

Building a Core Arabic Ontology About Iraqi News

Abdulhussein M. Abdullah, Miaad Raisan

Abstract - Nowadays, Iraqi newspapers spread on the WWW in a great and remarkable form. All of these Websites are belonging to traditional Web. Therefore, the search results are not perfect. In order to move with these Websites to Semantic Web generation, ontology must be created. Unfortunately, there is a lack in ontologies written in Arabic language because it is a difficult language. If some attempts in different domains exist, it is not available on the World Wide Web as Linked Data. This paper aimed to build a core ontology in Arabic language interested in Iraqi News domain to be used as a source data for Iraqi's newspapers. Through the study the proposed ontology includes classes in hierarchical form depend essentially on class called Event class which play with other classes also, these classes may play with each other. Predicates on classes are relationships between these classes, thus among their individuals. An inference feature is enabled by adding restrictions on predicates. ORM (Object Role Modeling) approach is used to design the verbalization conceptual model for our ontology. Ontology mapping is used for populating the proposed ontology by converting XML documents to OWL using XSLT.

Index Terms— Arabic Ontology, OWL, ORM, semantic web, XSLT, ontology population.

I. INTRODUCTION

Despite the role that is played by WWW in publishing and sharing information, the information cannot be retrieved to the user properly. Because the search process is performed by entering a word, this possibly leads to ignore some of the content that does not match this word. Actually, there is a lack in processing semantic meaning for information in both search engines and Web Pages. This can give rise to poor process to synonym and homonym words, as example. So, it becomes necessary to rely on the meaning of the contents which is contained within Web Pages. This idea was referred by Tim Berners Lee et al. [1] in an article published in the Scientific American Magazine, who described the Semantic Web technology as an informal definition:

“The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

In order to make the computer able to understand the meaning, it must have vocabularies as knowledge base which can rely upon in interpreting the meaning of the contents of

Web Pages. Such knowledge base is called an ontology, which represents the essential part for Semantic Web applications. There are different definitions for ontology and the popular one was given by Gruber as[2]:

“An explicit specification of a conceptualization; in which a conceptualization is an abstract, simplified view of the world that we wish to represent for some purpose.” On the other hand, W3C¹ is formally define it as[3]: *“A common set of terms that are used to describe and represent a domain.”*

Generally, Ontology can be classified into two types[4]:

- Upper Level Ontology represents a generic model for common concepts that are generally applicable across a wide range of domain ontologies; such as SUMO Ontology.²
- Domain Ontology represents a conceptualization of specific domain of the world, such as Music Ontology.³

There are a lot of ontologies in different languages such as English, German, Spanish and others with different rates. However, there is a lack in Arabic ontologies that available on the web, as shown in Fig 1. The reason of this lack is that Arabic language is considered as a difficult language. It has some specificity, such as complex morphological, grammatical, and semantic aspects since it is a highly inflectional and derivational language [6]. As example, to construct ontology or to populate it, in some cases, it must use some NLP technique. NLP tools are limited and cannot exactly meet with Arabic language, for it was originally designed for English language. Despite these obstacles, there are

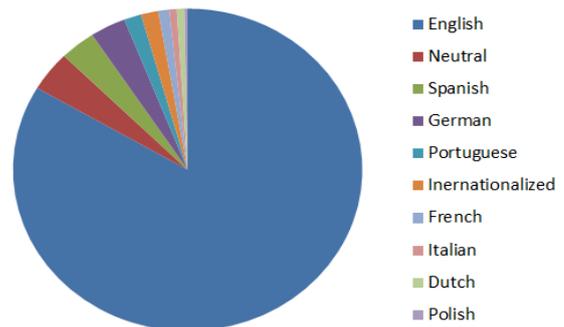


Fig 1 Representation of the most used languages for developing ontologies [5]

some attempts to build Arabic ontologies, such as the Arabic Ontology [7] and Legal Ontology [8]. Nowadays, Iraqi newspapers spread on the WWW in a great and remarkable form. All of these Websites are belonging to traditional Web. In order to move with these Websites to Semantic Web generation, an ontology must be created.

