%
5	Submitted to The University of Manchester Student Paper	<1%
6	Adebayo, Adekoya, Akinwale Adio, and Sofoluwe Adetokunbo. "A Conceptual Framework for an Ontology-Based Examination System", International Journal of Advanced Computer Science and Applications, 2011.	<1%

	Publication	
7	www.ibiblio.org Internet Source	<1%
8	www.cs.uga.edu Internet Source	<1%
9	Submitted to University of Westminster Student Paper	<1%
10	Journal of Knowledge Management, Volume 17, Issue 3 (2013-06-08) Publication	<1%
11	Kuck, G.. "Tim Berners-Lee's Semantic Web", SA Journal of Information Management, 2009. Publication	<1%
12	www.semicolon.no Internet Source	<1%
13	Submitted to Università degli Studi di Ferrara Student Paper	<1%
14	www.osti.gov Internet Source	<1%
15	Clement Yu. "Annotating Structured Data of the Deep Web", 2007 IEEE 23rd International Conference on Data Engineering, 04/2007 Publication	<1%
16	Attard, Judie, Fabrizio Orlandi, and Soren Auer. "Value Creation on Open Government	<1%

	Data", 2016 49th Hawaii International Conference on System Sciences (HICSS), 2016. Publication	
17	www.hwswworld.com Internet Source	<1%
18	www.questia.com Internet Source	<1%
19	www.ilrt.bris.ac.uk Internet Source	<1%
20	events.linkeddata.org Internet Source	<1%
21	Submitted to Greenwich School of Management Student Paper	<1%
22	Zidi, Amir, and Mourad Abed. "Towards a framework for ontology-based information retrieval services", International Journal of Services and Operations Management, 2014. Publication	<1%
23	Yu, Liyang. "RDFS and Ontology", A Developer's Guide to the Semantic Web, 2014. Publication	<1%
24	rassweb.org Internet Source	<1%

25	www.istp.org.in Internet Source	<1%
26	repositorio-aberto.up.pt Internet Source	<1%
27	Submitted to Al-Madinah International University (MEDIU) Student Paper	<1%
28	Pesce, Marcia Lucas, Karin K. Breitman, and Marco Antonio Casanova. "Surfacing scientific and financial data with the Xcel2RDF plug-in", 2012 Second International Workshop on Developing Tools as Plug-Ins (TOPI), 2012. Publication	<1%
29	thegovlab.org Internet Source	<1%
30	deri.ie Internet Source	<1%
31	Yang, S.J.H.. "A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network", International Journal of Human - Computer Studies, 200801 Publication	<1%
32	kemi.ac.ke Internet Source	<1%

33	Hermann Kaindl, Elmar P. Wach, Ada Okoli, Roman Popp, Ralph Hoch, Werner Gaulke, Tim Hussein. "Semi-automatic generation of recommendation processes and their GUIs", Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13, 2013 Publication	<1%
34	Yu, Liyang. "RDFS, Taxonomy, and Ontology", Introduction to the Semantic Web and Semantic Web Services, 2007. Publication	<1%
35	Submitted to University of East London Student Paper	<1%
36	Handbook of Semantic Web Technologies, 2011. Publication	<1%
37	www.information-online.com Internet Source	<1%
38	eprints.soton.ac.uk Internet Source	<1%
39	Submitted to Bournemouth & Poole College of Further Education Student Paper	<1%
40	Xiangji Huang. "Knowledge Retrieval (KR)", IEEE/WIC/ACM International Conference on	<1%

	Web Intelligence (WI 07), 11/2007 Publication	
41	medinform.jmir.org Internet Source	<1%
42	Submitted to University of Leicester Student Paper	<1%
43	www.unisonpartners.net Internet Source	<1%
44	Submitted to University of Duhok Student Paper	<1%
45	Submitted to KDU College Sdn Bhd Student Paper	<1%
46	www.ihub.co.ke Internet Source	<1%
47	Communications in Computer and Information Science, 2011. Publication	<1%
48	Fareedi, Abid Ali, and Syed Hassan. "The impact of social media networks on healthcare process knowledge management (using of semantic web platforms)", 2014 14th International Conference on Control Automation and Systems (ICCAS 2014), 2014. Publication	<1%
49	www.aimia.com.au	

