Egyptian Informatics Journal 23 (2022) 95–103

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

AutismOnt: An Ontology-Driven Decision Support For Autism Diagnosis and Treatment

Mariam M. Hassan *, Hoda M.O. Mokhtar

Faculty of Computers and Information, Cairo University, Egypt

A R T I C L E I N F O

Article history:
Received 13 November 2020
Revised 19 June 2021
Accepted 5 July 2021
Available online 31 July 2021

Keywords:
Autism spectrum disorder
Decision tree
Medical ontology
Decision support

A B S T R A C T

Autism Spectrum Disorder (ASD) is a deleterious neurodevelopmental disorder affecting 1 in 54 children. The complex interdisciplinary nature of ASD research, however, introduces challenges to the spread and accessibility of new discoveries among researchers in different disciplines; furthermore, this highly complex research environment makes it even harder for practicing physicians and primary care providers to keep up with recent advances, which would profoundly impact the extent toward which recent research impacts standard medical and caretake practices. In order to contribute toward bridging the gaps between researchers in different fields, practicing physicians, and primary caretakers, we have created the most expansive autism ontology up to date (AutimsOnt) through utilizing the Protégé ontological framework. With 676 classes and more than 124 properties, AutismOnt can serve as the foundation to support a wide range of applications ranging from decision support systems for practicing physicians to text annotation processes that would allow for the creation of an interdisciplinary research platform where investigators can easily share and retrieve scientific findings. The Ontology is available in the NCBO BioPortal.

1. Introduction

Autism Spectrum Disorders (ASD) is an umbrella term that covers a multitude of early-onset neurodevelopmental conditions characterized by a wide array of cognitive, behavioral, and social impairments [1]. The symptoms exhibited by ASD patients are diverse, as ASD is very heterogeneous in its manifestation, with a 4:1 male to female gender ratio and a wide range of symptom severity ranging from highly functioning individuals who can lead a relatively independent life as adults to severely mentally impaired individuals who would require constant caretaking and supervision throughout their lives [2,3]. Despite the diversity of symptoms, however, most ASD individuals suffer from a reduced quality of life, as even high functioning individuals still face social challenges, which can be a source of distress; furthermore, ASD also incurs a tremendous emotional and financial toll on the family of ASD individuals as well as society as a whole [4].

Research on ASD has progressed rapidly over the past two decades [5]. ASD research, however, is extremely multi-faceted, ranging from basic scientists seeking to understand the molecular and cellular underpinnings of the disorder through utilizing animal models and molecular neurobiology approaches to behavioral scientists working with patients in order to develop new approaches to help the parents and educators better take care of individuals afflicted with this disorder. This diversity in research makes it particularly challenging to bridge the gap between such disparate disciplines; furthermore, as the research environment grows increasingly complex, the development of techniques ensuring the accessibility of this research output to practicing physicians and primary caretakers becomes of paramount importance [6–8].

Biomedical ontologies have recently witnessed a remarkable increase in popularity as the tool of choice for bridging interdisciplinary gaps and ensuring the widespread accessibility and exchange of information between researchers and professionals with diverse backgrounds and a broad range of specializations [6–14]. The striking success of biomedical ontologies stems from

* Corresponding author.
E-mail addresses: m.mostafa@pg.cu.edu.eg (M.M. Hassan), h.mokhtar@fci-cu.edu.eg (H.M.O. Mokhtar).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.eij.2021.07.002
1110-8665/© 2022 Published by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University.

their ability to provide a comprehensive framework through which investigators can share their findings and where professionals can easily access the studies upon which their diagnostic and treatment approaches are based.

While several biomedical ontologies have been made for autism, they suffer from their focus on specific aspects of ASD diagnostic and risk factor information; therefore, their uses are limited in the scope of their potential applications [6–8]. In order to address the limitations of the currently available autism ontologies, we sought to develop a universal autism ontology (AutismOnt) that overcomes the existing limitations through exhaustively covering all aspects related to autism researchers and primary caretakers. The expanded hierarchical organization of AutismOnt, makes it one of the most reliable and versatile ontology for the development of future text annotation and decision support systems for ASD researchers, physicians, and caretakers.

Ontologies are used in the medical domain mainly for representing and reorganizing medical terms and terminologies. Building medical ontologies is a challenging task that requires a deep analysis in the medical domain; nevertheless, this analysis is the core of building the medical ontologies that handle issues in the medical field. In this study, the main contributions of AutismOnt ontology are:

- Developing a universal autism ontology that holds a large amount of the most useful autism knowledge from various sources and covers different autism research directions.
- Conceptualizing the available knowledge in different aspects of autism and enabling it to be shared and used by the clinical and research community.
- Providing a comprehensive semantic map for domain terms.

The autism ontology developed in this study unifies some domain terminologies and help practicing physicians and primary care providers in their diagnosis and treatment decisions, which will reflect positively on the autistic individuals' lives.

The rest of the paper is organized as follows: Section 2, summarizes related work. Section 3, methodology is explained. Section 4 the AutismOnt ontology is proposed. Finally, we conclude by pointing out some future research lines.

2. Related Work

Biomedical ontology is an important topic investigated by several researchers. It plays a fundamental role in the domain of medical informatics [6] by contributing, for instance, to knowledge management, data integration, and reasoning and decision support [7]. Biomedical ontologies can serve as a source of vocabulary to support various knowledge management tasks such as: annotating data and resources [8], accessing biomedical information [9], managing and extracting relevant information [10], and mapping across biomedical ontologies [11]. In addition, biomedical ontologies are a crucial component in semantic interoperability and data exchange in medicine, along with standardizing clinical information [12]. Finally, they can help in reasoning through: data selection, data aggregation, decision support [13], natural language processing applications, and knowledge discovery [7]. A few ontologies were developed in the autism domain for certain purposes.

