Adaptive Search and Selection of Domain Ontologies for Reuse on the Semantic Web

Jean Vincent Fonou-Dombeu
Department of Software Studies, Vaal University of Technology, South Africa
Email: fonoudombeu@gmail.com

Magda Huisman
School of Computer, Statistical and Mathematical Sciences, North-West University, South Africa
Email: Magda.Huisman@nwu.ac.za

Abstract—**Ontology plays an important role in Semantic Web applications. However, building ontology remains challenging due to the time, cost an effort required. Several studies have proposed the reuse of existing ontologies when building new ones. However, some challenges remain: (1) locating relevant domain ontologies for reuse, (2) determining appropriate concepts for searching targeted ontologies and (3) understanding the discovered ontologies. This study presents an adaptive strategy for searching and selecting domain ontologies for reuse on the Semantic Web. The strategy relies on ontology-based and generic search engines, and predefined ontology features to locate existing domain ontologies and related data sources. The data sources provide ontologies' specific concepts that enable their easy location over the Semantic Web. Finally, a set of criteria including semantic coverage, codification language, modularity and open availability are used to select the best reusable set of ontologies for the domain. The application of the framework in the e-government domain demonstrated its feasibility and yielded promising results.**

Index Terms - Semantic Web, Ontology Search, Ontology Selection, Ontology Reuse, E-government.

I. INTRODUCTION

The Semantic Web is an evolution of the current web that provides meaning to web contents to enable their intelligent processing by computers. The meaning of web contents is represented with ontology and described formally in logic-based syntaxes to facilitate their integration and interoperability. As such, ontology is a key component of any semantic web application. Ontology is commonly defined as an explicit specification of a conceptualization [1] i.e., a model of the real world domain such as medicine, geographic information systems, physics, e-government and so forth; which is explicitly represented with existing objects, concepts, entities and relationships between them.

Building ontology in Semantic Web remains a challenging task due to the demand in time, cost an effort. The solution lies in the reuse of existing domain ontologies when building new ones [2][3][4][5][6][7]. In fact, ontology reuse may (1) reduce human efforts required to formalized new ontologies from scratch, (2) increase the quality of the resulting ontologies because the reused ontologies have already been tested, (3) simplify the mapping between ontologies built using shared components of existing ontologies, and (4) improve the effi-

ciency of ontology maintenance [5].

However, existing domain ontologies are spread over the Internet and presented in different media including Semantic Web ontology files such as Resource Description Framework (RDF) and Web Ontology Language (OWL), text files (related research/project reports/published articles, program generated codes, etc.), Web pages, etc. Furthermore, ontology search engines enable the retrieval of ontology files based on keywords search; this presents some challenges: searching ontologies by keywords requires one to provide keywords that are likely to match those in the ontology files available in the indexes of the search engines [8]; but it is difficult to guess keywords of unknown domain ontologies; even if the domain ontology is known, it remains challenging to accurately guess keywords that are included in this ontology over the Internet. Moreover, semi-automatic and automatic ontology reuse solutions largely rely on ontology search engines for locating existing domain ontologies over the Semantic Web; consequently, they only focus on ontology files stored in the indexes of the search engines [4][9]; other data sources of existing ontologies such as related research/project reports, published articles, programme codes, Web page contents are left out. This results in many useful domain ontologies and information sources, including that of located ontologies, being ignored in these ontology reuse solutions; consequently, these solutions are directed towards experienced ontology engineers who are able to understand the located domain ontology files (RDF/OWL for example) to guide the process for building new ontologies.

The aforementioned challenges hinder the widespread reuse of existing domain ontologies and undermine the adoption of Semantic Web technologies in the respective domains. This study presents a framework for searching and selecting domain ontologies for reuse on the Semantic Web. The proposed framework may be applied in any application domains of Semantic Web such as e-commerce, e-business, e-learning, multimedia, e-government, etc., to identify and analyze existing domain ontologies for the purpose of knowledge sharing and reuse across domain specific Semantic Web applications. The framework uses an adaptive strategy that relies on ontology-based and generic search engines, and predetermined ontology features to locate existing domain ontologies and

related data sources.The result is a list of candidate domain ontologies along with sets of data sources. The data sources of an ontology may include semi-structured and unstructured data such as research and project deliverable reports, related published articles, ontology codes, plain texts on project web sites , ontologies repositories, etc. These data sources disclose valuable information that may support the widespread reuse and evolution of corresponding domain ontologies. Examples of such information are: (1) the purpose(s) for which the ontology was built, (2) the methodology employed to build the ontology, (3) the full or partial ontology graph(s), (4) theoretical explanation of the meaning of concepts and axioms, (5) full or partial code of the ontology, and (6) detailed description of the use of the ontology in real world semantic-based projects, etc. [10][11][12][13]. This information is certainly valuable for any reuse tasks, including the automatic or semi-automatic ontology reuse which requires the ontology engineer to have prior knowledge of existing domain ontology to be able to comprehend and guide the process for building new ontologies through the reuse of existing ones [3][4][5][7]. More importantly, the collected ontologies' data sources provide Semantic Web developers with specific concepts of the targeted ontologies to enable their easy location over the Semantic Web; furthermore the data sources provide useful information for analyzing, understanding and reusing the existing domain ontologies.