Manuscript published on 30 August 2015.

* Correspondence Author (s)

Abdulhussein M. Abdullah*, Computer Science, University of Basrah/ College of Science/ Basrah, Iraq.

Miaad Raisan, Computer Science, University of Basrah/ College of Science/ Basrah, Iraq.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

1 This acronym comes from "World Wide Web Consortium." Available at: <http://www.w3.org/>
2 This acronym comes from "Suggested Upper Merged Ontology." Available at: <http://web0.site.uottawa.ca:4321/sumo>.
3 <http://musicontology.com>.

Building a Core Arabic Ontology About Iraqi News

This ontology aims to the standardization of terminology with coherent description for concepts. So, we proposed building Arabic ontology which is interested in the Iraqi political news as a core, to be taken as a base including all Iraqi affairs.

The remainder of this paper is organized as follows: Sections II presents a brief of related work. Section III briefly describes the architecture of ontology. Section IV presents the construction of Iraqi News ontology includes their philosophy, modeling and how to populate it. Finally, section V presents the conclusion.

II. RELATED WORK

Some ontologies that are presented in the news domain used existing standards to refer to the topics. All of these standards are found in XML schema format.

Castells et al. [9] proposed an ontology includes generic categories, such as “politics” or “sport,” but it does not include classes and entities for these categories. This ontology uses IPTC¹ standards, which are converted into RDF² schema.

García et al. [10] proposed to build ontological framework based on journalism (NITF,³ NewsML,⁴ and NewsCodes⁵) and multimedia standards (MPEG-7)⁶. Moreover, they developed a tool to convert these standards from XML to OWL⁷ structure.

Fernández et al. [11] proposed an ontology to cover different types of metadata that can be attached to news items. This ontology uses ontologies, such as SUMO⁸ and its complement MILO,⁹ IPTC and Dublin Core,¹⁰ also it converts all standards to OWL structure. All ontologies in [9], [10], and [11] are created in Latin character set.

In the case of Arabic language there is an attempt to build an ontology for the Arabic news by Saleh and Al-Khalifa [12], in which they were classified Arabic news entities into four classes (person, organization, political location and company). Despite this, their classification is insufficient to cover the Arabic news and their events.

III. ONTOLOGY ARCHITECTURE

Generally, ontology is built from concepts which can be considered as general entities called classes. These concepts are usually arranged in a taxonomic hierarchy and associated together by relations. They are called properties or sometimes called slots or roles. Ontology schema consists of a set of axioms on properties to preform logical expression; they are called constraints or sometimes called restrictions or facets. Each class contains set of instances which represent individuals in ontology. Fig 2 illustrates ontology components.

OWL is an ontology language for the Semantic Web. It provides a representation for classes, properties, individuals, data values and restrictions. OWL document, knows as ontology, is stored as Semantic Web document designed to

represent knowledge about things, groups of things, and relations between things[13]. Moreover, OWL provides some properties to enable OWL ontology that published in WWW to use another OWL ontologies published in WWW, too. In OWL the property divided into two types [14]:

- 1) Object Property is a relation that links pairs of individuals;
- 2) Datatype Property is a relation that links individual with literals;
- 3) Annotation Property is a relation that can be used to add an annotation for ontology; both object and datatype Property consider semantic properties unlike annotation property.

Generally, there are three techniques for developing ontologies. They are: manually ontology development, ontology learning, and ontology mapping. The popular strategy according to manually ontology development was proposed by Noy and McGuinness. This strategy follows seven steps [15]:

1. Determine the domain and scope of the ontology;
2. Consider reusing existing ontologies;
3. Enumerate important terms in the ontology;
4. Define the classes and the class hierarchy;
5. Define the properties of classes—slots;
6. Define the facets of the slots;
7. Create instances.

