



Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)



Enhancing Semantic Interoperability in Bird Classification through XML/RDF and SPARQL

Asraa Mounaf Almousawy

Department of Computer Science , Faculty of Education for Girls ,University of Kufa , Najaf , Iraq

Israam.almoussawy@uokufa.edu.iq

ARTICLE INFO

Article history:

Received: 12/03/2024

Revised form: 20/04/2024

Accepted : 19/05/2024

Available online: 30 /06/2024

Keywords:

Ontology Learning (OL),
SPARQL, XML/RDF,
Information Extraction, Bird
Classification, Semantic
Interoperability, Knowledge
Organization.

ABSTRACT

Classification of birds is the area that is quite complex and involved such a lot of varieties that demands the correct as well as the professional organization of the information. This study addresses the problem of semantic interoperability in the bird categorization process by means of constructing an ontology which covers all the words used in taxonomic descriptions. The paper exploits XML/RDF standards for semantic web compatibility and Open Link Virtuoso SPARQL Query Editor that enables simple interaction with its tools and visualization in the process of querying and presentation of results. Attempt to develop a specific ontology for bird identification for the purpose of optimizing the accessibility and retrieval of heterogeneous bird-oriented data that ultimately help in building effective knowledge management system in this field. The effectiveness of an ontology is evaluated by its ability to make possible classification of diverse information. This research methodology is proposed for implementation in a broad range of researches, education, and conservation programs that can targeted to enhance their output and increase accuracy.

MSC..

<https://doi.org/10.29304/jqcm.2024.16.21540>.

1. Introduction

Classification of birds is a complex task that needs a precise way of storing data. This research study is focused on the problem of semantic interoperability in bird classification by creating a detailed ontology. Ontology is a philosophical term derived from two Greek words ontos (being) and logos (word), that was created for denoting the study of existing types of beings [1]. Through the years, Artificial Intelligence (AI) and Computational Linguistics (CL) lead to creation of ontologies that listing the

*Corresponding author

Email addresses:

Communicated by 'sub etitor'

concepts, properties, and relationships and produce structured knowledge. The combination of AI and ontologies has led to improvements in semantic search, automated engineering, and machine learning [2, 3].

Ontologies have leading roles in knowledge representation, knowledge management, as well as exchange and reuse of such knowledge in the systems [4]. There are many fields of study that use ontology as a source of knowledge like as: semantic web, search engines, e-commerce, natural language processing, knowledge architecture, information extraction and retrieval, multifactor systems, dynamic modeling of spatiotemporal systems, database design, geographic information science, and digital libraries [5]. Utilization of XML/RDF standards for semantic web compatibility and Open Link SPARQL Query editor ensures executing queries and outputting the results respectively. Through the development of the bird classification ontology the purpose is within is to establish a more unified, organized and retrievable bird-related information database pursuing of perfection in knowledge management in this particular domain. The usefulness of the ontology is gauged from the efficiency and precision in which it categorizes data coming from various sources. The research described herein looks at a method that may be of much help to experts, educators, and conservationists who want to classify birds in an accurate and speedy way.

1.1. The problems of ontologies with some solution

Ontology is recognized as $O = \{(C, R, A, Top)\}$, where C is the nonempty set of concepts, R is set of all relations that two or more concepts are related, A are statements that bound the ontology, and Top is the highest level of the hierarchy which also is the root. The R language is subdivided into the H and N partitions. The H partition consists of all hypotheses where the relationship involves taxonomy, and the N partition constitutes all assertions where the relation is non-taxonomy. Cooperative and goes to the two-way reflection, which connects R and reality [5].

A knowledge representation system is an important component of artificial intelligence applications. It stores complex and possibly incomplete domain models and uses reasoning techniques to answer queries about the knowledge. The properties of a knowledge representation system include the ability to express and describe ontology objects in different domains, dynamic expansion of different domains, flexible expression of knowledge using familiar vocabulary, and the generation of vocabulary packs based on an ontology language library [6].