Finally, various metrics including semantic coverage, open availability, codification language, and modularity are applied on the set of located candidate domain ontologies to evaluate and select the best reusable set of ontologies for the respective domain. The selected ontologies provide a good sharable and reusable conceptual representation and description of the domain. This may (1) promote their reuse across domain specific Semantic Web projects, (2) save the time and cost needed for building new ontologies from scratch in domain specific Semantic Web projects, (3) prevent inconsistency and confusion that may arise from multiple semantic representations of the same domain knowledge, and (4) strengthen the harmonization and adoption of Semantic Web technologies in the respective domain.

The proposed framework is simple and suitable to any Semantic Web developer who may like to search and locate existing domain ontologies on the Semantic Web, analyze, understand and reuse these ontologies in the process of building new ontologies either manually or with semi-automatic or automatic ontology reuse solutions [3][4][5][7]; this may promote the widespread reuse of existing domain ontology on the Semantic Web. The application of the framework in the e-government domain demonstrated its feasibility and yielded promising results.

The rest of the paper is structured as follows. Section 2 provides a formal specification of the framework of the search and selection strategy. The results of the application of the framework to the e-government domain is presented and discussed in Section 3. Section 4 discusses related studies and the last section concludes the paper.

II. FRAMEWORK OF THE SEARCH AND SELECTION STRATEGY

Let's D be a domain of knowledge such as e-commerce, e-business, e-government, etc. The aim is to investigate available data sources on semantic web initiatives (real world projects, academic research works, etc.) in that particular domain. The sources of information may include technical research / deliverable reports, published articles, programming codes, data repositories, plaintext pasted on websites, etc.

Let's L_D be the set of identified semantic based data sources gathered in the domain D. L_D is defined as in Equation (1).

$$L_D = A \cup P \qquad (1)$$

where, A is the set of data sources that are related to research carried out for academic purposes and P is the set of sources that are related to business projects for building semantic web applications. A and P are defined as in Equation (2) and (3).

$$A = \{a_i\}, 1 \le i \le N \qquad (2)$$

$$P = \{p_j\}, 1 \le j \le M \qquad (3)$$

where, N and M are the cardinalities of A and P respectively.

Let's l_{Dk} be the list of domain keywords to be used for the search of data sources, l_{Op} the list of ontology features required to guess the presence of any ontology activities in the data sources, and l_{Oc} the list of ontology specific concepts identified in data sources. Moreover, let's L_C and C be the list of data sources susceptible to content domain ontologies and the set of candidate domain ontologies respectively. C is defined as in Equation (4).

$$C = \{O_k\}, 1 \le k \le n \qquad (4)$$

where, O_k is the candidate ontology number k; it is assumed that there are up to a number n candidate ontologies in the domain.

Finally, let's C_r be the set of predefined criteria for selecting an ontology for the domain D and d_o the final set of selected ontologies. To fulfill the goal of the framework which is to search and select domain ontologies in the domain D, the following tasks are manually or semi-automatically performed: online search, group data sources, analyze data sources, online specific search, find candidate domain ontology, and select domain ontology. A brief definition of each of these tasks is provided below.

- **Online Search** - This task uses ontology search engines such as Swoogle, Watson, OntoSearch, OntoSearch2, OntoKhoj [9], etc. and generic search engines such as Google, Google Schoolar, IEEE Explore, ISI Web of Knowledge, etc. to gather diverse data sources on existing semantic-based research and projects, based on the list of domain keywords in l_{Dk}. The result is the set L_D of all identified semantic-based data sources.
- **Group Data Sources** - The set L_D of all data sources is used in this task; evidences of relatedness are searched

in the data sources; this enable to group the data sources. Two data sources are related if they were produced under the same project or study. At this stage, the result is a collection of folders containing data sources related to the same semantic-based academic research (the set A) or real world semantic-based project (the set P).

- **Analyze Data Sources** - This task uses the set of targeted ontology features l_{Op} and either an element of A (a folder containing data sources related to an academic research project) or an element of P (a folder holding data sources pertaining to a real world semantic-based project). Ontology features include ontology graphs and concepts of the semantic web ontology languages such as RDF/RDFS and OWL. These concepts may have been used in a semi-formal definition of an ontology (simple definition of concepts and relationships in the form of texts), in the graphical representation of an ontology or within different axioms representing a formal ontology (machine generated codes). OWL constructs targeted could include Class, SubClassOf, Equivalent-Class, DisjointWith, ObjectProperty, Property, Domain, Range, etc., whereas, RDF constructs could encompass Class, SubClass, SubProperty, Domain, Range, Object, Predicate, Type, Literal, etc. The result of this task is a list of ontology concepts l_{Oc}. l_{Oc} may be empty or not, depending on whether the targeted ontologies features in l_{Op} where found or not.

- **Online Specific Search** - The set of specific ontology concepts l_{Oc} obtained with the previous task is used in this task to perform further search with ontology search engines; aiming at finding the codes of the targeted ontologies. The results of the search are used to update the sets $a_i \subset A$ or $p_j \subset P$ of semantic-based data sources.

- **Find Candidate Domain Ontology** - This task consists of scrutinizing each data source $a_i \subset A$ or $p_j \subset P$ where ontology features were found to identified candidate domain ontology. The result is a candidate ontology O_k. O_k is added to the set of candidate ontologies C.

- **Select Domain Ontology** - A candidate ontology $O_k \subset C$ and the set of predefined criteria C_r for selecting domain ontologies in the domain D are used in this task. Further analysis of the ontology $O_k \subset C$ data sources is then performed to tell whether the candidate ontology $O_k \subset C$ meet the selection criteria. Based on the works in [14] and [15], it is suggested that the elements of the set C_r of predefined criteria for selecting a domain ontology $O_k \subset C$ be: codification language, semantic coverage, modularity and open availability. These criteria are defined below.

