Evaluating the Quality of Authoritative Linked Data Models

Authors:

Hanan Muhajab Dr Alia Abdelmoty

Cardiff University

Presenter:

Hanan Muhajab

Muhajabh@Cardiff.ac.uk

GEOProcessing 2023

The Fifteenth International Conference on Advanced Geographic Information Systems, Applications, and Services.

About Me

- I am Hanan Muhajab, a PhD student specializing in the development of computational models that incorporate artificial intelligence and knowledge graphs. Currently, I am working on geographic knowledge graphs and linked data on the web.
- Attained a Bachelor of Science degree in Computer Science from Taif University, Saudi Arabia. Then, I graduated from Kent State University, Ohio USA, with a master's degree in Computer Science - thesis track. During 2017, I received a Golden Key membership from Kent State University, Ohio, USA. In 2019, I earned a Master's degree in Data Science at Kent State University, Ohio, USA.
- Work as a lecturer at Jazan University, Saudi Arabia.

Introduction

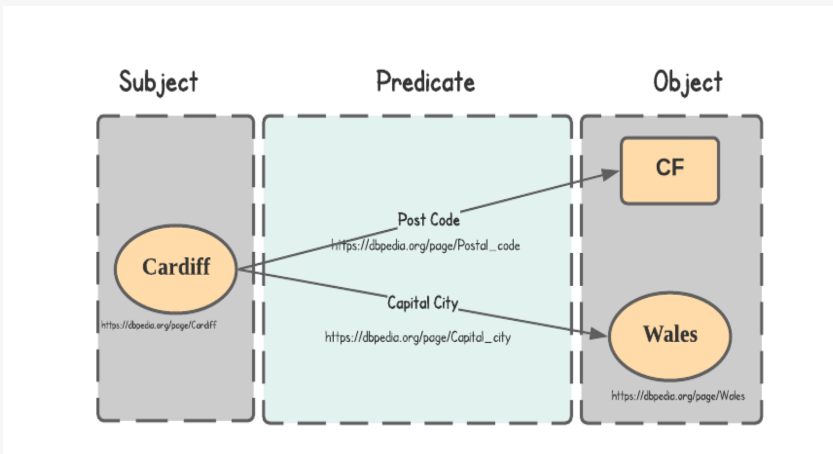
- Linked Data

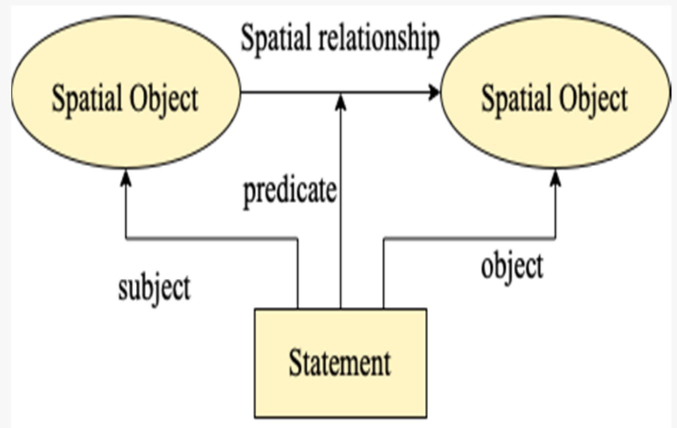
Large scale integration of, and reasoning on, data on the Web.

It organizes knowledge with a set of concepts, relations, and facts, which are associated by two types {entity, relation, entity} and {entity, attribute, attribute value}

Semantic Web and Linked Data technologies have been considered for the representation and sharing of authoritative geographic data sets.

- Example of semantic web data - RDF triples





Triple statement representation

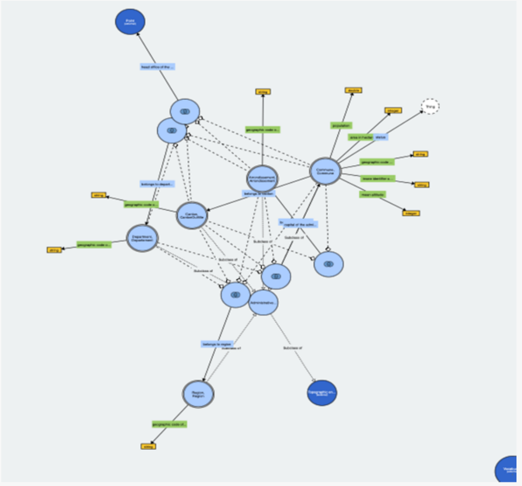
Core Semantic Web Technologies

- **Uniform Resource Identifiers (URIs)** used to identify the concepts (people, places, things, abstract..) and properties (data relationships)
- **Resource Description Framework (RDF)** provides a W3C standard way to describe logical statements about relationships.
- **Ontologies** are like data dictionaries with additional logical annotations (to say how properties and resources are related).