	Internet Source	<1%
50	ir.inflibnet.ac.in Internet Source	<1%
51	users.encs.concordia.ca Internet Source	<1%
52	www.practicalecommerce.com Internet Source	<1%
53	Submitted to University of North Texas Student Paper	<1%
54	www.websurfingguide.com Internet Source	<1%
55	www.jcheminf.com Internet Source	<1%
56	citeseerx.ist.psu.edu Internet Source	<1%
57	Submitted to University of Pretoria Student Paper	<1%
58	Submitted to Staffordshire University Student Paper	<1%
59	Benny, S. Prabhakar, S. Vasavi, and P. Anupriya. "Hadoop Framework For Entity Resolution Within High Velocity Streams", Procedia Computer Science, 2016.	<1%

	Publication	
60	"Semantic Web Challenges", Springer Nature, 2016 Publication	<1%
61	web.uniroma1.it Internet Source	<1%
62	Katrin Weller. "Frontmatter", Walter de Gruyter GmbH, 2010 Publication	<1%
63	Kornkanok Phoksawat, Massudi Mahmuddin. "Hybrid Ontology-based knowledge with multi-objective optimization model framework for Decision Support System in intercropping", Advances in Science, Technology and Engineering Systems Journal, 2017 Publication	<1%
64	www.politesi.polimi.it Internet Source	<1%
65	www.knoesis.org Internet Source	<1%
66	vidont.org Internet Source	<1%
67	docs.marklogic.com Internet Source	<1%
68	www.eifl.org	

	Internet Source	<1%
69	Submitted to Coventry University Student Paper	<1%
70	www.slideshare.net Internet Source	<1%
71	Submitted to University of South Africa Student Paper	<1%
72	Dragoni, M.. "A conceptual representation of documents and queries for information retrieval systems by using light ontologies", Expert Systems With Applications, 20120915 Publication	<1%
73	Farag Ahmed. "Evaluation of n-gram conflation approaches for Arabic text retrieval", Journal of the American Society for Information Science and Technology, 2009 Publication	<1%
74	ses.library.usyd.edu.au Internet Source	<1%
75	Akerkar, Rajendra. "Ontology : Fundamentals and Languages", Applied Semantic Web Technologies, 2011. Publication	<1%
76	Danjuma, Sani; Jun Zhou; Hongyan Mei; Aliyu,	

	Ahamd and Waziri, Usman. "Design, Analysis and Implementation of Semantic Web Applications", International Journal of Computer Science Issues (IJCSI), 2013. Publication	<1%
77	Submitted to iGroup Student Paper	<1%
78	Submitted to University of Ulster Student Paper	<1%
79	Hogan, Aidan, Marcelo Arenas, Alejandro Mallea, and Axel Polleres. "Everything you always wanted to know about blank nodes", Web Semantics Science Services and Agents on the World Wide Web, 2014. Publication	<1%
80	Submitted to Macquarie University Student Paper	<1%
81	Encyclopedia of Social Network Analysis and Mining, 2014. Publication	<1%
82	en.wikipedia.org Internet Source	<1%
83	eli.elc.edu.sa Internet Source	<1%
84	Submitted to University of Hull Student Paper	<1%

85	answers.semanticweb.com Internet Source	<1%
86	Calegari, S.. "Granular computing applied to ontologies", International Journal of Approximate Reasoning, 201003 Publication	<1%
87	www.nesc.ac.uk Internet Source	<1%
88	Submitted to 90152 Student Paper	<1%
89	Submitted to Midlands State University Student Paper	<1%
90	docnum.univ-lorraine.fr Internet Source	<1%
91	videolectures.net Internet Source	<1%
92	docplayer.org Internet Source	<1%
93	mehmetalierturk.com Internet Source	<1%
94	krazytech.com Internet Source	<1%
95	Sajja, . "Web intelligence", Chapman & Hall/CRC Data Mining and Knowledge	<1%