Tu et al. [14] developed an autism ontology based on the Open Biomedical Ontologies Foundry (OBO) principles and which conforms to the Basic Formal Ontology (BFO) [15]. The ontology supports the integration and annotation of terminologies to enable users perform queries and inferences for autism phenotypes from the National Database for Autism Research (NDAR) repository. This ontology consists of 34 classes describing phenotypes, together with four classes representing autism diagnostic instruments: Autism Diagnostic Interview-Revised (ADI-R) [16], and two out of four modules from the Autism Diagnostic Observation Schedule (ADOS) [17]. Moreover, the ontology contains 15 Semantic Web Rule Language (SWRL) [18] giving the ontology the ability to infer phenotypes for the provided patient data. One year later, the same team built a project based on this ontology and the main goal was to assist researchers to access clinical data in NDAR [19].

McCray et al. created a phenotype ontology based on 24 instruments. The selected instruments are the most important and commonly used ones, ranging in their structure and type between interviews, direct assessment, and questionnaires [20]. Their goal was building an ontology to help in accessing autism data, assessing ASD instruments and comparing these instruments for better evaluation. In order to define ontology concepts, various diagnosis instruments and tools have been collected and evaluated for phenotype data. Various metrics were applied to build the phenotype ontology, for instance, existing metrics at the National Center for Biomedical Ontology (NCBO BioPortal) are used to find structural weaknesses, in addition, they developed further metrics in order to analyze ontology content and provide suggestions for additional ontology development, iteratively applying the metrics over the ontology to highlight structural imbalances, inconsistencies, and areas that need a review.

Mugzach et al. extended the ontology developed by Tu et al. [14] by adding the ability to infer autism phenotypes based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) 4th and 5th edition (DSM-IV, DSM-5) [21,22] criteria [23]. In addition, they integrate McCray's phenotypic hierarchy [20] into BFO hierarchy used in their ontology. They enriched the ontology by using instruments such as DSM-IV and ADI-R [16]. The ontology was expanded based on the DSM-IV and DSM-5 criteria by adding 443 classes and 632 rules that represent phenotypes, together with their synonyms, environmental risk factors, and comorbidities. They represent all DSM criteria as OWL classes arranged in a hierarchy form.

Although, Tu's ontology [14] contains SWRL rules, the ontology does not include diagnosis criteria; thus, it cannot assist the diagnosis process. In addition, the ontology was built for research purposes, thus all classes and sub-classes are defined from literature keywords and abstracts which results in an inaccurate comprehensive ontology compared to the latest autism ontologies. In addition, their project that is based on this ontology is not helpful for other users (i.e., physicians or autistic child family members). The system can't provide information like decision making in the diagnosis or treatment for doctors or therapists. Likewise, the Autism Spectrum Disorder Phenotype ontology only covers three main topics in autism domain: the personal traits, social competence, and medical history [20]. Treatment and other domain topics such as: biology (i.e., genes), services, and lifespan issues are not covered, and autism risk factors are not organized in proper groups and only a few of them are included. Both ontologies didn't capture a wide range of domain knowledge nor did they unify a large set of terminologies. On the other hand, the most recent ontology focuses on the diagnosis of ASD [23]. Alternatively, the ontology does not cover any other topics related to autism domain such as, the effect of different autism treatments, and various disorders and disabilities that occur to autistic people. Nevertheless, the proposed AutismOnt ontology is offering a panoramic view of ASD research main categories through 8 sub-classes covering: Diagnosis, Risk Factors, Treatments, Strength and Weakness, Services, Lifespan Issues, Profile, and Family Relationships. The ontology classes indicate the effect of the diagnosis on treatment and services. In addition, it provides a diagnosis decision support through reasoning over SWRL rules.

3. Materials and Methods

3.1. Materials

Building the ontologies to describe the medical domain or representing the medical terminologies is a challenging task, as analyzing the domain and structuring the medical terminologies should take place before building the ontology. The medical domain has its own terminologies and lexicons that represent the patient information and the medical knowledge in an efficient way. We started enumerating important terms in the ontology by searching for the most accurate and efficient top-level hierarchy of autism domain. The most suitable top-level structure, which has been chosen for this work is the classification of the "Autism Spectrum Disorder Research: The Global Landscape of Autism Research 2012[1]" by The Interagency Autism Coordinating Committee (IACC). In addition, the terms and terminologies in diagnosis and medical history instruments available in NDAR are used for sub classes and data properties. For a further enhancement, autism encyclopedia are used while building the AutismOnt class hierarchy since it offers descriptions and definitions of the array of terminologies associated with autism research and other pervasive developmental disorders (PDDs) [24–26]. Finally, the DSM-5 classification and diagnosis criteria was included in the ontology.

In considering reusing existing ontologies, we searched in ontology libraries: NCBO BioPortal[2], OBO[3], and the Ontology Lookup Service (OLS)[4] for available autism ontologies. The resulting three ontologies are: Autism Spectrum Disorder Phenotype[5] (1), Autism DSM-ADI-R ontology[6] (2), and Ontology of Language Disorder in Autism[7] (3). The ontology used while constructing our proposed ontology was the Autism Spectrum Disorder Phenotype Ontology developed by McCray et al. [20], because of its similarity to our concept representations. Since autism is heritable, it is necessary to include family relations with their medical history. Thus, we used a full family relationship tree using the Family Health History Ontology[8], the relationships have almost all family relations to help in the future enhancement of autism ontology.

Finally, the data used in the machine learning model to generate SWRL rules are obtained from the NDAR repository, consist of Subject Medical History (ace-subjmedhist)[9] and Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2) modules 1, 2, 3, and 4[10,11,12,13]. Then records from those files are mapped to generate a new dataset which consists of both medical history features and diagnosis results for subjects. The mapping process was especially challenging since each subject can have multiple records in those files which occurred on different dates. The diagnosis column in the generated dataset will be used as a label in the decision tree model. The total number of subjects is 1534 (1316= Autism, 211= controls). Age ranges between 16 and 254 months. 1024 of the subjects are males, and 510 are females. The data attributes are categorized under 13 categories: Diagnostic History (1), Prenatal/Early Postnatal History (2), Developmental History (3), Current Medications (4), Review of Systems (5), Pulmonary (6), Cardiovascular (7), Gastrointestinal (8), Endocrine/Metabolic (9), Hematologic (10), Immunologic (11), Genetic (12), and Mental Health (13).