  "An ontology is an explicit specification of conceptualization."
- **SPARQL** query language enables a query to combine machine-readable data from multiple sources and also allows **new** data relationships to be constructed (*inferred*) from existing data.

Authoritative Geographic linked data

- Authoritative Geographic Linked Data is open government data to provide the most accurate information.
- For example, the Ordnance Survey, the mapping agency of Great Britain:
- Five defined ontologies
- Open data sets of approximately 64,342,201 triples.
- Norway, Germany, Ireland, and Spain are published, and presented at the Knowledge Graph in Action conference (KGiA) [2].



osi Ordnance Survey Ireland

Ordnance Survey

Ontology Evaluation

- The process of deciding on the quality of an ontology in respect to a particular criterion.
- Ontology Evaluation is classified into two concepts:

Validation

Verification

- Evaluation perspective:

Provider

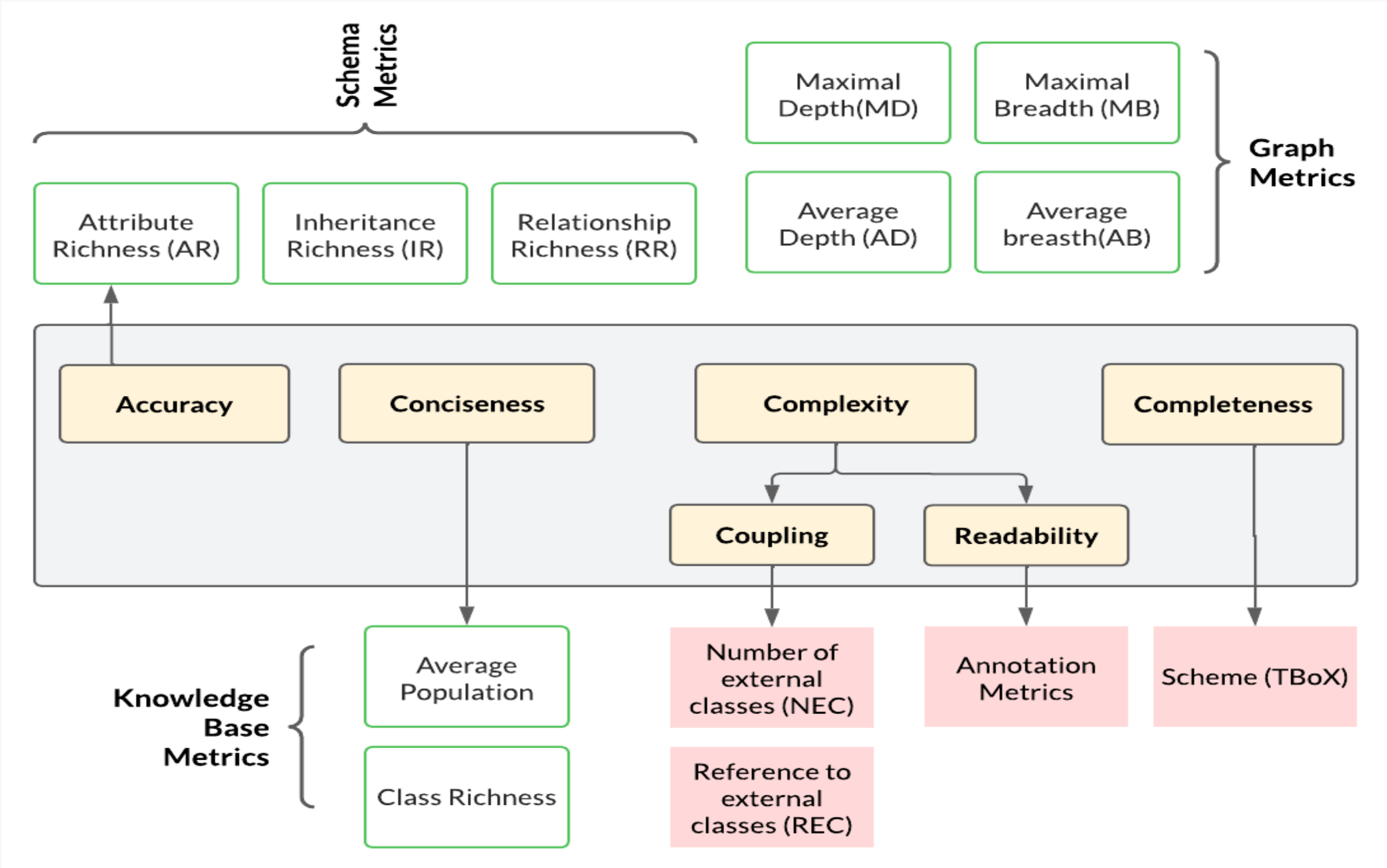
Consumer

- Debattista et al. [3] evaluated Ordnance Survey Ireland (OSI) using the Luzzu and OOPS platforms.
- There is a need to assess the authoritative geographic ontologies from the **user's perspective** to better facilitate their understanding and reuse, as recommended in KGiA .

Dataset

Index	Weblink	Classes	Individuals
O_1	https://data.ordnancesurvey.co.uk/ontology/ [accessed: 2023-03-03]	53	2021346
O_2	https://triplydb.com/osi/adminitrative-units [accessed: 2023-03-03]	18	659333
O_3	http://linkedopendata.gr/dataset. [accessed: 2023-03-03]	9	2914
O_4	http://data.ign.fr/def/geofla/20190212.en.htm [accessed: 2023-03-03]	8	132567

Method



Schema Metrics
Maximal Depth(MD)
Maximal Breadth (MB)