Modeling knowledge related to a particular area of interest is a challenging task for researchers. It takes a great deal of time and money, which is compounded when we consider that concepts and relationships within an ontology must be updated on a regular basis. As a result of these problems, researchers have developed semi-automated methods for creating and maintaining ontology such as: ontology learning (OL), which uses linguistic resources as well as natural language processing (NLP) capabilities to build and deploy an ontology. [7, 8].

Getting an ontology manually is a time consuming and inefficient process. It entails an in-depth understanding of an area, and the output is often inadequate or wrong. Manually generated ontologies

are expensive, time-consuming, error-prone, developer-biased, and inflexible, particularly with regard to the reason for which they are created. Researchers try to solve the defects of manual ontology construction by using semi-automatic or automated methods. Automating the creation of an ontology not only saves time and money, but also results in an ontology suitable for its intended use. Several methodologies and systems for learning ontology have been introduced in the past decade. They use two methods of constructing an ontology. One way is to create ontology building tools such as Protege-2000 and Edit, which are used by knowledge engineers and domain professionals. Another option is to develop an ontology semi-automatically or automatically using data from several sources [8].

The key issues that are addressed in the development of ontologies are:

- a. Lack of criteria for merging and reusing existing ontology.
- b. A fixed hierarchy based on fixed categories based on a single point of view would be too rigid to allow for a vastly diverse perception of our reality.
- c. Absence of fully automatic ways of acquiring knowledge[5].

2. Ontology Learning: Step by Step

In recent years, a number of methodologies and systems for learning ontology have been developed. These techniques generally fall into two categories: manual construction with tools such as Protégé and Edit, designed for use by knowledge engineers and domain professionals, or automated and semi-automated approaches that learn ontology from diverse knowledge sources. Learning ontology involves extracting existential elements, or conceptual knowledge, from input and using them to construct an ontology. Like Cimiano et al. (2005), this process is a critical step in facilitating knowledge sharing and reuse [5].

Machine learning, knowledge acquisition, natural language processing, information retrieval, artificial intelligence, inference, and database management are just a few of the areas used in ontology learning.

Ontology learning systems can be classified based on the types of data they are learned from. So it can learn from unstructured, semi-structured and structured data. Documents written in natural language, such as books and magazines, are an example of unstructured data, while HTML and XML files are examples of semi-structured data. Databases and dictionaries are examples of structured data [6].

The methodologies divide the ontology-building process into a number of stages, each with a set of activities to complete. The relevance of a particular activity within methodology is primarily determined by factors such as the characteristics of the ontology-based application, the complexity of the ontology to be developed, the availability of information sources, and the experience of the ontology architects.

The maintenance and application of ontology are post-development tasks. In parallel with the main development processes, ontology support activities such as Knowledge Acquisition (KA), Evaluation, Reuse, and Documentation are carried out.

The methodologies also define the roles of the participants in the ontology development process. They distinguish between domain experts who provide knowledge about the domain to be designed, ontology

architects who specialize in topics such as knowledge representation or ontology tools, and users who use ontology for a specific purpose. [10]

Ontology learning is interdisciplinary in nature, attracting researchers from a wide range of disciplines, including knowledge representation, logic, philosophy, databases, machine learning, natural language processing, image processing, and so on. As a result, learning ontology has benefited from a massive exchange of ideas and approaches that have formulated a slightly different perspective on the challenge of knowledge acquisition.. [9].

Manual ontology construction and management is expensive and time-consuming, and the ontologies that are created are only static and context-dependent representations of knowledge. The goal of ontology learning is to develop a sufficient amount of knowledge, describing difficult-to-processable resources, automatically (or semi-automatically) and dynamically, and describe it in machine-processable manner as ontologies. In quickly expanding domains, ontologies must be taught, enriched, populated, and evaluated constantly, hence ontology learning, automatic evolution, and assessment approaches are critical.

Ontology development can be done by building an ontology from scratch, or by expanding (filling in or enriching) an existing ontology; By integrating previously developed ontology; And by specializing or enriching general ontology, in order to adapt it to a specific field. Ontology learning uses a large variety of methods and algorithms from several fields such as machine learning, knowledge acquisition, natural language processing, information retrieval, text mining, artificial intelligence, reasoning, and database management [10].