3.2. Methods

Ontology building methodologies are very important, it reflects the ability of an ontology to obtain efficient results and its ability to grow and integrate. For building the AutismOnt ontology we used a hybrid method of the Uschold and King et al. methodology [27] and the guide official for the ontology development that Stanford University recommended [28]. Based on these two methodologies, the AutismOnt ontology is built based on a set of guidelines and steps: Identify the purpose of the ontology (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 (object and data properties) (6), Define the Semantic Web Rule Language (SWRL) (7), and Create ontology instances (8). For ontology editing, Protégé was chosen due to its powerful abilities and its full support to OWL ontologies. The ontology was represented in OWL language, and SWRLTab[14] is used for defining and running SWRL rules representing rules obtained form decision tree algorithms over Subject Medical History.

Fig. 1 shows the methodology workflow, which consists mainly of: building the structure of the ontology which includes concepts with their properties and relations (i.e., Descriptive Knowledge), and the logical relation that allows more explicit knowledge that enables the drawing of inference from descriptive knowledge, through SWRL rules (i.e., Procedural Knowledge). The methodology is further discussed in the following subsections.

3.2.1. Descriptive Knowledge

The terminologies and concepts of the ontology define the autism domain knowledge. The data discussed in the Material section are organized in a hierarchical form to build a firm relation among main classes of Autism, and, at the same time, each class can be considered as a stand-alone ontology with its own subclasses; nevertheless, the object properties and relation within them is the coherence of the ontology, as ontology classes cannot provide enough information solitarily. There are two kinds of properties: object properties and datatype properties. Object properties point at a class of the ontology. The datatype properties contain a value and describe the relationships between instances (individuals) and data values.

3.2.2. Procedural Knowledge

SWRL is used to add rules into an ontology to provide an additional layer of expressiveness. We write SWRL rules from valid relationships between AutismOnt ontology concepts to detect high presence of autism. Rules are generated from machine learning model and converted into SWRL format. The machine learning model used in this study is the decision tree algorithm, each path from node to leaf represent a single SWRL rule.

The data set is split equally. The first part is used as input in the decision tree model, while the other part is stored as instances in the AutismOnt ontology, in order to be used in evaluating the generated SWRL rules. Before running the model, pre-processing stage took place to clean and standardize the data, for example unify the term "Autism" (i.e., a positive diagnosis of autism could be written by "AUTISM", "YES", "1", "Aut", "Autistic", and "autism"). In this study, we use the 'Decision Tree Classifier' from scikit-learn, which uses an optimized version of the Classification And Regression Trees (CART) algorithm. In evaluating the classifier performance,

1 https://iacc.hhs.gov/publications/publications-analysis/2012/
2 https://bioportal.bioontology.org/
3 http://www.obofoundry.org/
4 https://www.ebi.ac.uk/ols/index
5 https://bioportal.bioontology.org/ontologies/ASDPTO
6 https://bioportal.bioontology.org/ontologies/ADAR
7 https://bioportal.bioontology.org/ontologies/LDA/?p=summary
8 https://bioportal.bioontology.org/ontologies/FHHO/?p=summary
9 https://nda.nih.gov/data_structure.html?short_name=ace_subjmedhist01
10 https://nda.nih.gov/data_structure.html?short_name=ados1_201201
11 https://nda.nih.gov/data_structure.html?short_name=ados2_201201
12 https://nda.nih.gov/data_structure.html?short_name=ados3_201201
13 https://nda.nih.gov/data_structure.html?short_name=ados4_201201

14 https://protegewiki.stanford.edu/wiki/SWRLTab

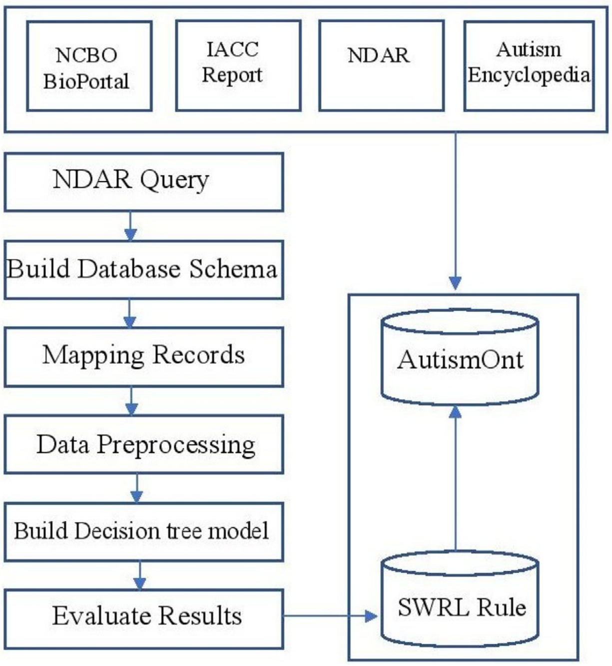

Fig. 1. AutismOnt building methodology workflow.

we used different metrics and measures. In addition to calculating accuracy, we calculated the sensitivity which measures the model's ability to correctly detect autism patients. Other metrics are used for evaluation: Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Likelihood Ratios (LRs), Positive LR (LR+), and Negative LR (LR-). Specificity refers to the model's ability to correctly reject non-autistic subjects. PPV is the proportion of the number of positive results over number of positive results predicted by the model, also known as precision. NPV is the probability that subjects without autism truly don't have autism. LRs are used for assessing the performance of a diagnostic test. LR+ is the ratio of the proportion of patients who have autism and test positive to the proportion of patients without autism who also test positive, and LR- is the ratio of the proportion of patients who have autism and test negative to the proportion of patients without autism who also test negative. In order to compare results from the scikit-learn classifier, four decision tree classifiers (ADTree, CDT, J48, and LADTree) were picked to be compared with the 'Decision Tree Classifier' from scikit-learn.