Graph Metrics
Attribute Richness (AR)
Inheritance Richness (IR)
Relationship Richness (RR)
Average Depth (AD)
Average breasth(AB)
Accuracy
Conciseness
Complexity
Completeness
Coupling
Readability
Knowledge Base Metrics
Average Population
Class Richness
Number of external classes (NEC)
Reference to external classes (REC)
Annotation Metrics
Scheme (TBoX)

Accuracy

- The accuracy criterion measures the extent to which an ontology models its real-world domain.
- Attribute Richness (AR): indicates the number of attributes (slots) defined for each class, which can be used to infer the quality of the ontology design.

$$AR = \frac{|ATT|}{|C|}$$

Number of attributes for all classes(AT T)
Number of classes C

- Inheritance Richness (IR): shows the distribution of information across different levels of ontology.
- It defined as the average number of subclasses per class

$$IR = \frac{\sum_{C_i \in C} |H^C(C_1, C_i)|}{|C|}$$

- Relationship Richness (RR): indicates the diversity of ontology relationships.

$$RR = \frac{|P|}{|H| + |P|}$$

Number of relationships(P)
Number of subclasses (H)

- Graph metrics include Average Depth (AD), Average Breadth (AB), Maximal Depth (MD), and Maximal Breadth (MB).

Conciseness

- The conciseness criterion measures the degree of usefulness of the ontology knowledge.
- This quality criterion correlates with Average Population (AP) and Class Richness (CR).
- AP represents the average distribution of instances across all classes.

$$CR = \frac{|U|}{|C|}$$

Number of classes used in the base (U)
Number of classes(C)

- CR is a measure of how instances are distributed among classes. Therefore, it indicates how many instances are related to the classes defined in the schema.