  - **Codification Language** This characteristic refers to the language employed for the formal representation of the ontology. In fact, it is expected that the codification language of a selected ontology be one of the standard ontology languages for the Semantic Web,

such as Resource Description Framework (RDF) or Web Ontology Language (OWL).
  - **Semantic Coverage** The value of this characteristic is low, medium or high, thereby indicating the level of semantic richness of the ontology; the semantic richness is assessed based on the ontology features such as the number of concepts, supsumption (is.a), meronymy (part-of), etc.; in brief, a selected ontology should not be built as a simple taxonomy, it must further be formed of rich semantic features.
  - **Modularity** This characteristic tells whether the ontology is formed of a single or many components. An ontology with several modules enables: (1) easy reuse of smaller parts, (2) distributed and collaborative development, (3) smooth and efficient evolution, and (4) easy replacement of parts of the ontology [16].
  - **Open Availability** Here, it is shown whether the ontology is publicly available or not. The accessibility of the selected ontologies to the public is of prime importance as the major aim of the study is to foster the reuse of the selected domain ontologies in Semantic Web projects in the domain D.

In light of the above, the pseudo-code of the framework's algorithm is drawn in Table 1. In the next section, the framework described above and formalized in the algorithm in Table 1 is applied on the e-government domain.

III. APPLICATION IN E-GOVERNMENT

A. Online Search of Domain Ontologies

First of all, it became necessary to investigate and choose amongst existing ontology search engines those that are suitable for the task at hand. The researchers benefited from the work in [9]. In fact, in [9] a detailed comparative analysis of the commonly used [9] semantic web search engines including Swoogle, Watson, Sindice, Falcons and Semantic Web Search Engine; the study revealed that Swoogle and Watson are the state-of-the-art of all ontology search engines. Consequently, the Swoogle and Watson ontology search engines were adopted in this study. Thereafter, the following e-government domain keywords were chosen to perform the search in Swoogle and Watson search engines: government, citizen, service, business, tax, procurement, law, department, agency, civil servant, and life event.

These keywords were not exhaustive, but the aim was to perform the search and appreciate the nature of the results obtained. Furthermore, the abovementioned keywords were grouped into triplets as in Fig. 1 with the aim of improving the quality of the search results [9].

Although Swoogle and Watson search engines could return hits on OWL and RDF ontology files, some general problems surfaced. Firstly, searching ontologies by keywords requires one to provide keywords that are likely to match those in the ontology codes available in the indexes of the search engines [8]; but it is difficult to guess keywords of unknown domain

TABLE I
PSEUDO-CODE OF THE ONTOLOGY SEARCH AND SELECTION ALGORITHM

$Inputs: D; l_{Dk}; l_{Op}; C_r$
1. $L_D = Online\ search\ with\ domain\ keywords\ in\ l_{Dk}$
2. $A = Group\ academic - based\ data\ sources\ from\ L_D$
3. $P = Group\ real\ world\ projects\ data\ sources\ from\ L_D$
4. **For** All academic research a_i in A
5. $l_{Oc} = Analyse\ a_i\ data\ sources\ with\ the\ ontolgy\ features\ in\ l_{Op}$
6. **If** specific ontology concepts were found i.e. l_{Oc} isn't empty **Then**
7. $a_i = update\ a_i\ data\ sources\ with\ a\ specific\ online\ search\ with\ l_{Oc}$
8. **EndIf**
9. $O_k = analyse\ a_i\ data\ sources\ to\ identify\ corresponding\ domain\ ontology$
10. $C = update\ the\ set\ of\ domain\ ontologies\ C\ with\ the\ new\ ontology\ O_k$
11. **EndFor**
12. **For** All project p_j in P
13. $l_{Oc} = Analyse\ p_j\ data\ sources\ with\ the\ ontolgy\ features\ in\ l_{Op}$
14. **If** specific ontology concepts were found i.e. l_{Oc} isn't empty **Then**
15. $p_j = update\ p_j\ data\ sources\ with\ a\ specific\ online\ search\ with\ l_{Oc}$
16. **EndIf**
17. $O_k = analyse\ p_j\ data\ sources\ to\ identify\ corresponding\ domain\ ontology$
18. $C = update\ the\ set\ of\ domain\ ontologies\ C\ with\ the\ new\ ontology\ O_k$
19. **EndFor**
20. **For** All candidate domain ontologies O_k in C
21. $Use\ selection\ criteria\ in\ C_r\ to\ analyse\ O_k\ data\ source$
22. **If** O_k matches the selection criteria in C_r **Then**
23. $d_o = update\ the\ set\ d_o\ of\ selected\ domain\ ontologies\ with\ O_k$
24. **EndIf**
25. **EndFor**
$Output: d_o$

ontologies; even if the domain ontology is known, it remains challenging to accurately guess keywords that are included in this ontology over the Internet. Secondly, the number of hits returned for certain keywords entered in the search engines was high; then, it becomes impractical to click and visually assess each hit; furthermore, a large number of hits returned were not related to useful ontologies for the domain [9]. Finally, the ontology codes downloaded from the search did not provide enough information on the target ontologies; in general only concepts of the ontologies and their semantic structures (axioms) are provided in these codes. Although the Watson search engine could provide some Meta data such as the size of the ontology, its number of statements, classes, properties, individuals, etc. little information was provided in these ontology codes on the discovered ontologies such as the purposes and circumstances for which they were built, the available documentation such as the deliverable reports of projects in which they were built, the related published articles, etc. This information may provide important insights for analysing and reusing these ontologies. In fact, a good documentation on an existing ontology would certainly ease its reuse and evolution. In light of the above mentioned challenges, it becomes necessary to complement the results of the ontology search engines (Swoogle and Watson) with that of robust and generic search engines. To this end, a generic search was carried out in several search engines including ISI Web of Knowledge, IEEE Explore, Google Scholar and Google. The keywords employed were "e-government ontology" and "semantic e-government". These generic searches produced 202 e-government domain semantic-based documents presenting ontology codes, semantic-based published articles, research

and projects' deliverable reports, and ontology repositories. These ontologies' data sources are grouped in the next subsection.