3.2.3. Ontology Evaluation

In this study, Pellet is used as the reasoner in AutismOnt, having widespread support for constructing ontologies. The Subject Medical History dataset mapped with the ADOS-2 four modules from the NDAR are used to evaluate the ontology model. In the beginning, we converted the patient's data stored in the MySQL database into the AutismOnt ontology using the Resource Description Framework (RDF) triple (subject–predicate–object) format. Then, we run the reasoner and the results generated based on the stored SWRL rule was evaluated using different metrics: accuracy, specificity and sensitivity.

4. Results

4.1. AutismOnt Hierarchy

The AutismOnt maps the concepts and knowledge of the autism domain. The top-level classes of the AutismOnt ontology and their sub-classes (shown in Fig. 2) are extracted based on the terms and terminologies in medical ontologies, alongside trusted medical literature and autism encyclopedia. The following paragraphs describe main classes (i.e., Top- level classes) and their subclasses (Table 1). The ontology main metrics that are shown in Table 2 demonstrate that the maximal depth is 7. In addition, 2,347 axioms were used to structure AutismOnt. The majority of these axioms used *is-a* relations. In NCBO BioPortal, 35 concepts are common to AutismOnt and Autism Spectrum Disorder Phenotype Ontology, 27 are common to Logical Observation Identifier Names and Codes, and 4 to the Ontology of Drug Neuropathy Adverse Events.

Diagnosis Class includes: DSM-5 Criteria, DSM-5 Result, Diagnostic Instruments, and signs and symptoms of ASD, and such classes can help in early intervention. Based on DSM-5 diagnosis criteria, ASD signs fall under two main categories: social communication and restricted repetitive patterns of interest and behavior. Autistic individuals' symptoms in the area of social communication include: deficiencies in emotional and social reciprocity, deficiencies in non-verbal communication behaviors utilized for social interaction, and deficiencies in maintaining and developing proper relationships. The expression of social and communication deficiencies ranges a lot.

Risk Factor Class has three sub-classes. First, Genetic Risk Factors include: genes that are involved in increasing the ASD risk and genes that can help in understanding autism better. A screenshot of gene class is shown in Fig. 3. In addition, an example of gene descriptive logic axioms is shown below (1). Second, Environmental Risk Factors include: pre and postnatal environmental risk factors. Third, Epigenetic Processes include: gene function that happened without changing the DNA sequence. The second and third factors are related because the environmental factors can alter the epigenetic processes by affecting the molecules that bind DNA and control gene function.

$$\begin{aligned} genes \sqsubseteq\ & Genetic - Risk - Factors \sqcap \neg Envirmomental - Risk \\ & - Factors, \exists has - 17q21.33.chromosomal \\ & - alterations \\ \equiv\ & diagnosed - by.Ristricted - and - Repetitive \\ & - Behaviors \end{aligned} \tag{1}$$

Treatments Class represents a list of treatments that can help autistic individuals throughout their life. The treatment becomes one of the important topics especially for the autistic patients and their family members because developing effective treatments will significantly reduce the disability and help enhance the quality of their lives. Various autism treatments are available with the aim of reducing disability symptoms such as: (1) Behavioral: for example, cognitive behavioral therapy and Applied Behavior Analysis (ABA). (2) Complementary and Dietary: for example, research on nutritional supplements, acupuncture, acupressure, cholesterol supplementation, antioxidants. (3) Educational: for example, special education programs, positive behavioral supports, math/reading training, and inclusive education programs.

Services Class represents high-quality, evidence-based, available and accessible services that supports ASD individuals. Special Education Services for autistic students include: Accommodation Plan 504, Individual Education Plan (IEP), and early intervention. Autistic children and teenagers can easily reach support and services from their family and community. On the other hand, Autistic adults' statistics shows that only 15 to 20 percent of them live independently and work without external support, without taking in consideration their cognitive functioning level. Therefore, mapping needed services to Autistic adults is very important as studies shows that many of them faces mal-employment, where they work in jobs beneath their qualifications.