Individual Ontology Approaches Individual ontology techniques use a single global ontology that provides a common vocabulary for semantic specifications. All data sources are linked to a single global ontology. Single ontology techniques can be used to solve integration difficulties where all data sources to be combined have an almost identical domain representation. However, if an information source has a distinct perspective on a domain, because by presenting a different level of detail, defining the minimum existential commitment can be difficult. It is much easier to combine sources of information that link to catalogs of similar products. Changes in sources of information may also affect understanding of the field represented in ontology, leaving individual ontology approaches vulnerable. Changes in one information source can lead to changes in global ontology and mappings to other information sources, depending on the nature of the changes [11].

Multiple ontologies: allow for the representation of different information resources with their own unique ontologies. This approach does not require a universal ontology that all sources agree on, making it easier to make changes to individual sources or add/remove sources. However, the lack of consistent terminology makes it difficult to compare ontologies from different sources. To address this challenge, additional representational formalisms have been proposed to define inter-ontology mapping. These formalisms aim to explore linguistically related concepts from multiple ontologies and provide a solution for comparing and aligning ontologies from different sources.[12, 13]

Hybrid ontologies have been developed to address the limitations of single and multiple ontology approaches. These hybrid technologies combine the use of multiple ontologies with a single global shared vocabulary, allowing for the semantics of each source to be described by its own ontology. This approach enables the source ontologies to be comparable and facilitates interoperability between different systems and domains. The use of a shared vocabulary helps to overcome the challenges of integrating diverse ontologies and promotes consistency and coherence in hybrid intelligent systems[14]

The task of ontology learning is divided into four sub-tasks: ontology representation, ontology acquisition, ontology evaluation, and ontology maintenance [15].

2.1. Ontology Representation

Before we start working on ontology, we need to know what kind of representation we want to use. This decision is important not only for human and computer understanding, but also for ontology efficiency. Indeed, a good representation improves the expression of the thinking system.

The standardization of ontology representation languages has sparked a lot of interest in OL in recent years.

2.2. The acquisition of ontologies

After choosing a representation, it is important to extract and collect relevant content to characterize a particular domain in order to construct an ontology. This material typically consists of concepts, properties, and relationships, and can be manually developed by subject matter professionals or extracted semi-automatically from a corpus using Natural Language Processing (NLP) techniques. The generated/extracted content must then be translated into an ontology-compatible format. The main processes of ontology acquisition are summarized as follows:

1. Information extraction: information about a certain domain must be recognized and extracted from a variety of sources (e.g., web sites, corpora, books, dictionaries, thesauri, etc.) and converted into a suitable format for ontology discovery.
2. Ontology discovery: once the data is collected, the irrelevant data must be removed and the relevant data must be added to the ontology.
3. Ontology organization: once the information has been added to the ontology, it must be arranged in order for the knowledge to be more usable. Synonyms can be clustered, inverse relations can be deduced, centroids can be discovered, taxonomies can be created, and so on.

2.3. Evaluation of ontologies

Once the ontology is complete, its performance must be evaluated. Likewise, this process can be done manually or semi-automatically. The latter technique is used frequently because it saves money and enables comparison of the performance of different ontologies.

It satisfies the three basic criteria for gold-standard-based assessments: we allow assessment along multiple dimensions, (ii) the distance between the correct and given answer is considered, and (iii) the scale interval is used equally, in accordance with analytical and empirical assessment. [16].

Evaluation of the quality of ontology acquisition is an important aspect of intelligent web technology because it allows researchers and practitioners to assess the validity of the resulting ontology at the lexical level, coverage at the concept level, validity at the taxonomic level, and adequacy at the non-lexical level. In the event that ontology does not meet the user's unique requirements, evaluation of ontology acquisition allows for improvement and reconfiguration of the entire ontology learning process. As mentioned earlier, ontology learning is a multi-level process, which makes the evaluation of ontology extraction difficult. Due to the difficulty in evaluating domain ontology, a large number of evaluation strategies have been introduced in the past few years, and this field is in continuous development. All proposed methodologies fall into one of these categories, which are categorized according to the type of target ontology and the objective of the assessment.