B. Group Data Sources

It was discovered that several documents downloaded with the generic searches were related to the same semantic-based projects or study. Then, a strategy based on the analysis of their contents was used to group related documents. To this end, each downloaded document was searched for the acknowledgement section. In fact, where found, the acknowledgement section provided information on the project or study in which the research was undertaken. Furthermore, the deliverable reports of various e-government projects, mainly European based projects, were scrutinized to discover more semantic-based e-government projects. As a result, all the documents downloaded were grouped into 21 folders, corresponding to 19 e-government projects and several academic studies. The analysis of the discovered ontology data sources is explained in the next subsection.

C. Analyse Data Sources

The semantic-based researches and projects documents downloaded in the previous task were further scrutinized to identify the projects and research studies which have employed ontology to address a particular aspect of e-government services delivery. This was done by checking ontology features in these documents. Let's recall that ontology features include ontology graphs and concepts of the semantic web ontology languages such as RDF/RDFS and OWL. These concepts may have been used in a semi-formal definition of an ontology (simple definition of concepts and relationships

Fig. 1.   Triplets of Domain Keywords Employed for E-government Domain Ontologies Search

in the form of texts), in the graphical representation of an ontology or within different axioms representing a formal ontology (machine generated codes). The identified candidate ontologies were recorded along with their authors, date of publication and where applicable, the project in which they were developed. Out of the 19 semantic-based projects initially identified, 12 projects remained (See Table 2 and Table 3); the related published papers and reports provided enough evidence (conceptual part of domain ontology, informal description of domain ontology, and/or sample code of ontology) of ontology development in these projects. The next subsection performs a specific search using the specific ontology concepts discovered in the data sources.

D. Search Specific Ontology Codes

The ontology features discovered within the ontologies data sources in the previous task provided in some cases, specific concepts of the candidate ontologies. At this stage, some of these concepts were used in Swoogle and Watson ontology search engines to attempt to retrieve the full codes of these ontologies. Fig. 2 depicts the concept $lkif-core$ obtained from the data sources on the FEA-RMO ontology along with the OWL files of 4 FEA-RMO modules retrieved with the search in Swoogle; the URLs in Fig. 2 disclose that the ontology modules were developed under the Estrella e-government project. Furthermore, Table 5 shows selected e-government domain ontologies along with the Web links to their full OWL codes, retrieved from the Web with specific keywords search. Ontology selection is done in the next subsection.

E. Select E-government Domain Ontologies

With the list of candidate e-government domain ontologies in Table 2 and Table 3, their data sources including eventual full codes, predefined criteria such as codification language, semantic coverage, modularity and open availability [14][15] are applied to select the best set of ontologies for the e-government domain as in Table 4. The next subsection presents and discusses the complete results of the application of the framework in Section 2 in the e-government domain.

F. Results and Discussions

Table 2 and Table 3 list 62 discovered candidate e-government domain ontologies along with selected data sources on these ontologies as well as the e-government research and projects in which they were developed. This provide any e-government developer interested in reusing these existing domain ontologies with relevant information for analyzing, understanding and reusing these domain ontologies for building new ontologies, even with existing automatic and semi-automatic ontologies reuse solutions [3][4][5][7] that required ontology engineers to guide the process.

Further, Table 2 and Table 3 shows that most of e-government projects employ several domain ontologies for the Semantic Web development of e-government systems. Moreover, one can notice in Table 2 that some candidate ontologies are being repeated in different projects with the same name to serve the same purposes; for instance, the life-event ontology have been developed in 6 projects and the service ontology in 3 projects; this shows a lack of ontology reuse culture in the Semantic Web e-government development community.

Table 4 presents the candidate ontologies that were selected as the best set of ontologies for the e-government domain, based on their codification language, semantic coverage, modularity, and open availability as defined in the Section 2. A brief presentation of these selected e-government domain ontologies obtained from their data sources is provided below.

The selected e-government domain ontologies in Table 4 were developed within real world e-government projects in the United States [10], European countries [12][11][17][13], and Palestine [16]. This indicates that these ontologies have been well thought of, consistently designed and published. In particular:

- The LKIF-core ontology [12] describes the law and regulations that government the public administration domain through basic legal concepts; it is formed of 150 concepts and built with intensive semantic features (hyponymy, supsumption, etc.).
- The government ontology [16] is composed of 15 modules describing public administration entities such as address, bank, local government unit, natural and non-natural person, company, partnership company, shareholder company, driving licence, etc.; these set of ontologies model processes and enable systems interoperability

Fig. 2.   Screenshot Showing how the LKIF-core Concept was used to Retrieve the Modules of the LKIF-core ontology from Swoogle

in e-government.