	Discovery Series, 2012. Publication	
96	www.mexabet.biz Internet Source	<1%
97	jhaycee9811.wordpress.com Internet Source	<1%
98	Lecture Notes in Computer Science, 2012. Publication	<1%
99	www.ghxiao.org Internet Source	<1%
100	uclab.khu.ac.kr Internet Source	<1%
101	hsss.slub-dresden.de Internet Source	<1%
102	acva2010.cs.drexel.edu Internet Source	<1%
103	researcharchive.vuw.ac.nz Internet Source	<1%
104	Ghaleb, Fayed F. M., Sameh S. Daoud, Ahmad M. Hasna, Jihad M. Jaam, and Hosam F. El-Sofany. "A Web-Based E-Learning System Using Semantic Web Framework", Journal of Computer Science, 2006. Publication	<1%

105	www.billganz.com Internet Source	<1%
106	Xie, Bin, Denghui Zhang, Le Yu, Dengrong Zhang, Jianya Gong, and Huayi Wu. "", International Conference on Earth Observation Data Processing and Analysis (ICEODPA), 2008. Publication	<1%
107	www.901am.com Internet Source	<1%
108	www.aifb.uni-karlsruhe.de Internet Source	<1%
109	Mauricio Barcellos Almeida. "Ontologies in knowledge management support: A case study", Journal of the American Society for Information Science and Technology, 10/2009 Publication	<1%
110	Hart, . "Appendix A", Linked Data A Geographic Perspective, 2013. Publication	<1%
111	www.w3.org Internet Source	<1%
112	Nie, Ping, Yanhui Jiang, and Zhangxi Lin. "Hierarchy structure of XBRL and financial information data mining", The 2nd International	<1%

	Conference on Information Science and Engineering, 2010. Publication	
113	Lecture Notes in Computer Science, 2013. Publication	<1%
114	researchspace.ukzn.ac.za Internet Source	<1%
115	Lecture Notes in Computer Science, 2008. Publication	<1%
116	link.springer.com Internet Source	<1%
117	Lecture Notes in Computer Science, 2014. Publication	<1%
118	Samsuzzaman, M.; Islam, M. T.; Rashid, Sharmin; Ahmed, Faysal and Khan, Ridgewan. "Proposed Model Of E-Learning Management System Using Semantic Web", Journal of Applied Sciences Research, 2012. Publication	<1%
119	www.mindswap.org Internet Source	<1%
120	Submitted to The Hong Kong Polytechnic University Student Paper	<1%
121	Goutam Kumar Saha. "Web ontology language	

	(OWL) and semantic web", Ubiquity, 2007 Publication	<1%
122	mro.massey.ac.nz Internet Source	<1%
123	D LEE. "Data Mining", Database and Data Communication Network Systems, 2002 Publication	<1%
124	wellsdigest.com Internet Source	<1%
125	seal.ifi.unizh.ch Internet Source	<1%
126	start.org Internet Source	<1%
127	Casado-Lumbreras, C.. "PsyDis: Towards a diagnosis support system for psychological disorders", Expert Systems With Applications, 20121001 Publication	<1%
128	Letia, Ioan Alfred, and Anca Goron. "Model checking as support for inspecting compliance to rules in flexible processes", Journal of Visual Languages & Computing, 2015. Publication	<1%
129	World Journal of Science, Technology and Sustainable Development, Volume 7, Issue 1	<1%

(2012-08-06)
Publication

130	Lecture Notes in Computer Science, 2015. Publication	<1%
131	Lecture Notes in Computer Science, 2005. Publication	<1%
132	Semantic Web Information Management, 2010. Publication	<1%
133	Merelli, Ivan Perez-Sanchez, Horacio Ges. "Managing, analysing, and integrating big data in medical bioinformatics: open problems and future pe", BioMed Research International, Annual 2014 Issue Publication	<1%
134	www.ifi.unizh.ch Internet Source	<1%
135	web.it.kth.se Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	Off		