- Gold standard-based assessment
- Application based evaluation
- Data-driven assessment
- Human evaluation [17]

2.4. Ontology maintenance

Existent ontologies must be updated on basis since knowledge is constantly changing. Maintenance can also be done manually or semi-automatically, with the latter being the recommended method [16].

3. PROPOSED WORK

There are many ways to extract keyword from text or even from images using a specific algorithm. In this case, there is no way that there will be errors in extracting the information, even if it is from 100 documents or from a number of sites.

Information extraction is a process that brings data in order to obtain information about a specific field, and information extraction can be from text, image, and audio. We collect information from various websites: wikipedia. org, rspb. org. uk and garden-birds.co.uk. To build an ontology several steps are performed as shown in Figure 1 and described as follows:



Fig. 1. Snippet of classes and subclasses added

First step is adding categories and subcategories as shown in Figure 2: We define a category for Birds and Habitat, and four more categories: City, Coast, Countryside, and Nature Reserve which are subcategories of the Habitats category. The second step is to add the characteristics and sub-characteristics and their domain and scope as shown in Figure 3. The third step is to add the affirmations which serve as a description of each bird as shown in Figure 4.

```

<!-- Classes and Sub Classes -->
<rdfs:Class rdf:about="&ex;bird"/>
<rdfs:Class rdf:about="&ex;habitat"/>
<rdfs:Class rdf:about="&ex;town">
<rdfs:subClassof rdf:resource="&ex;habitat"/>
</rdfs:Class>
<rdfs:Class rdf:about="&ex;contryside">
<rdfs:subClassof rdf:resource="&ex;habitat"/>
</rdfs:Class>
<rdfs:Class rdf:about="&ex;Nature_reserves">
<rdfs:subClassof rdf:resource="&ex;habitat"/>
</rdfs:Class>
<rdfs:Class rdf:about="&ex;Coastal">
<rdfs:subClassof rdf:resource="&ex;habitat"/>
</rdfs:Class>

```

Fig. 2. Snippet of classes and subclasses added

```

<!--properties and sub Properties-->
<rdf:Property rdf:about="&ex;habitatIn">
<rdfs:domain rdf:resource="&ex;bird"/>
<rdfs:range rdf:resource="&ex;habitat"/>
</rdf:Property>
<rdf:Property rdf:about="&ex;length">
<rdfs:domain rdf:resource="&ex;bird"/>
<rdfs:range rdf:resource="http://www.w3.org/2000/XMLSchema#integer"/>
</rdf:Property>
<rdf:Property rdf:about="&ex;sound">
</rdf:Property>
<rdf:Property rdf:about="&ex;beak">
</rdf:Property>
<rdf:Property rdf:about="&ex;familyname">
<rdfs:domain rdf:resource="&ex;bird"/>
<rdfs:range rdf:resource="http://www.w3.org/2000/XMLSchema#string"/>
</rdf:Property>

```


Fig. 3 Snippet of properties and sub properties added

```
<rdf:Description rdf:about="&ex;bird1">
<rdf:type rdf:resource="&ex;bird"/>
<ex:name>Arctic skua</ex:name>
<ex:Familyname>Skuas</ex:Familyname>
<ex:habitatIn rdf:resource="&ex;Coastal"/>
<ex:eat >
    <rdf:Bag>
        <rdf:li>fish</rdf:li>
        <rdf:li>eggs</rdf:li>
        <rdf:li>birds</rdf:li>
    </rdf:Bag>
</ex:eat>
<ex:length rdf:parseType="Resource">
<ex:length_min>16in</ex:length_min>
<ex:length_max>19in</ex:length_max>
</ex:length>
```

Fig. 4 Snippet of assertions

In order to support semantic web applications, SPARQL is the W3C's candidate recommendation for accessing RDF/OWL data graphs. We use SPARQL to execute queries after converting the raw semantic data into RDF/OWL format. We apply SPARQL query shown in Figure 5 by Open Link website.

```
Default Graph URI
http://                               /bird/new.xml

Query Text
prefix ex: <http://www.rspb.org.uk/>
select ?Colourbars ?bird
where
{
  ?n ex:Colourbars ?Colourbars.
  ?n ex:name ?bird.
  filter regex(?Colourbars,"red").
}
```

Fig. 5 An example of SPARQL query to find the bird's information

The results of the query is shown below in Figure 6.