- The FEA-RMO [10] ontology is a set of 5 modules namely performance, business, services, technology and data reference models ontologies; these ontologies were developed to enable the interoperability of the US government's federal agencies; they basically provide common reference models for modelling federal agencies' business processes, thereby, supporting their interoperability.

- The SAKE ontology [11] is formed of 3 modules including: process and profile, information, and decision making quality ontologies; these ontologies were developed as support to an agile knowledge management system for e-government. In particular, the process and profile ontology models the business process and related activities that might involve a public administration user; it is formed of 47 concepts including input, output, date, creation-date, last-modification-date, process-model, and so forth and fully represented in an is-a hierarchy. The information ontology describes metadata such as subject, description, title, creator, publisher, format, location, and the like; overall, it contents 33 concepts describing storable information; these concepts were designed after a meticulous analysis of existing metadata standards and their harmonization. The decision making quality ontology models concepts that might be used as performance evaluation parameters of a process in a public administration organization; these concepts are in total 33 and include: metric, accountability, cost, quality, and many more.

- The GEA ontology [17][18] is a single abstract model that describes the public administration semantic as well as the overall e-government domain; it includes concepts such as governance-entity, political-entity, admin-level, service-provider, public-administration-service, law, outcome, and so forth. It is also used to enable the auto-

matically mapping of citizens' needs to suitable public services.

- The life-event-ontology [13] is a single generic ontology model as well; it models the public administration services with 18 concepts related to life-events (e.g., get married, change address) of citizens with the public administration systems; these concepts include: public-service, input, output, profile, document, citizen, family-status, education-level, job-category, gender, and the like.

In light of the above, the selected e-government domain ontologies in Table 4 are largely formed of several modules that are publicly available; this may promote their reuse and evolution in the Semantic Web e-government development community [16].

Tables 5 provides the Web links to chosen data sources of the selected e-government domain ontologies in Table 4; these Web links are directed to either the ontology codes, deliverable reports or published research articles from projects in which these domain ontologies were developed. It is worth mentioning that in some cases, the ontology codes were not found with a keywords search in Swoogle and Watson search engines; instead, the full codes of some of the domain ontologies discovered were found in deliverable reports of corresponding projects with generic search engines; this shows the effectiveness of the adaptive search strategy presented in this study for locating domain ontologies and their data sources on the Semantic Web.

Furthermore, the deliverable and research reports of projects provided valuable information on the identified ontologies such as: (1) the purpose(s) for which the ontologies were built, (2) the methodologies employed to build the ontologies, (3) the full or partial ontology graphs, (4) theoretical explanations of the meaning of concepts and axioms, (5) full or partial codes of the ontologies, (6) detailed descriptions of the use of these ontologies in real world semantic-based projects, etc.

TABLE II
CANDIDATE E-GOVERNMENT DOMAIN ONTOLOGIES PART I

Code	Ontology	Selected Data Sources	Project
O_1	**DIP ontology**	Gugliotta et al. [19]	DIP
	Legacy ontology		
	Workflow ontology		
	Service ontology		
	Life-event ontology		
	E-government domain ontology		
O_2	**3 kinds of ontologies**	Sabucedo & Rifon [20]	Academic work
	Life-event ontology		
	Variable ontology		
	Legal document ontology		
O_3	E-government Business ontology	Xiao et al. [34]	Academic work
O_4	LKIF-core ontology	Breuker et al. [12]	Estrella
O_5	Social care ontology	Barthes & Moulin [21]	TerreGov
O_6	Life-event ontology	Sanati & Lu [22]	Academic work
O_7	FEA-RMO ontology	Allemang & Hodgson [10]	OSERA
	PRM ontology		
	BRM ontology		
	SRM ontology		
	TRM ontology		
	DRM ontology		
O_8	**Access-eGov ontology**	Hreno et al. [25]	Access-eGov
	Life-event ontology		
	Service profiles ontology		
	Domain ontology		
O_9	Life-event ontology	Todorovski et al. [13]	OneStopGov
O_{10}	Process document ontology	Puustjarvi [26]	Academic work
O_{11}	**SAKE ontology**	Butka et al. [11]	SAKE
	Public Administration ontology		
	Process and Profile ontology		
	Information ontology		
	Decision making quality ontology		
O_{12}	**OntoGov ontology**	Apostolou et al. [23], [24]	OntoGov
	Legal ontology		
	Organizational ontology		
	Life-cycle ontology		
	Domain ontology		
	Service ontology		
	Life-event ontology		
	Profile ontology		
	Web Service Orchestration ontology		
O_{13}	**3 kinds of ontologies**	Chen et al. [27]	Academic work
	E-government ontology		
	Regulatory ontology		
	Service ontology		
O_{14}	E-government services ontology	Fraser et al. [28]	SmartGov
O_{15}	GEA ontology	Goudos et al. [17]	SemanticGov

[10][12][11][13]; this may promote the reuse and evolution of the corresponding domain ontologies.

Finally, Table 6 provides the URLs of Web sites of e-government projects under which the selected ontologies in Table 4 were developed; these Web links may provide the interested reader access to more information on the selected e-government domain ontologies in Table 4. Related studies are discussed in the next section.

IV. RELATED WORK

In [9] the Swoogle ontology search engine is used to search multimedia ontologies on the Semantic Web; the search in Swoogle is based on domain keywords and their combinations; the data sources of the targeted multimedia ontologies are not considered for selecting ontologies specific keywords that are likely to improve the search results.