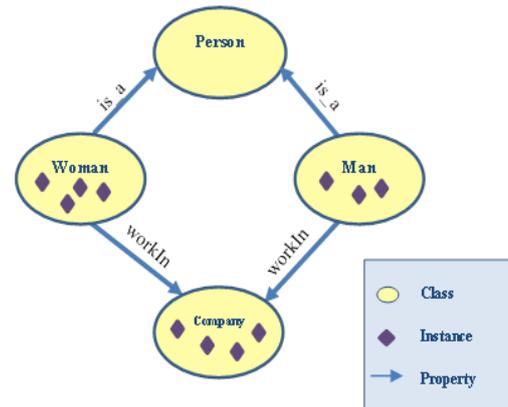


Fig 2 Sample of Ontology

In this article, we used this strategy to build the proposed ontology schema, at first, and then using ontology mapping technique for populating it.

IV. IRAQI NEWS CORE ONTOLOGY (IQNCONTO)¹¹ CONSTRUCTION

Before embarking to our ontology development process, there is a preparation process that preceds it, which aimed to identify terms. This step is based on a number of sources from famous Iraqi agencies such as Iraqi Media Network,¹² Alsumaria News,¹³ and Albaghdadia News.¹⁴ For three months, the articles have been collected manually and in a personal effort. Actually, one took this time in order to give a chance for diversity news. As a final outcome, about 1000 articles have been collected all of them focus upon the political affairs of Iraq.

1 This acronym comes from “International Press and Telecommunications Council.” Available at: <http://www.iptc.org>.
2 This acronym comes from “Resource Description Framework.”
3 This acronym comes from “News Industry Text Format.” Available at: http://www.iptc.org/site/News_Exchange_Formats/NITF
4 <http://www.newsml.org>
5 <http://www.iptc.org/site/NewsCodes>
6 <http://mpeg.chiariglione.org/standards/mpeg-7>
7 This acronym comes from “Web Ontology Language.”
8 This acronym comes from “Suggested Upper Merged Ontology.” Available at: <http://web0.site.uottawa.ca:4321/sumo>.
9 This acronym comes from “Mid-Level Ontology.”
10 <http://dublincore.org>.

11 <http://purl.org/net/iqnconto>
12 <http://www.imn.iq/>
13 <http://www.alsumaria.tv/news>
14 <http://www.albaghdadianews.com/>

The proposed ontology is based on manual methodology to build a sample from it. At the beginning of the study, it is necessary to give answers for the questions, which are raised by Noy and McGuinness [15]. These questions help us to give a clear vision to design our ontology as follows:

1) What is the domain that the ontology will cover?

The domain of the proposed ontology will cover Iraqi political news involved with events, persons, organizations and the effects with each other.

2) For what we are going to use the ontology?

We plan to use our ontology in Semantic Web applications, such as using it inside the structure of Websites (e.g. newspapers).

3) For what types of questions the information in the ontology should provide answers?

Our proposed ontology aimed to provide answers to questions related to Iraqi news and their events, like:

a) Who are the politicians and what their party affiliations?

b) Who have been targeted by assassinations, arrests, or explosion?

c) What are the ingredients of party alliances in the political blocs?

d) For security events, which affected certain cities in a particular time? What are they? And what tools are used in these events?

4) Who will use and maintain the ontology?

The proposed ontology will be available for any developer to reuse it, or for any application that is interested in News Domain, especially Iraqi news. Also, it must be clear, i.e. to be understandable for developers, and scalable to add any possible developments for this domain.

According to reusing an existing ontology represents an important thing in the ontology creating process. However, one can decide to reuse a class form an existing ontology known FOAF¹

A. Iraqi news ontology philosophy

First off, the concepts have been collected from news headlines only without incursion in the articles details. Because the articles in these agencies are usually short and the headlines give sufficient expressions.