$$P = \frac{|I|}{|C|}$$

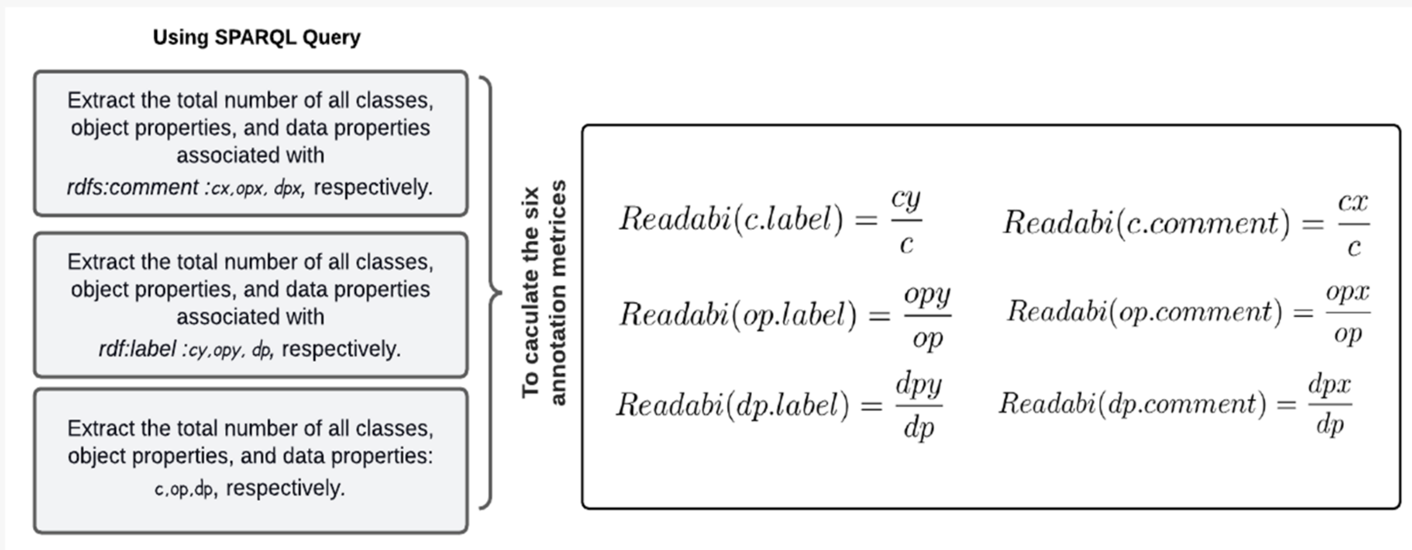
Number of instances of the KB(I)
Number of classes defined in the ontology schema(C)

Complexity

- **Readability**

  A measure of readability is the average number of names (labels) and descriptions (comments) per ontology entity, such as classes and properties.

1)Extract all the classes *C* associated with rdf:label and rdfs:comment.
2)Extract all the object properties *OP* associated with rdf:label and rdfs:comment.
3)Extract all the data properties *DP* associated with rdf:label and rdfs:comment.



GraphDB and ***SPARQL*** queries to compute the annotation metrics.

Coupling

- Coupling reveals the number of external classes from imported ontologies referenced in the local ontology.
- It measures the relatedness between the local ontology and other existing ontologies or vocabularies used to construct the ontology.

$$Coupling(O) = REC/NEC$$

NEC is the distinct number of external classes,
REC is the number of references to external classes

- By parsing the OWL file, we calculated the number of distinct external classes defined in the ontology and the number of references to external classes.

- The code is available online [8].

Completeness

- Assessment of completeness considered the schema level and not the instance level of representation.
- Spatial completeness of the ontologies was done by considering the standard set of possible spatial relationships between data types.
- A completeness score for the ontology is computed in terms of the completeness score of its spatial classes.
- Each class in the ontology is checked for completeness, and then equation is used to compute the result.

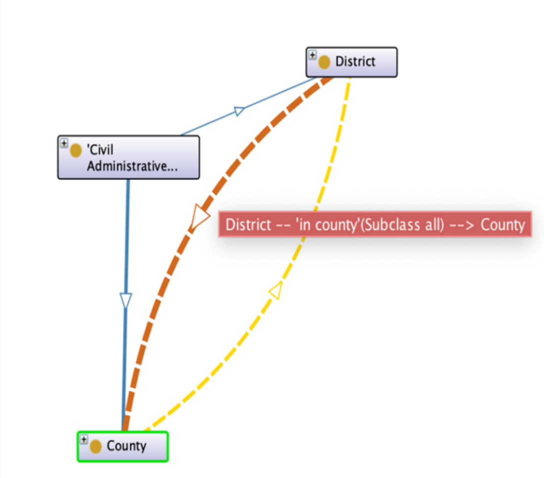
$$Completeness = Comp/C$$

where Comp is the sum of the completeness score of all the spatial classes

and C is the total number of spatial classes in the ontology.