A strategy for searching biomedical ontologies is presented in [8]; the strategy relies on the keywords search in Swoogle; the keywords used are extracted from related Web pages retrieved with domain keywords search in Google; the data sources on the targeted domain ontologies that may help identifying ontologies specific concepts for the search are not considered.

In [4] an infrastructure for searching and reusing distributed ontologies is presented . The proposed infrastructure is composed of many ontology servers or nodes that store and maintain ontologies; a domain ontology to be searched is described in a meta-ontology with information such as the ontology author, ontology location and used ontology language; the meta-ontology is further improved with a list of ontology terms by matching each ontology concept to the WorldNet lexical semantic net; finally, the meta-ontology

TABLE III
CANDIDATE E-GOVERNMENT DOMAIN ONTOLOGIES PART II

Code	Ontology	Selected Data Sources	Project
O_{16}	**Real-estate transaction ontology** Real-estate ontology Person ontology Organizational ontology Legislation ontology Location ontology Tax ontology Contract model ontology Jurisprudence ontology Civil personality ontology Real-estate transaction verification ontology	Ortiz-Rodriguez & Villazon-Terrazas [29]	Reimdoc
O_{17}	**Government ontology** Address ontology Association ontology Bank ontology Company ontology Currency code ontology Driving licence ontology Legal person ontology Local government unit ontology Natural person ontology Non Natural ontology Partnership company ontology Professional association ontology Shareholding company ontology Vehicle ontology Vehicle engine ontology	Jarrar et al. [16]	Zinnar

TABLE IV
SELECTED E-GOVERNMENT DOMAIN ONTOLOGIES

Code	Ontology	Codification Language	Semantic Coverage	Modularity	Open Availability
O_7	FEA-RMO ontology	OWL	High	5 domain ontologies	publicly available
O_4	LKIF-core ontology	OWL	High	15 domain ontologies	Publicly available
O_9	Life-event ontology	OWL	High	1 generic	Publicly available
O_{11}	SAKE ontology	OWL	High	3 modules	Publicly available
O_{15}	GEA ontology	OWL	High	1 generic model	Publicly available
O_{17}	Government ontology	Not publicly available	High	15 domain ontologies	Publicly available

is stored in an ontology registry, providing a compact representation for efficient search and reuse of related ontologies. However, to build a meta-ontology for searching targeted domain ontologies, the ontology engineer need to have prior knowledge of the targeted ontologies; but, it is unclear in the study how such prior knowledge could be acquired. The available data sources of ontologies in the domain could be of help to the ontology engineer in this case.

The underlying algorithms of ontology and semantic search engines including Swoogle, OntoSearch and OntoKhoj are presented in [30][31][32], respectively. However, the search in these search engines is based on keywords [8]; but, the scope of these studies do not address the issue of selecting relevant domain and specific ontology keywords for the search. This study performs a content analysis of ontology data sources based on predefined ontology features to guess specific concepts for searching domain ontologies on the Semantic Web.

Ontology editors such as Protégé allow the reuse of an existing ontology in another ontology being designed [6];

furthermore, the Web Ontology Language (OWL) offers the possibility to import an OWL ontology into a new ontology under development [33][6]; both ontology reuse solutions require the ontology engineer to have good knowledge and understanding of the existing domain ontologies to be integrated or imported; once more, locating existing domain ontologies and their data sources may be of assistance to the ontology engineer in these cases.

Other solutions for semi-automatic and automatic ontology reuse are presented in [4][5][7]. However, there remains some general challenges in these ontologies reuse solutions: (1) locating relevant domain ontologies for reuse [4], (2) determining appropriate concepts for searching targeted ontologies and (3) understanding the discovered ontologies. This study may be used as a pre-investigative task to existing semi-automatic and automatic ontology reuse solutions in the sense that it enables the ontology engineer to search and retrieve existing domain ontologies along with their data sources; this information may help the ontology engineer in analyzing, understanding and reusing the discovered ontologies. Furthermore, in [2] the authors described the process of reusing

TABLE V
SELECTED E-GOVERNMENT DOMAIN ONTOLOGIES AND WEB LINKS TO THEIR DATA SOURCES

Ontology	Links to Data Sources
Government ontology	http://zinnar.pna.ps/ontologyServer/
	http://www.jarrar.info/publications/JDF11.pdf
LKIF-core ontology	http://www.estrellaproject.org/lkif-core/lkif-core.owl
	http://www.estrellaproject.org/lkif-core/legal-role.owl
	http://www.estrellaproject.org/lkif-core/lkif-rules.owl
	http://www.estrellaproject.org/lkif-core/legal-action.owl
	http://www.estrellaproject.org/doc/D1.4-OWL-Ontology-of-Basic-Legal-Concepts.pdf
FEA-RMO ontology	http://protege.cim3.net/file/work/ontology/FEARMO/
	http://www.osera.gov/owl/2004/11/fea/brm.owl
	http://www.osera.gov/owl/2004/11/fea/prm.owl
	http://www.osera.gov/owl/2004/11/fea/srm.owl
	http://www.osera.gov/owl/2004/11/fea/trm.owl
Life-event ontology	http://islab.uom.gr/onestopgov/index.php?name=UpDownload&req=getit&lid=459
	http://islab.uom.gr/onestopgov/index.php?name=UpDownload&req=getit&lid=460
SAKE ontology	www.sake-project.org/fileadmin/filemounts/sake/DeliverableD6b.pdf
GEA ontology	http://islab.uom.gr/semanticgov/index.php?name=UpDownload&req=getit&lid=454
	http://islab.uom.gr/semanticgov/index.php?name=Web_Links&req=visit&lid=65