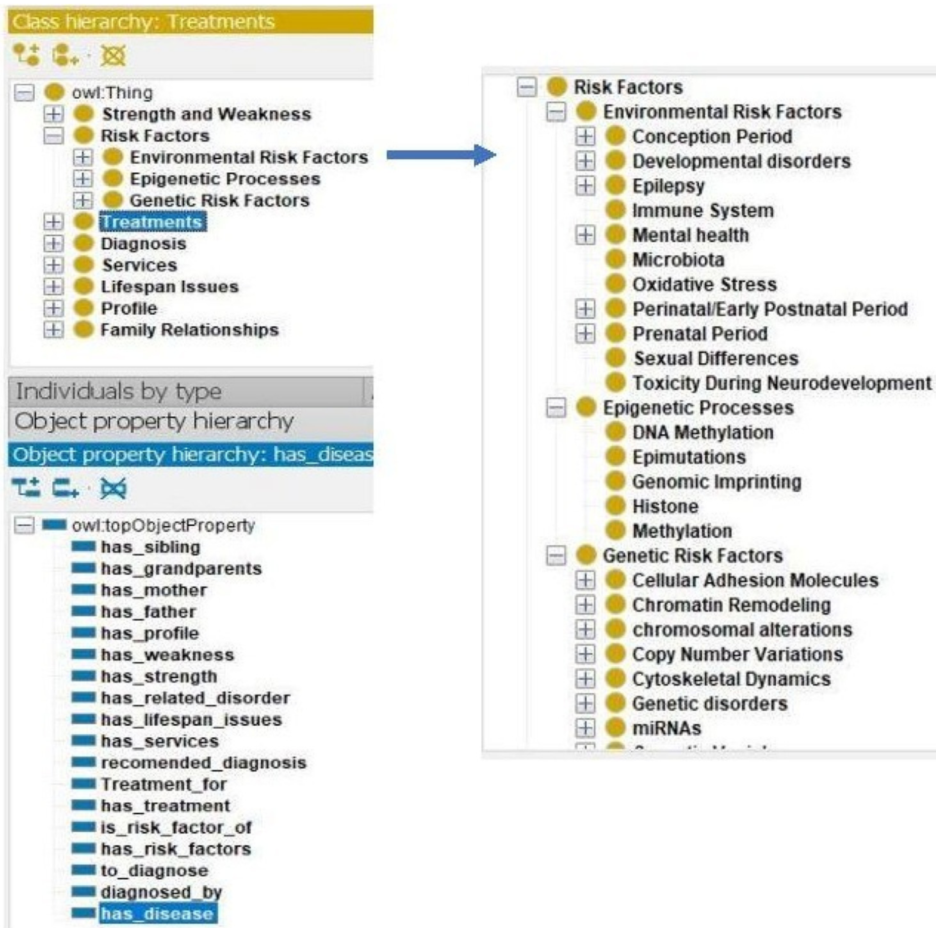

Fig. 2. AutismOnt top-level class hierarchy.

Lifespan Issues Class represents general categories in this field that reflect the changing needs in ASD individual's life from kindergarten through college. Although numerous numbers of autistic children are diagnosed and detected throughout the preschool years, many of them start their school without being diagnosed. Life span issues class classified difficulties autistic individuals faced based on age and academic level: school aged, middle and high school, and college. Difficulties in middle and high school could be social challenges, bullying, victimization, mental health issues (e.g., anxiety and depression), pace with academic demands, and challenges in independent living skills. While in college life challenges could be academic accommodations, self-awareness, self-advocacy, self-disclose, social skills, relationship skills, organization, time management, control emotions, and stress.

Strength and Weakness Class autism is said to be a spectrum because symptoms vary a lot from very mild to very severe, consequently various support and functionating are required for Autistic individuals across the spectrum. Though there exists this broad variability throughout autism, it is observed that there exist specific patterns of strengths and weaknesses related to autism. Strengths of autistic individuals may include the capability of recalling and memorizing various information and facts. In addition, they are good followers to procedures and rules, they use it with their talent on visual learning. On the other side, they suffer from patterns of weakness, to start with, they do not have flexibility, nor can they manage/organize their time. Another drawback are the challenges they face while working within groups and in situations where they are required to communicate. The strength and weakness class is classified into general strengths and weaknesses, such as: memorize facts, follow concrete rules, visual learning, and flexibility, organize, manage time, work within groups, social communication, auditory processing, verbal information. In addition to strength and weakness in school subjects (e.g., reading, writing, and math) and processing (e.g., attention, memory, and executive functions)

Profile Class consists of age, blood type, gender, race, dietary restrictions, dietary supplements, vaccination, and medical history of the individuals. Medical history contains 61 related disease/comorbidities sub-classes organized in tree hierarchy, such comorbidities refer to related disorders and disabilities to ASD. Usually, comorbidity is two or more illnesses or disorders happening in the same individual at the same time interval or one after the other. In the AutismOnt Ontology comorbidities were linked to their related signs and symptoms, as signs and symptoms of some comorbidities may overlap with ASD, such as attention deficit hyperactivity disorder (ADHD). It was estimated that 30% to 80% of autistic individuals meet the diagnostic criteria of ADHD. Such overlap had a great importance to capture while diagnosis, since they may have some biological roots that will affect the treatment prescribed for the subject.

Finally, **Family Relationships Class:** contains biological family relationships from 1st, 2nd, and 3rd degree who share behavioral, environmental, and genetic risk factors of Autism.

4.2. AutismOnt Properties

The AutismOnt maps concepts and knowledge in autism domain. The ontology classes cannot provide enough information; therefore, creating object properties and data properties provide more meaning to the ontology. Object properties play an essential

Table 1
Description of AutismOnt main classes.

Class	Description	Subclass
Diagnosis	Categorize list of signs and symptoms of autism to enable early intervention and maximize the possibility to reduce disability.	DSM-5 Criteria, DSM-5 Result, Diagnostic Instruments, Signs
Risk Factors	List of environmental, genetic, and epigenetic risk factors.	Genetic Risk Factors, Environmental Risk Factors, Epigenetic Processes
Treatments	Different treatments and interventions that are safe and effective to be applied over autistic individuals across their lifespan.	Behavioral, Complementary and Dietary, Educational, Medical and Pharmacologic, Model Systems/Therapeutic Targets, Occupational, Technology Based
Strength and Weakness	Patterns of strength and weakness in school subjects and processing.	Strength, Weakness, School Subjects, Processing
Services	High standard services and supports needed by autistic individuals.	Student Disability Service Centers, Housing Supports Tool Kit, Supplemental Security Income, Division of Vocational Rehabilitation, Developmental Disability Waiver, Department of Rehabilitation Services, Special Education Services, Early Intervention
Lifespan Issues	Requirements and issues that faces autistic individuals across their lifespan.	School aged, Middle and High School, College
Profile	General individual information and medical history profile	Race, Blood Type, Gender, Age, Dietary Restrictions, Dietary Supplements, Vaccination, Medical History
Family Relationships	Capture different family relationships	First Degree Relation, Second Degree Relation, Third Degree Relation, Female Blood Relation, Male Blood Relation

Table 2
AutismOnt Metrics.

Ontology Metrics	Number
Classes	676
Properties	124
Upper-level terms	8
Maximal depth	7
Maximum number of children	8
Classes with a single child	0
Classes with more than 10 children	4

role in linking ontology classes. In object properties, there are implicit relations such as is-a and part-of relations, they refer to relationship between classes and their sub-classes, more examples of object properties defined in the AutismOnt are shown at the left-bottom in Fig. 2.

