

Ontology based Multimodal semantic Transformation method for Biomedical data

Sanjib Pitchaipillai, Indadul Nayar

State Key Laboratory of Bioelectronics, School of biological science and medical engineering, Cartagena University

Abstract: An ontology-driven data transformation model is used to build a biomedical dataset integration database, and a shared semantic identifier for heterogeneous resources is given. Supports semantic Web lattice biomedical library, can realize the automatic transformation and mapping of data sets.

Keywords: Semantic Noumenon; Biological Medicine; enantiomorphism; Multi-modal

Introduction

The semantic Web can provide the data format recognized by the computer. This body is responsible for the formal representation and semantic interpretation of the data, and shares and reuses the structural description of a particular application field. Check the logical consistency of data content. Biomedical databases based on semantic Web, such as BioPortal^[1], contain more than 500 biomedical ontologies, vocabulary, Domain terminology. Open semantic data sets in the face semantic Web community are represented in semantic formats such as rdf or owl, but are widely distributed and loosely distributed. It is necessary to standardize the semantic format of biomedical data set to reduce the space-time cost of data integration and interoperability.

In the field of biomedical database, the mature gene ontology, such as the UniProt platform developed by the European Institute of Biological Information, is used to use the ontology based data semantic annotation model. A biomedical resource model based on RDF has been developed by using ontology logical locus for data annotation modeling. Bio2rdf^[3]. In the area of electronic health archives, semantic health projects use domain ontology-based electronic medical records as the modeling language for electronic health files, and implement semantic interoperability. These studies do not solve the problem of logical error detection in the field of biomedical databases, For the cross-semantic data fusion process, there is a lack of dynamic adjustment ability. Based on the criterion of biomedical data, the authors intend to establish a data access and integration framework, and dynamically define the process of data mapping and integration by using the multi-modal of ontology transformation. Past data

Integrate and interoperate the data conversion model of heterogeneous data sources.

1. Normalization of biomedical data

1.1 Biomedical data preprocessing

The widely used biomedical databases are usually stored in relational databases in XML format, just as the source data representation tools orthoxml and seqxml^[4]. There is a lack of necessary semantic connection between data, so it is necessary to fuse biomedical information from different data sources to achieve differences. Interoperability of construction information.

Copyright © 2018 Sanjib Pitchaipillai *et al.*

doi: 10.18063/bc.v2i2.

This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License

(<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Clinical models are used to specify clinical record information and capture clinical data (laboratory tests, blood pressure measurements, and blood pressure measurements) by constrained information model structures.

Pharmacotherapy, etc., is stored and exchanged in the form of Prototype Definition Language (ADL)^[5], providing an extensible clinical data sharing platform.

Biomedical databases contain a large number of complex and dynamic biomedical entity information, such as protein information widely existing in different databases. Because biomedical entities in heterogeneous resources can not share information identifiers with similar semantics, there is a huge semantic gap between data sharing and fusion. The application of knowledge engineering theory and technology largely depends on the acquisition, discovery and fusion of medical knowledge, which requires such processes as the description framework of medical domain knowledge, the transformation between heterogeneous information sources, the conceptual modeling of domain knowledge, the resolution and orientation of conflict domains among multiple sources of knowledge. The key technologies of knowledge discovery such as Semantic Discovery support.

1.2 Access and integration of biomedical data

Ontology-based Data Access (OBDA) allows the use of traditional data formats based on semantic technology. Data access is performed by separating the user from the original data source. Ontology provides user-oriented data information for access. XML data and relational data. Separate semantic information is processed by SPARQL query language. Semantic links between XML and RDF are established to realize data conversion and information query.

In relational database, OBDA method differs greatly in mapping relation expression, query language translation and reasoning ability because of different systems. OBDA supports semantic reasoning and uses logical bits to solve semantic conflicts. DA method is used to meet the needs of open data integration. Data warehouse-oriented integration method is adopted to integrate data semantic link function, and external data sets are constructed by defining data transformation rules and mapping rules.

Data integration overcomes redundancy or inconsistency among data sources and provides more data control with a data warehouse-oriented approach. Data sets (XML and relational data) are created in a common data warehouse. In order to preserve raw data information while integrating semantic knowledge bases, data integration processes To merge or link instances with equivalent semantic relations.

2. The above modal ontology conversion mode

In order to get a data set with rich semantic information, we need to change the way the data type is transformed to meet two requirements: (1) the rules defined can handle the relationship between multiple inputs and multiple ontology classes; (2) allow additional information to be added to make the input data rich in semantics. Source.^[8] For this purpose, a multi-modal ontology data transformation pattern involving partial or complete ontology class semantic description is adopted. From the point of view of mapping definition, the object of the schema transformation is mainly XML class entities and ontology classes for easier implementation of semantic links. Data genera are designed by using OWL ontology classes. A generic template for properties, object attributes, and constraints that binds entities, attributes, or relationships to corresponding instances, classes, and relational variables.