TABLE VI
URLs OF PROJECTS WEBSITES OF THE SELECTED E-GOVERNMENT DOMAIN ONTOLOGIES

Code	Ontology	Projects	Websites Links
O_7	FEA-RMO ontology	OSERA	http://osera.modeldriven.org/projects/fearmo.htm
O_4	LKIF-core ontology	ESTRELLA	http://www.estrellaproject.org/
O_9	Life-event ontology	OneStopGov	http://islab.uom.gr/onestopgov/
O_{11}	SAKE ontology	SAKE	http://www.sake-project.org/
O_{15}	GEA ontology	SemanticGov	http://islab.uom.gr/semanticgov/
O_{17}	Government ontology	Zinnar	http://zinnar.pna.ps/

and applying existing ontologies and concluded that reusing ontologies is far from an automatic process and requires significant effort from the knowledge engineer; this assertion is also supported in [3].

V. CONCLUSION

This study presents a framework that uses an adaptive technique based on ontology and generic search engines, and predefined ontology features to search and locate domain ontologies and their data sources over the Semantic Web. The predefined ontologies features are used to learn ontology specific concepts from the data sources; these concepts are further employed to improve the quality of the search results.

The application of the framework in the e-government domain permitted the discovery of 62 candidate e-government domain ontologies; furthermore the framework enabled the application of predefined criteria including semantic coverage, open availability, codification language, and modularity on the candidate ontologies to select the best reusable set of ontologies for the e-government domain. The selected ontologies provide a good sharable and reusable conceptual representation and description of the public administration domain as well as the electronic services delivery processes; this may promote their reuse across semantic-based e-government projects.

The study may be used as a pre-investigative task to existing automatic and semi-automatic ontologies reuse solutions which require the ontology engineers to have prior knowledge of the targeted ontologies to guide the process for building new domain ontologies from existing ones.

The framework of the study may be applied in any application domains of Semantic Web such as e-commerce, e-business, e-learning, multimedia, etc., to identify and analyze existing domain ontologies for the purpose of knowledge sharing and reuse across domain specific Semantic Web applications.

The future direction of the research will be to conceptualize and build a generic ontology model for the e-government domain through the reuse of the discovered domain ontologies.

REFERENCES

[1] T. R. Gruber, "Toward Principles for the Design of Ontologies used for Knowledge Sharing," *International Journal of Human-Computer Studies*, Vol. 43, pp. 907-928, 1993.

[2] M. Ushold, M. Healy, K. Williamson, P. Clark and S. Woods, "Ontology Reuse and Application," *In Proceedings of the 1st International Conference on Formal Ontology and Information Systems - FOIS'98*, Trento, Italy, pp. 179-194, 1998.

[3] Maedche, A. & Staad, S. (2001). Ontology Learning for the Semantic Web. IEEE Intelligent Systems, 16, 72–79.

[4] A. Maedche, B. Motik, L. Stojanovic, R. Studer and R. Volz, "An Infrastructure for Searching, Reusing and Evolving Distributed Ontologies," *In Proceedings of the World Wide Web Conference (WWW 2003)*, Budapest, Hungary, pp. 439-448, 2003.

[5] Y. Ding, D. Lonsdale, D. W. Embley, M. Hepp and L. Xu, "Generating Ontologies via Language Components and Ontology Reuse," *In Proceedings of the 12th International Conference on Applications on Natural Language to Information Systems (NLDB'07)*, Paris, France, pp. 131-142, 2007.

[6] P. Doran, V. Tamma and L. Lannone, "Ontology Module Extraction for Ontology Reuse: An Ontology Engineering Perspective," *In: International Conference on Information and Knowledge Management (CIKM'07)*, Lisboa, Portugal, pp. 61-69, 2007.

[7] M. d'Aquin, M. Sabou and E. Motta, "Reusing Knowledge from the Semantic Web with the Watson Plugin," *In Proceedings of the 7th International Semantic Web Conference (ISWC)*, Karlsruhe, Germany, 2008.

[8] H. Alani, N. Noy, N. Shah, N. Shadbolt and M. Musen, "Searching Ontologies Based on Content: Experiments in the Biomedical Domain," *In Proceeding of the Fourth International Conference on Knowledge Capture (K-Cap)*, Whistler, BC, Canada, pp.55-62, 2007.

[9] G. A. Atemezing, "Analyzing and Ranking Multimedia Ontologies for their reuse," *MSc Dissertation*, Unversidad Politechnica de Madrid, Madrid, Spain, 2010.

[10] D. Allemang, R. Hodgson and I. Polikoff, "Federal Reference Model Ontologies (FEA-RMO)," *White Paper*, 2005.

[11] P. Butka, A. Gabor, A. Ko, M. Mach, S. Ntioudis, A. Papadakis, N. Stojanovic, R. Vas and T. Zelinsky, "Semantic-enable, Agile, Knowledge-based e-Government (SAKE)," *Deliverable No. 3*, 2006.

[12] J. Breuker, R. Hoekstra, A. Boer, K. Van der berg, G. Sartor, R. Rubino, A. Wyner, T. Bench-Capon and M. Palmirani, "OWL Ontology of Basis Legal Concepts (LKIF-Core)," *Deliverable 1.4*, 2006.

[13] L. Torodovski, M. Kunstelj, D. Cukjati, M. Vintar, I. Trochidis, E. Tambouris, OneStopGove: D13 Life-event Reference Models, Deliverable No. 13, 2007.