Data properties capture the relationship between ontology classes and their individuals. Data properties describe the relationships between class instance and their data values. Examples of data properties include: Age-at-diagnosis, Gross-Motor, has-Abnormal- Breathing and has-Birthmarks are added to the AutismOnt ontology to connect instances with their classes and assign values to instances in classes, sample of data properties are shown in Table 3. Furthermore, the reasoner plays an important role in building and using the AutismOnt. We applied Pellet reasoner on the AutismOnt ontology to verify the consistency of ontology classes and properties.

4.3. The Implementation of SWRL Rules

The classification performance was tested on the Subject Medical file for the five decision tree algorithms on 69 features. The scikit-learn shows the best accuracy (91.1%), while sensitivity, specificity, PPV, NPV, LR+, and LR-, results were 0.91, 0.71, 0.98, 0.41, 5, 0.12 respectively. Fig. 4 shows a comparison between scikit-learn and other decision tree algorithms. The resultant decision tree is too big to be visualized here; therefore, Fig. 5 shows a segment of it. Each path of the decision tree is converted into a SWRL rule, generating 12 SWRL rules.

SWRL rule includes a high-level abstract syntax for Horn-like rules in OWL Lite and OWL DL (Description Logics). SWRL rule can mainly cover everything from diagnosis to treatment, however, in this study we focus on diagnosis rules. The symptoms and laboratory indexes of autism screening and the specific conditions of patients could lead to early diagnosis which in turn leads to some early intervention which can make significant results on autistic individuals. Given that genes are not easily available in patients' medical records, genes are not taken into account in the diagnosis rules. Examples of SWRL rules are given in Table 4.

Finally, instances (i.e., the patient's medical history) are stored, then we run the reasoner and store results generated based on the stored SWRL rule. The performance analysis for AutismOnt in terms of accuracy of disease prediction/diagnosing autism was 85%, it is shown that the diagnosis specificity is 90.2%, while sensitivity reveals a value of 87.18%.

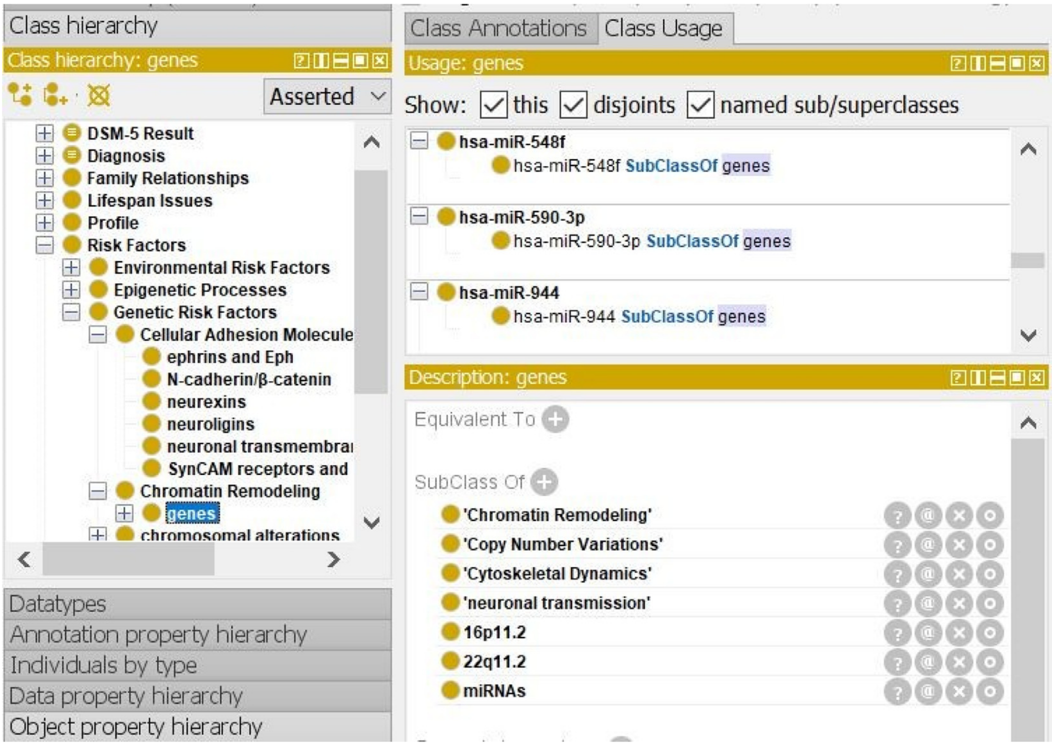

Fig. 3. gene class hierarchy and its properties.

Table 3
AutismOnt Data Properties.

Data property	Domain	Range
Age_at_3-word_sentences	Family_Relations	int
Age_at_diagnosis	Family_Relations	Int
Gross_Motor	Family_Relations	String
has_Abnormal_Breathing	Family_Relations	Boolean
has_Abnormal_Rhythm_or_Heart_Rate	Family_Relations	Boolean
has_Anemia	Family_Relations	Boolean
has_Autism_Disorder	Family_Relations	Boolean
has_Birthmarks	Family_Relations	Boolean
has_Bleeding_Disorder	Family_Relations	Boolean
has_Developmental_Delay	Family_Relations	Boolean
has_Language_Delay	Family_Relations	Boolean

5. Conclusion and Future Work

In this paper, we: conceptualized knowledge, unified terminologies and provided a comprehensive semantic map for domain terms in a form of a formal ontology accessible by both machines and humans, therefore facilitating access to ASD knowledge and supporting professionals and physicians in their clinical decisions (1). In addition, it enables, reuses, and benefits from the knowledge obtained from different studies (2). We focused on the diagnostic process, the constructed rules and patterns from machine learning model enable physicians and experts in their diagnosis. The limitations of this study are: the absence of the graphical user interface