A transformation schema can be defined as a binary? $S, V?$, where S represents class and numeric attributes, object attributes, all instances in the OWL ontology and their subsets, and V is a set of variables associated with class instances in S . A schema is a set of attribute variables that contain an instance of an input schema entity by joining them together. Together, new semantic entities can be created in the OWL ontology to further enrich the ontology semantics. The schema has the functions of allowing the creation of new instances, assigning values to numeric attributes, connecting instance sets through object attributes, and providing specifications for specific content without depending on input data sets.

2.1 Data mapping modes

The mapping submodel between OrthoXML data and OWL ontology is the basis of data conversion and integra-

tion. It is also closely related to the corresponding mode of ontology transformation. The overall operating quality of the system will be greatly affected.^[9] To minimize the dependence of the mapping process on rules, we create an in-body spreadsheet. Semi-automatic mapping of core elements (real, generic, and relational) of OrthoXML data using ontology assistive technology To reduce mapping time and improve mapping quality. Mapping different types of data models to OWL ontology, there is no standard definition language. This model is based on formal ontology alignment technology. The relational database and the XML data schema are mapped to the OWL ontology. If the same OWL ontology is used in the conversion process, the mapping can be repeated.

The rules of semantic mapping are established to construct the OWL instance from the relational number database and the XML number data module. When the content of the OWL ontology is limited or not enough to complete the mapping input mode, The created ontology class can effectively complement the semantic information. In order to generate ontology class method by semi-automatization, it is necessary to bind the variable set associated with the schema to the ontology class instead of the instance set. Mapping rules may be based on ontology instances or classes. In fact, the same semantics and different representations of data are allowed in this model Because OWL-DL may define different uses of the same operative language, an instance or class can have the same URI, The same data base contains information about genes, proteins, and biological expression data. In the database, they are represented as examples. At the semantic level, it corresponds to classes.

Figure 1 illustrates the mapping relationship between the XML schema (left) of molecular information and the molecular domain ontology (right). Ontology class molecules and atoms are associated with the XML schema through mapping, but there are no classes in the ontology that represent chiral molecules, so the related content is added to meet the needs of semantic mapping. In OWL, a chiral molecule is defined as a Molecule-like Chirality with an attribute relationship has_chemical_property. A chiral molecule is represented by an attribute name isChiral whose attribute value is defined in Val with a value of 0 or 1. Define variables in opp1217? "Chiralmolecule" and generate the corresponding ontology axiom code, the code is: ? ChiralMolecule:Individual Begin Add? Molecule instance of (Molecule and has_chemical_properties of some Chirality) End

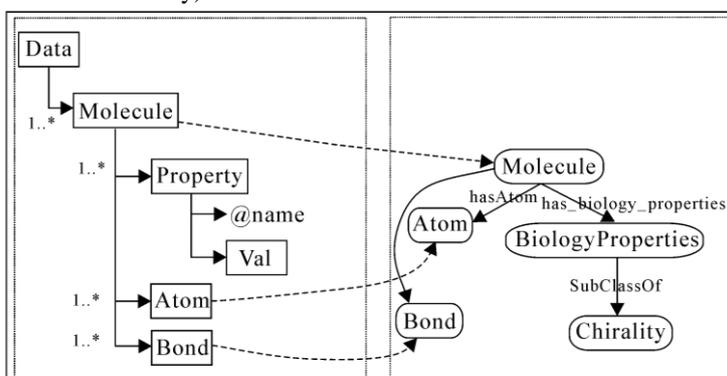


Figure 1; Mapping relationship between XMLSchema and molecular domain ontology

On the basic reflection gauge: the definite meaning of the function of the General Real body is entity_rule (molecule,) On_condition (property ("@ name =" is- chiral ") / r value 1). When the number of ischiral is 1, The Molecule class of molecule and OWL has an enantiomeric relation. The instance information from OWL is incomplete and does not contain chiral attribute information because of There is a semantic inconsistency between this input instance and the ontology individual. To solve this problem, This schema allows you to define the mapping relationship between variables? "chiralmolecule" and additional information.

2.2 Data conversion modes

The data transformation submodel checks all the generated formal ontologies to ensure that semantically consistent data content is generated, The use case and extensible numeral-data interface in which the semantic relation is relatively unique. The logic of using OWL to protect the internal capacity of the evidence. O WLDL, Output RDF or OWL format data. If inconsistent content is found in the mapping data, these contents are automatically discarded and corresponding

ontology classes are added to the mapping result set from the source body.