After studying the Iraqi news, six classes have been categorized for proposing ontology. The ontology focuses on the Event class (حدث) due to the fact that news is events influenced by or influence people, organizations, etc. In addition, an event may produce another event and it may need some tool for doing so. Moreover, it needs place and time to happen. Actually, all of these possibilities were taken into consideration when the classes have been proposed. Table I shows the proposed classes for Iraqi News Ontology.

According to Person Class, it is created with base URI (<http://xmlns.com/foaf/0.1/person>). Also, one can give additional rdfs:label value of "شخص"@ar, where "ar" represents ISO 6392 code for Arabic language. If someone in Spain gave an additional rdfs:label value of "persona"@es. So, the data that are created, using foaf:person refer to the

same class. But the information displays an appropriate label for it in Arabic and Spanish. The classes have been organized in a hierarchical taxonomy (see Fig 3). These classes are more general with the possibility of dividing them into subclasses according to future requirements. Actually, classes alone in the ontology schema will not give a sufficient vision of how the ontology components interact with each other. So, it should define properties to reflect the internal structure of ontology schema.

Class	Description
حدث (Event)	Represents everything happens by someone, organization or another event in Iraqi political area such as (explosion, election, conference, etc.)
شخص (Person)	Represents any (leader of political party, minister, president, journalist, etc.)
منظمة (Organization)	Represents any (government regulation, political party, the ministry, institution, university, etc.)
مكان (Place)	Represents every stats, province, city, town ... etc.
زمان (time)	Represents the time that an event happened
اداة الحدث (eventTool)	Represents anything that assist an event to be, such as (bomb is a tool that caused an explosion)

Table I Iraqi News Ontology Classes

As has been presented in previous the properties in OWL can be divided into Object Property and Datatype Property. Therefore, in this thesis 24 object properties have been proposed as table II illustrates.



Fig 3 Iraqi News Ontology Classes in Hierarchy

Taking into account, when one design ontology schema, there might be an inverse characteristic for each property. This means, there exists a property in opposite direction with any property.

Table II Iraqi News Ontology Object Properties

No.	Property	Description
-----	----------	-------------

¹ This acronym comes from "Friend of a friend."

² ISO 639 is a set of standards by the International Organization for Standardization that is concerned with representation of names for language and language groups.

Building a Core Arabic Ontology About Iraqi News

1	أحد_أعضائها (hasMember)	Object property links organization (منظمة) to person (شخص) who belongs to it
2	أحد_مكوناتها (hasComponent)	Object property links organization (منظمة) to organization (منظمة) which belongs to it
3	أحدث (generate)	Object property links the eventTool class (اداة_الحدث) to the event class (حدث), that caused by this tool
4	أدى_إلى (ledTo)	Object property links an event (حدث) with another event (حدث) that generated from it
5	بسبب (dueTo)	Object property links an event (حدث) with another event (حدث) that caused by
6	جزء_من (partOf)	Object property links an organization (منظمة) to another organization (منظمة), which is part of it
7	حدث_عن_طريق (isDoneBy)	Object property links the event class (حدث) with the eventTool class (اداة_الحدث) that caused this event
8	حدث_في (OccurredIn)	Object property links event (حدث) to place (مكان) where this event happened
9	أستهدف (targeted)	Object property links an event (حدث) to person (شخص) affected by this event
10	حدث_في (isPlaceOccurrence)	Object property links place (مكان) to event (حدث) that happened in this place
11	حدث_من_قبل (happenedByOrg)	Object property links the event class (حدث) to the organization class (منظمة), which is responsible for this event
12	عضو_في (member)	Object property links the person class (شخص) to the organization class (منظمة), which he belongs to
13	قام_بـ (dose)	Object property links the person class (شخص) to the event class (حدث), which was done by him
14	قامت_بـ (do)	Object property links the organization (منظمة) to the event class (حدث), which was done by this organization
15	له_هدف (hasTarget)	Object property links the event class (حدث) to the organization class (منظمة), which is affected by this event
16	لها_مسؤول (hasResponsible)	Object property links the organization class (منظمة) to the person class (شخص), who is (manager, leader, etc.) for this organization
17	مسؤول (responsible)	Object property links the person class (شخص) to the organization class (منظمة), which it managed by this person
18	مستهدف_من (targetOf)	Object property links the person class (شخص) to the event class (حدث), which targeted this person
19	مستهدفة_من_قبل (targetedBy)	Object property links the organization (منظمة) to the event class (حدث), which targeted this organization