Colourbars	bird
"red"	"Linnet"
"red"	"Bullfinch"
"red"	"Limosa lapponica"
"red"	"Calidris ferruginea"
"red"	"Chaffinch"

Fig. 6 Results of SPARQL query to extract bird's information

4. Conclusion

In conclusion, we have successfully developed an ontology architecture for bird classification using XML/RDF standards for semantic web interoperability. Our ontological approach provides a comprehensive, structured classification of birds, including their characteristics, habitats, and relationships. The effectiveness of an ontology is evaluated by its ability to make possible classification of diverse information. This ontology of birds need extra development to be a valuable resource for researchers and conservationists. As a future work, an integration of the developed ontology into an expert system for bird classification will be conducted. The knowledge representation scheme will be combined with inference mechanisms and decision-making technique to make decisions in the domain of bird classification.

References

- [1] Ž. Turk, "Construction informatics: Definition and ontology," *Advanced engineering informatics*, vol. 20, no. 2, pp. 187-199, 2006.
- [2] M. Neji, F. Ghorbel, B. Gargouri, N. Mimouni, and E. Metais, "An Ontology based Smart Management of Linguistic Knowledge," *Journal of Data Mining & Digital Humanities*, 2022.
- [3] U. Bharambe, C. Narvekar, and P. Andugula, "6 Ontology and Knowledge Graphs for Semantic Analysis in Natural Language," *Graph Learning and Network Science for Natural Language Processing*, p. 105, 2022.
- [4] H. Sebbag, N.-e. el Faddouli, and S. Bennani, "Recommender System to Support MOOCs Teachers: Framework based on Ontology and Linked Data," in *Proceedings of the 13th International Conference on Intelligent Systems: Theories and Applications*, 2020, pp. 1-7.
- [5] M. Shamsfard and A. A. Barforoush, "Learning ontologies from natural language texts," *International journal of human-computer studies*, vol. 60, no. 1, pp. 17-63, 2004.
- [6] H.-T. Nguyen, V.-H. Nguyen, and V.-A. Vu, "A knowledge representation for Vietnamese legal document system," in *2017 9th international conference on knowledge and systems engineering (kse)*, 2017: IEEE, pp. 30-35.
- [7] C.-R. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, and L. Prevot, *Ontology and the lexicon: A natural language processing perspective*. Cambridge University Press, 2010.
- [8] M. Hazman, S. R. El-Beltagy, and A. Rafea, "A survey of ontology learning approaches," *International Journal of Computer Applications*, vol. 22, no. 9, pp. 36-43, 2011.
- [9] P. Buitelaar, P. Cimiano, and B. Magnini, "Ontology learning from text: An overview," *Ontology learning from text: Methods, evaluation and applications*, vol. 123, 2005.
- [10] T. Ivanova, "Ontology learning technologies-brief survey, trends and problems," in *proceedings of the International Conference on Information Technologies*, 2012, pp. 245-255.
- [11] 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 ..., 2001.
- [12] X. Guo, A. Berrill, A. Kulkarni, K. Belezko, and M. Luo, "Merging ontologies algebraically," *arXiv preprint arXiv:2208.08715*, 2022.
- [13] V. M. Tayur and R. Suchithra, "Multi-Ontology Mapping for Internet of Things (MOMI)," in *Prediction and Analysis for Knowledge Representation and Machine Learning*: Chapman and Hall/CRC, 2022, pp. 125-141.

- [14] S. F. Pileggi, "Ontology in Hybrid Intelligence: a concise literature review," *arXiv preprint arXiv:2303.17262*, 2023.
- [15] L. Zhou, "Ontology learning: state of the art and open issues," *Information Technology and Management*, vol. 8, no. 3, pp. 241-252, 2007.
- [16] K. Dellschaft and S. Staab, "Strategies for the Evaluation of Ontology Learning," *Ontology Learning and Population*, vol. 167, pp. 253-272, 2008.
- [17] M. N. Asim, M. Wasim, M. U. G. Khan, W. Mahmood, and H. M. Abbasi, "A survey of ontology learning techniques and applications," *Database*, vol. 2018, 2018.