The semantic representation of the semantic transformation data needs to be explicitly annotated in the XML schema or in the corresponding data table, The computational complexity of semantic transformations depends on the number of mapping instances, attributes, and relationships. Because the number of axiomatic instances to be generated is large, the conversion time is generally longer. The transition time is relatively short. This model introduces conditional judgment by using exchanged data sets, that is, data satisfying the rule of identity. If two XML input entities are mapped to the same ontology class, there is no need to automatically infer to generate the mapping data set, omitting the relevant types of axiomatic instances. Reduce the time and space complexity of model operation.

3. Data integration and interoperability

The data integration sub-model integrates data from different data sources. The main goal is to detect equivalent data instances to reduce redundancy, to ensure that the resulting repository has logical and semantic consistency and supports OWL ontology reasoning.

In the process of integration, identity rule is the basis to reduce data redundancy, pass through instance attributes, identify and deal with ontology with the same semantics.

Class instances. For example, if protein attribute information is combined, different databases may use different identifiers (mainly URI attributes of instances) to distinguish protein instances from different data sources. These resources may use human genome nomenclature to name the related gene sequences of the protein. If the gene name is the same Sex rules for protein identification can minimize the occurrence of protein heteronyms.

The goal of interoperability is to produce semantic link relationships through the URI of a protein. Take protein resources as an example, assign different native URI, to each protein entity The goal is to integrate proteins from two resources into the same protein identifier. The number is converted and assembled on the basis of a OWL ontology that contains a protein class. Generally speaking, The semantic interconnection described above may occur as follows: (1) the properties of egg white matter have the same values as those that satisfy the identity rule in two resources; (2) the protein class and the attribute satisfying the identity rule have the same values in two kinds of resources. (3) proteins do not have association rules.

Under the condition of case (1), after data transformation, there will be two kinds of existence modes between entities with semantic relations: when owl: sameAs relationship exists between two instances, the protein markers will be retained, and whether the semantic chains are merged or annotated will be further confirmed by identity rules. Then, when two instances are merged, only one URI resource can be retained by merging instances with different URIs. In this case, one of the "identifiers + prefixes" will be selected to distinguish different instance information. If this happens (2) or (3), The mapping rules allow input data to be converted into RDF or OWL formats that support semantic Web and associated open data clouds, enabling data integration and interoperability.

4. Conclusion

At present, in the field of biomedical semantic modeling, transformation methods such as regular transformation-based relational database, XML, RDF and OWL language are widely used to perform syntax conversion in traditional formats, which makes it difficult to store the acquired data sets in semantic interoperability and data fusion. The mapping modeling tool of domain ontology knowledge uses the domain knowledge of OWL ontology as the data source corpus to guide the semantic representation of the input data set and construct an ontology-driven biomedical semantic data model. In order to solve the problem of dynamic data fusion, we mainly deal with the data mapping model with complex heterogeneous morphology.

References

1. NOYNATALYAF, SHAHNIGAMH, WHETZELPATRICIAL, *et al.* BioPortal: Ontologies and Integrated Data-re-sources at The Click of a Mouse [J]. *Nucleic Acids Research* 2009; 37:170-173.
2. SHAHNIGAMH, TENENBA UMJESSICAD. The Coming Ageof Data-Driven Medicine: Translational Bioinformatics 'Next Frontier [J]. *Journal of the American Medical Information Association* 2016; 19(e1): 2-4.

3. GALPERINMICHAELY, RIGDENDANIELJ, FERN NDEZ-SU REZXOS M. The 2015 Nucleic Acids Research Database Issue and Molecular Biology Database Collection [J]. Nucleic Acids Res.
4. BERNSTEINFRANCESC, KOETZLETHOMASF, WILLIAMSGRAHEMEJB, *et al.* The Protein Data Bank. A Computer - Based Arch Ival File for Macromolecular Structures [J]. European Journal of Biochemistry 2013; 80(2): 319-324.
5. JUPPSIMON, MALONE JAMES, BOLLEMANJERVEN, *et al.* The EBIRD FPlatform: Linked Open data for the Life Sciences [J]. Bioinformatics 2014; 30(9): 1338-1339.
6. Cai Lihong, Ma Jing. An Ontology Mapping Experiment Based on Synthesis Method [J]. Journal of Information Science 2010; 29(5): 820-825.
7. MART NEZ-COSTAC, SCHULZS. Ontology Content Patternsas Bridge for the Semantic Representation of ClinicalIn- Formation [J]. Applied Clinic Information 2014; 5(3): 660-679.
8. Teng Guangqing, Bi Qiang. Thinking of what should be and what should be: Ontology and Ontological Process of Knowledge [J]. Information Theory and Practice 2011; 34(12): 24-28.
9. FERN_NDEZ-BREISJT, LEGAZ-GARC_AMDC, CHIBAH, *et al.* Towards the Semantic Standardization of Ortholo-gyConte NT [J]. Semantic Web Applications and Tools for Life Science 2015; 134 (41): 74-83.