TABLE III: The Spatial Relationships and the Geometry Types

Spatial Relation	Equals	Disjoint	Touches	Within	Contains	Overlaps
$P \leftrightarrow P$	✓	✓		✓	✓	
$P \leftrightarrow L$		✓	✓	✓	✓	
$P \leftrightarrow Po.$		✓	✓	✓	✓	
$L \leftrightarrow L$	✓	✓	✓	✓	✓	✓
$L \leftrightarrow Po.$		✓	✓	✓	✓	
$Po. \leftrightarrow Po.$	✓	✓	✓	✓	✓	✓



Results

Index	Weblink	Classes	Individuals	AR	IR	RR	AD	MD	AB	MB	AP	CR
O_1	https://data.ordnancesurvey.co.uk/ontology/ [accessed: 2023-03-03]	53	2021346	0.321	5.35	0.11	2.38	3	4.33	8	38138.60	0.339
O_2	https://triplydb.com/osi/adminitrative-units [accessed: 2023-03-03]	18	659333	0	0.93	0.166	1.93	2	8	15	36629.61	0.777
O_3	http://linkedopendata.gr/dataset. [accessed: 2023-03-03]	9	2914	0.444	0.88	0.272	1.88	2	4.5	8	323.77	0
O_4	http://data.ign.fr/def/geofla/20190212.en.htm [accessed: 2023-03-03]	8	132567	0.409	0.5	0.56	1.54	3	5.5	15	6025.77	0.409

The administrative units for the UK (O1), Ireland (O2), Greece (O3), and France (O4).

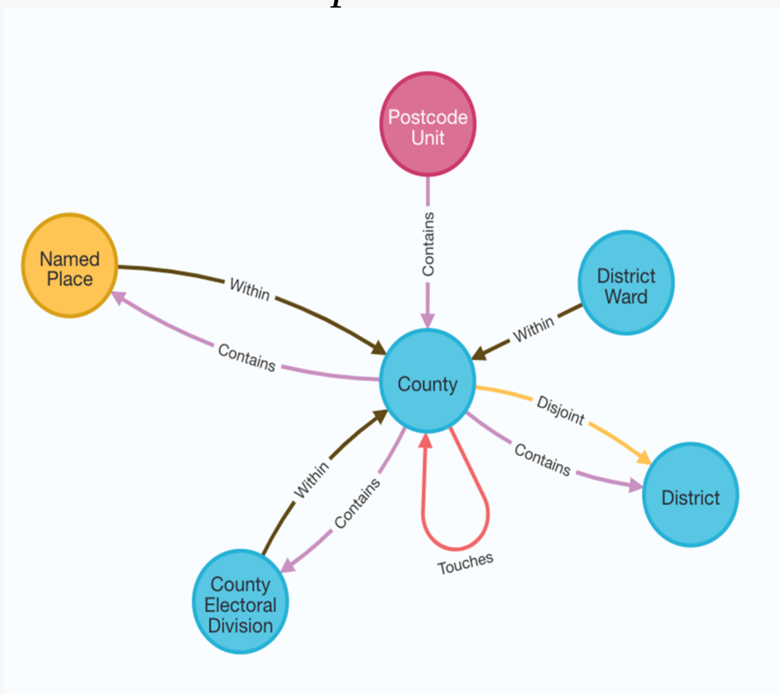
Index	Coupling	Readability						Completeness
		c.comment	*c.label*	*op.comment*	*op.label*	*dp.comment*	*dp.label*	
O_1	0	0.75	1	0.84	0.84	0.6	0.8	0.56
O_2	15	0.93	1	1	1	0	0	0.3
O_3	8	0	1	0	0	0	1	0.3
O_4	0	0.75	1	0.85	0.85	1	1	0.3