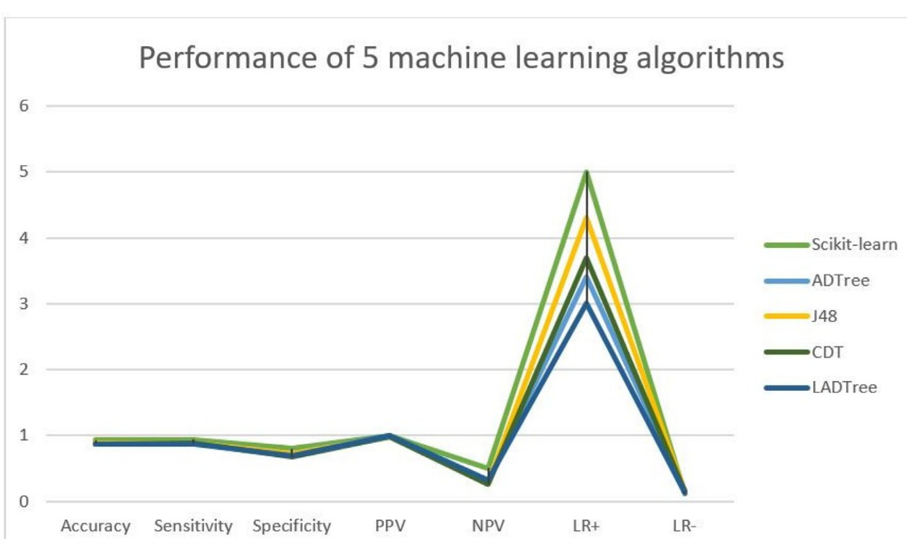

Fig. 4. Performance of 5 machine learning algorithms evaluated for classifying autism cases and non-spectrum controls over the Subject Medical History data.

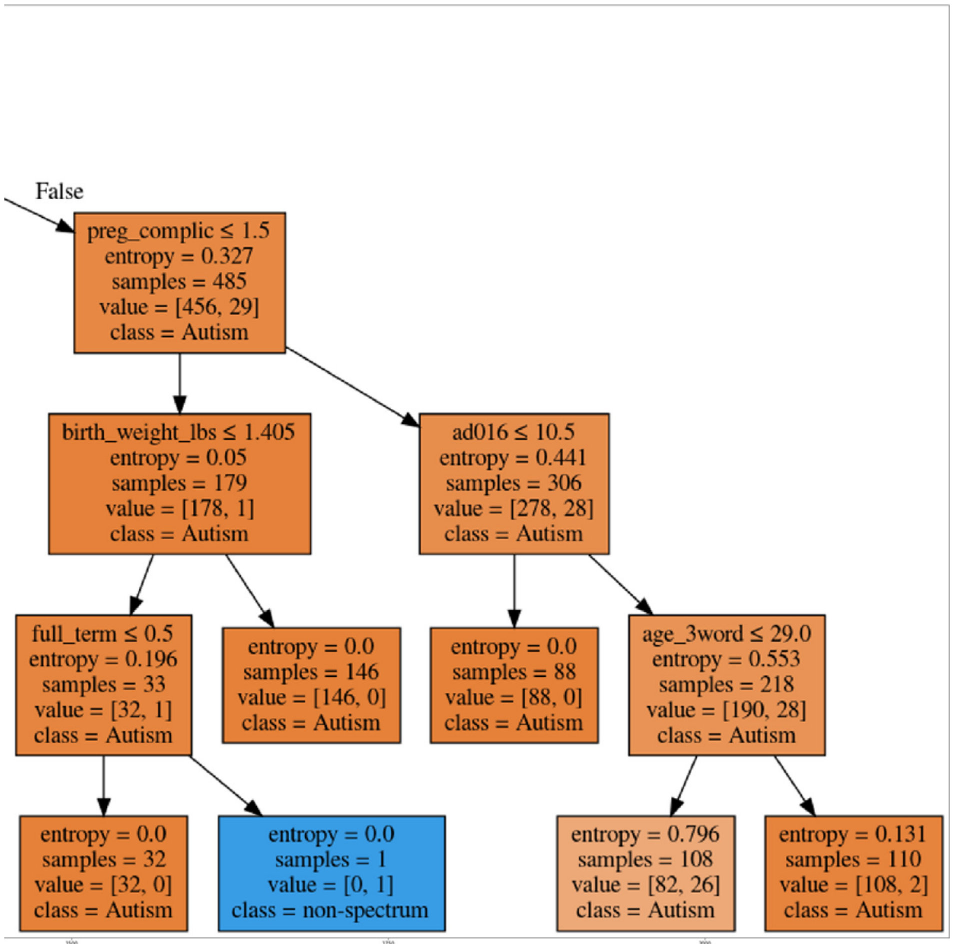

Fig. 5. Part of the Subject Medical History decision tree.

Table 4
Subject Medical History SWRL rules.