[14] R. Fitterer, U. Greiner and F. Stroh, "Towards Facilitated Reuse of Ontology Results from European Research Projects - A Case Study," *In: 16th European Conference on Information Systems (ECIS)*, Galway, Ireland, pp. 1929-1940, 2008.

[15] A. Esposito, M. Zappatore and L. Terricone, "Evaluating Scientific Domain Ontologies for the Electronic Knowledge Domain: A General Methodology," *International Journal of Web & Semantic Technology*, Vol. 2, 1-18, 2001.

[16] M. Jarrar, A. Deik and B. Farraj, "Ontology-Based Data Process Governance Framework The Case of e-Government Interoperability in Palestine," *In IFIP International Symposium on Data-Driven Process Discovery and Analysis (SIMPDA'11)*, Campione, Italy, pp. 83-98, 2011.

[17] S. K. Goudos, V. Peristeras and K. Tarabanis, "Mapping Citizen Profiles to Public Administration Services Using Ontology Implementations of the Governance Enterprise Architecture (GEA) models,"*In Proceedings of the 3rd Annual European Semantic Web Conference*, Budva, Montenegro, pp. 25-37, 2006.

[18] S. K. Goudos, V. Peristeras, N. Lutas and K. Tarabanis, "A Public Administration Domain Ontology for Semantic Discovery of e-Government Services," *In Proceedings of the 2nd IEEE Conference on Digital Information Management 2007 (ICDIM 2007)*, Lyon, France, pp. 260-265, 2007.

[19] A. Gugliotta, L. Cabral, J. Domingue and V. Roberto, "A Conceptual Model for Semantically-Based E-government Portal," *In Proceedings of the International Conference on e-Government 2005 (ICEG 2005)*, Ottawa, Canada, 2005.

[20] L. A. Sabucedo and L. A. Rifon, "Semantic Service Oriented Architectures for E-government Platforms," *American Association for Artificial Intelligence*, 2006.

[21] J. P. Barthes and C. Moulin, "Impact of e-Government on Territorial Government Services," *Deliverable No. 1.4*, 2005.

[22] F. Sanati, J. Lu, "Multilevel Life-event Abstraction Framework for E-government Service Integration," *In Proceedings of the 9th European Conference on E-government 2009 (ECEG 2009)*, London, UK, pp. 550-558, 2006.

[23] D. Apostolou, L. Stojanovic, T. P. Lobo, J. C. Miro,and A. Papadakis, "Configuring E-government Services Using Ontologies," *IFIP International Federation for Information Processing*, Springer Boston, Vol. 2005, pp. 141-155, 2005.

[24] D. Apostolou, L. Stojanovic, T. P. Lobo and B. Thoensen, "Towards a Semantically-Driven Software ngineering Envirionment for E-government," *IFIP International Federation for Information Processing*, M. Bohlen (Eds), Vol. 3416, pp. 157-168, 2005.

[25] J. Hreno, P. Bednar, K. Furdk and T. Sabol, "Integration of Government Services using Semantic Technologies," *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 6, pp. 143-154, 2011.

[26] J. Puustjarvi, "Using Knowledge Management and Business Process in E-government," *In Proceedings of the Information Integration and Web-based Applications and Services 2006 (iiWas2006) Conference*, Yogyakarta, Indonesia, pp. 331-339.

[27] D. Chen, G. Nie and P. Liu, "Research Knowledge Sharing of E-government Based on Automatic Ontology Mapping," *In Proceedings of the 6th Wuhan International Conference on E-Business*, Business, China, pp.105-111, 2008.

[28] J. Fraser, N. Adams, A. Mckay-Hubbard, A. Macintosh and R. Canadas, "A Framework for e-Government Services," *Deliverable No. 71*, 2003.

[29] Ortiz-Rodriguez, F. & Villazon-Terrazas, B. (2006). EGO Ontology Model: Law and Regulation Approach for E-government. In Proceedings of the Workshop on Semantic Web for E-government 2006, Workshop at the 3rd European Semantic Web Conference (pp. 13-23). Budva, Serbia and Montenegro.

[30] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, Y. Peng, P. Reddivari, V. C. Doshi and J. Sachs, "Swoogle: A Search and Metadata Engine for the Semantic Web," *In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management*, Washington, DC, USA, pp. 1-8, 2004.

[31] Y. Zhang, W. Vasconcelos, D. Sleeman, "Ontosearch: An ontology search engine," *In Procceedings of the International Conference on Innovative Techniques and Applications of Artificial Intelligence*, Cambridge, UK, pp. 1-12, 2004.

[32] C. Petel, K. Supekar, Y. Lee and E. K. Park, "OntoKhoj: A Semantic Web Portal for Ontology Searching, Ranking and Classification," *In Proceedings of the Workshop On Web Information And Data Management*, New Orleans, Lousiana, USA, pp. 58-61, 2003.

[33] J. Z. Pan, L. Serafini, Y. Zhao,"Semantic Import: An Approach for Partial Ontology Reuse," *In Proceedings of the 1st Workshop on Modular Ontologies (WoMO06)*, Athens, GA, USA, pp. 1-12.

[34] Y. Xiao, M. Xioa and H. Zhao, "An Ontology for E-government Knowledge Modelling and Interoperability," *In Proceedings of the IEEE International Conference on Wireless Communications, Networking and Mobile Computing, (WiCOM 2007)*, Shanghai, pp. 3600-3603, 2007.