Discussion

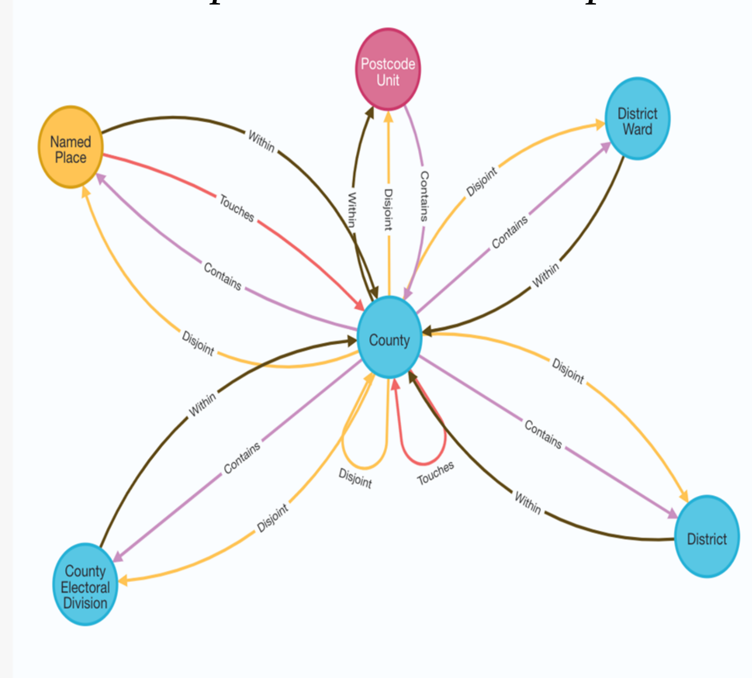
- The O2, O3, and O4 cover more specific details (depth), while O1 defines the domain broadly.
- Ontologies O1, O2 and O3 have a low RR score, as they represent mostly one type of relationship; namely the subclass relationship.
- In addition, Attribute Richness (AR) values indicate that the ontologies O1, O3 and O4 contain more attribute information about the classes than O2.
- The O1, O2 have a large number of instance per class, indicating a good fit for the class representation in the ontology.
- The results of the CR indicate that O1 , O2 , and O4 have more instances than O3.
- As shown in Table II, O2 and O3 have high complexity due to the coupling and readability values.
- The result indicates hat O2 has a strong coupling, which makes it more difficult to understand and maintain than O1.

Completeness Graph

Incompleteness



All possible relationships



Results show that O1 is 56% complete, O2, O3, and O4 are 30%,
making O1 more capable of reasoning and retrieving the geographic information.

CONCLUSION

- We evaluated authoritative geographic ontologies using metrics-based methods. Analysis of metrics result indicates that geographic ontologies contain enough data to facilitate knowledge usage.
- Results confirm that Uk ontology covers a wide range of information and show that the ontologies have a good hierarchy.
- A high score for incomplete spatial relationships leads to fewer inferred geographical details in France, Greece, and Ireland.
- The UK ontology has very low complexity, which indicates that the model is easy to understand by the user.
- The future research direction would be to develop a unified data model to integrate the authoritative ontologies

References

1] J. Goodwin, C. Dolbear, and G. Hart, “Geographical Linked Data: Theadministrative geography of Great Britain on the semantic web,” Trans-actions in GIS, vol. 12, pp. 19–30, 2008.

[2] Bucher, et al. “Spatial linkeddata in Europe:Report from spa-tiallinked data sessionat knowledge graphin action,” Official Publication-EuroSDR 73, 2021.

[3] J. Debattista,´E. Clinton, and R. Brennan. “Assessing thequal-ity of geospatial linked data–experiences from ordnance surveyire-land (osi),”, 2018.

[4] M. Poveda-Villal´on, A. G´omez-P´erez, and M. C. Su´arez-Figueroa,“Oops! (ontology pitfall scanner!),” International Journal on SemanticWeb and Information Systems, vol. 10, no. 2, pp. 7–34, 2014.

[5] B. Lantow, “OntoMetrics: Putting metrics into use for ontology evalua-tion,” Proceedings of the 8th International Joint Conference on Knowl-edge Discovery, Knowledge Engineering and Knowledge Management,2016.

[6] “Ontotext graphdb,” Ontotext, 13-Jan-2023. [Online]. Available:https://www.ontotext.com/products/graphdb/. [Accessed: 04-Mar-2023].

[7] L. Ouyang, B. Zou, M. Qu, and C. Zhang, “A method of ontologyevaluation based on coverage, cohesion and Coupling,” 2011 EighthInternational Conference on Fuzzy Systems and Knowledge Discovery(FSKD), 2011.

[8] H. Muhajab, “HMUHAJAB/coupling,” GitHub. [Online]. Available:https://github.com/hmuhajab/Coupling. [Accessed: 06-Mar-2023]


THANK
YOU