20	اثناء (during)	Object property links the place class (مكان) with the time class (زمان), it defines the time of the event in specific place
21	في (in)	Object property links the time class (زمان) with the place class (مكان), it defines the scene of event in specific time
22	وقعية (TimeOccurrence)	Object property links the Time class (زمان) to the Event class (حدث) that happened at this time
23	وقع_في (OccurredAt)	Object property links the Event class (حدث) to the Time class (زمان) when this event happened
24	وقع_من_قبل (happendBy)	Object property links the event class (حدث) to the person class (شخص), who is responsible for this event

A transitive characteristic has also been added to partOf (جزء_من) and hasComponent (أحد_مكوناتها) properties. Because some organizations are parts of the other in series. So, it has been added to declare a part of whole and contains relationships. Surely, adding a characteristic like this will help to infer information through a query.

In addition to the object property, five Datatype properties have been proposed as illustrated in Fig 4. They help us to describe the type data for individuals. In Iraqi news ontology the range for all Datatype property is a string, while the domain in each case represents one class in our proposed ontology.



Fig 4 Iraqi News Ontology Datatype Properties

B. Iraqi News Ontology Modeling

Ontology modeling step provides us a clear vision for ontology structure. In general, the main maintenance and management of ontology in graphical model are easier than them in RDF/XML serialization.

Object Role Modeling (ORM) has been chosen to model Iraqi News Ontology as a conceptual modeling language, due to the advantage of its verbalization. Moreover, some of its constraints are similar to OWL restrictions, thus leading to high flexibility in mapping between these two approaches. The complete conceptual schema for Iraqi news ontology is shown in Fig 5.



8. S. Zaidi, M. Laskri, and K. Bechkoum, "A cross-language information retrieval based on an Arabic ontology in the legal domain," in Proceedings of the International Conference on Signal-Image Technology and Internet-Based Systems (SITIS'05), 2005, pp. 86-91.
9. P. Castells, F. Perdrix, E. Pulido, M. Rico, R. Benjamins, J. Contreras, et al., "Neptuno: Semantic web technologies for a digital newspaper archive," in The Semantic Web: Research and Applications, ed: Springer, 2004, pp. 445-458.
- 10.R. García, F. Perdrix, and R. Gil, "Ontological infrastructure for a semantic newspaper," in Semantic Web Annotations for Multimedia Workshop, SWAMM, 2006.
11. N. Fernández, D. Fuentes, L. Sánchez, and J. A. Fisteus, "The NEWS ontology: Design and applications," Expert Systems with Applications, vol. 37, pp. 8694-8704, 2010.
- 12.L. M. B. Saleh and H. S. Al-Khalifa, "AraTation: an Arabic semantic annotation tool," in Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services, 2009, pp. 447-451.
- 13.W3C. OWL 2 Web Ontology Language Primer. Available: <http://www.w3.org/TR/2009/PR-owl2-primer-20090922/>
- 14.J. Hebel, M. Fisher, R. Blace, and A. Perez-Lopez, Semantic web programming: John Wiley & Sons, 2011.
- 15.N. F. Noy and D. L. McGuinness, "Ontology development 101: A guide to creating your first ontology," ed: Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, 2001.
- 16.H. Bohring and S. Auer, "Mapping XML to OWL Ontologies," Leipziger Informatik-Tage, vol. 72, pp. 147-156, 2005.