SWRL Rules	Prediction Class
asd:Patient(?p) ∧*asd*: *has_Cleft_P alate*(*?p*, "1" ∧ ∧*xsd*: *boolean*) ∧ *has_Developmental_Delay*(*?p*, "1" ∧ ∧*xsd*: *boolean*)∧*Birth_weight_pounds*(*?p*, "< 6.45"∧∧*xsd*: *int*)	Autism
asd:Patient(?p) ∧*asd* : *has_Cleft_P alate*(*?p*, "1" ∧ ∧*xsd*: *boolean*) ∧ *has_Developmental_Delay*(*?p*, "1" ∧ ∧*xsd*: *boolean*) ∧ *Birth_weight_pounds*(*?p*, " greater than 6.45" ∧ ∧*xsd* : *int*) ∧ *Recurrent_Otitis_M edia*(*?p*, "0" ∧ ∧*xsd*: *boolean*)	Autism
asd:Patient(?p) ∧*asd* : *has_Cleft_P alate*(*?p*, "1" ∧ ∧*xsd*: *boolean*) ∧ *has_Developmental_Delay*(*?p*, "1" ∧ ∧*xsd*: *boolean*) ∧ *Birth_weight_pounds*(*?p*, " > 6.45" ∧ ∧*xsd*: *int*) ∧ *Recurrent_Otitis_M edia*(*?p*, "1" ∧ ∧*xsd* : *boolean*)	non-spectrum
asd:Patient(?p) ∧*asd* : *has_Cleft_P alate*(*?p*, "1" ∧ ∧*xsd*: *boolean*) ∧ *has_Developmental_Delay*(*?p*, "0" ∧ ∧*xsd*: *boolean*) ∧ *T otal_days_baby_stay_in_hospital*(*?p*, " < 1.5" ∧ ∧*xsd*: *int*) ∧ *Age_at_3_word_sentences*(*?p*," < 27" ∧ ∧*xsd*: *int*)	non-spectrum

that enables researchers to easily query over the ontology and not annotating research studies to ontology classes, leaving this for future work. For future uses, the ontology can provide supportive decisions to physicians in: (1) the diagnosing process, (2) finding the best treatment plan, (3) and creating a roadmap for the patient, by finding strength and weakness points and services needed for him/her, based on the patient's signs and symptoms, risk factors, medical history, life issues, and family medical history.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Data used in the preparation of this manuscript were obtained from the NIH supported National Database for Autism Research (NDAR). NDAR is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in autism. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the Submitters submitting original data to NDAR.

References

[1] Loomes R, Hull L, Mandy WPL. What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 2017;56(6):466–74.
[2] B. DiCicco-Bloom, The new normal: Families, caretakers, and adults with autism (2014).
[3] Marsack-Topolewski CN, Church HL. Impact of caregiver burden on quality of life for parents of adult children with autism spectrum disorder. American journal on intellectual and developmental disabilities 2019;124(2):145–56.
[4] Rogge N, Janssen J. The economic costs of autism spectrum disorder: a literature review. J Autism Dev Disord 2019;49(7):2873–900.
[5] S. Mughal, R. M. Faizy, A. Saadabadi, C. Doerr, Autism spectrum disorder (nursing) (2021).

[6] Musen MA et al. Medical informatics: searching for underlying com- ponents. Methods Inf Med 2002;41(1):12–9.
[7] O. Bodenreider, Biomedical ontologies in action: role in knowledge management, data integration and decision support, Yearbook of medical informatics (2008) 67.
[8] Doms A, Schroeder M. Gopubmed: exploring pubmed with the gene ontology. Nucleic Acids Res 2005;33(Web Server):W783–6.
[9] Bratsas C, Koutkias V, Kaimakamis E, Bamidis P, Maglav- eras N. Ontology based vector space model and fuzzy query expansion to retrieve knowledge on medical computational problem solutions. In: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2007. p. 3794–7.
[10] Ali F, El-Sappagh S, Islam SMR, Ali A, Attique M, Imran M, et al. An intelligent healthcare monitoring framework using wearable sensors and social networking data. Future Generation Computer Systems 2021;114:23–43.
[11] Bodenreider O. The unified medical language system (umls): integrating biomedical terminology. Nucleic Acids Res 2004;32(90001):267D–270.
[12] M. Gong, Z. Wang, Y. Liu, H. Zhou, F. Wang, Y. Wang, N. Hong, Toward early diagnosis decision support for breast cancer: Ontology based semantic interoperability. (2019).
[13] Ali F, El-Sappagh S, Islam SMR, Kwak D, Ali A, Imran M, et al. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Information Fusion 2020;63:208–22.
[14] S. W. Tu, L. Tennakoon, M. O'Connor, R. Shankar, A. Das, Using an integrated ontology and information model for querying and reasoning about phenotypes: the case of autism, in: AMIA Annual Sympo- sium Proceedings, Vol. 2008, American Medical Informatics Association, 2008, p. 727.
[15] Grenon P, Smith B. Snap and span: Towards dynamic spatial ontology. Spatial cognition and computation 2004;4(1):69–104.
[16] Lord C, Rutter M, Le Couteur A. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994;24(5):659–85.
[17] Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. J Autism Dev Disord 2000;30(3):205–23.
[18] Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M, et al. Swrl: A semantic web rule language combining owl and ruleml. W3C Member submission 2004;21(79):1–31.
[19] Young L, Tu SW, Tennakoon L, Vismer D, Astakhov V, Gupta A, Grethe JS, Martone ME, Das AK, McAuliffe MJ. Ontology driven data integration for autism research. In: 22nd IEEE Inter- national Symposium on Computer-Based Medical Systems. IEEE; 2009. p. 1–7.
[20] McCray AT, Trevvett P, Frost HR. Modeling the autism spectrum disorder phenotype. Neuroinformatics 2014;12(2):291–305.
[21] Association AP et al. Diagnostic and statistical manual of mental disorders. BMC Med 2013;17:133–7.
[22] Association AP et al. Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub 2013.
[23] Mugzach O, Peleg M, Bagley SC, Guter SJ, Cook EH, Altman RB. An ontology for autism spectrum disorder (asd) to infer asd phenotypes from autism diagnostic interview-revised data. J Biomed Inform 2015;56:333–47.
[24] Neisworth JT, Wolfe PS. The autism encyclopedia. Paul H Brookes Publishing; 2005.
[25] Boutot EA, Tincani M. Autism encyclopedia: The complete guide to autism spectrum disorders. Sourcebooks Inc; 2009.
[26] Turkington C, Anan R. The encyclopedia of autism spectrum disorders. Infobase Publishing 2007.
[27] Uschold M, King M. Towards a methodology for building ontologies. Citeseer 1995.
[28] N. F. Noy, D. L. McGuinness, et al., Ontology development 101: A guide to creating your first ontology.