Investigating Semantics-driven Query Rewriting Using Plausible Patterns for Medical Knowledge Discovery and Decision Support

Hossein Mohammadhassanzadeh[1], Samina Raza Abidi[2] and Syed Sibte Raza Abidi[1]

[1] NICHE Research Group, Faculty of Computer Science, Dalhousie University, Canada
{hassanzadeh, raza.abidi}@dal.ca
[2] Medical Informatics, Faculty of Medicine, Dalhousie University, Canada
samina.abidi@dal.ca

Abstract.

Established computational techniques for medical decision making try to keep pace with the increasingly large and complex health data. However, incompleteness of medical data challenges the ultimate success of these techniques. To deal with partial health data, daily clinical decision-making process involves hypothesis generation exploiting both health data gathered during recognition stage and physician's tacit knowledge. Disregarding any element of the process impacts the outcome. Therefore, the integration of physicians' thinking process with machine intelligence seems inevitable for the success of medical decision making. Plausible reasoning is the manifestation of the "plasticity" element of human capability to reason over incomplete data and discover unknown associations by leveraging semantics of concepts. SEmantics-based Data ANalytics framework (SeDan) proposed an approach that integrates plausible reasoning with fine-grained biomedical ontologies to transform an initial query with no answer to an expanded version aiming to infer new knowledge. In this paper we investigate the efficiency of SeDan in a real world medical setting by posing intelligent medical queries form BioASQ challenges over Semantic MEDLINE database. We have developed a Sematic Web-based framework that stores data from databases into an RDF storage and the semantics from two biomedical OWL ontologies conduct the query rewriting.

Keywords: Plausible Reasoning, Query Expansion, Semantic Web Reasoning, Semantic Analytics.

1 Introduction

The emergence of P4-medicine (Predictive, Preventive, Participatory, Personalized) [1] is providing opportunities to analyze large volumes of healthcare data to discover interesting patterns and relationships that may not necessarily be omnipresent in medical knowledge-bases. Given the scope of medical knowledge required for medical decision making, the potential to derive new knowledge using data-driven approaches is not just interesting but extremely useful as it is enabling to extend the knowledge closure of medical knowledge bases with data-derived associations. These data-driven associations may not be congruent with established medical knowledge in terms of their absolute truth/determinism, yet it is worth noting that such relations —plausible as opposed to deterministic— are derived from observations of actual clinical practices that have resulted in positive health outcomes.

Medical data analytics discovers the trends, relations and patterns that might not be known even by experienced clinicians [2]. In knowledge intensive domains, such as health care and life sciences, critical decision-making is typically performed by domain-expert since they are quite adept at generating hypothesis and problem-solving skills—i.e. they (i) understand abstract concepts, semantics and their relationships, and (ii) can apply flexible forms of reasoning when confronted with incomplete knowledge. The same cannot be said for knowledge-centric decision support systems as they require a 'complete' knowledge-base and rather deterministic reasoning algorithms to infer solutions. Given that medical knowledge completeness is relative, especially when considering the open-world assumption, and with the availability of large volumes of health data about clinical practices and processes, there is a need to investigate additional reasoning methods that can infer 'plausible' solutions from the available knowledge.

In this paper, we argue for extending the knowledge coverage of medical knowledge-bases using reasoning methods that can identify 'plausible knowledge' from medical datasets is required—the intent is to employ plausible reasoning approaches to supplement problem-specific knowledge in situations when the available medical knowledge is deemed to be incomplete to solve the problem. Within the health context, plausible reasoning therefore relates to expert's problem-solving process to analyze available knowledge and establish semantic relationships that are inherent within the data to solve complex problems [3], [4]. In this paper, we elaborate the proposed SEmantics-based Data ANalytics (SeDan) framework [4] for medical knowledge discovery by implementing Plausible reasoning methods that work with biomedical ontologies to identify plausible associations from large medical dataset to support medical decision-support.

SeDan leverages the Semantic Web (SW) technologies for knowledge representation and reasoning. To semantically represent the annotated data at various levels of expressivity, we use Resource Description Framework (RDF), RDF Schema (RDFS) and Web Ontology Language (OWL). SPARQL is applied as the semantic query language to retrieve and manipulate the data stored in triple stores and implement the proposed query rewriting algorithm. Linked Open Data (LOD) methods can be used to link data sources to extend data coverage for solving complex problems. We use built-in Description Logic (DL) based reasoning to discover semantic associations in repositories.

Nevertheless, the SW reasoning does not fully support reasoning with uncertainty and incompleteness, which is an irresolvable part of clinical decision making [5]. In our work, we have developed a set of plausible patterns, using the OWL semantics, that provide a plausible extension to OWL [4] to improve reasoning with incomplete knowledge—i.e. to extend the knowledge coverage of a medical knowledge-base.

We demonstrate the functionality of SeDan [4] by developing a healthcare and life science scenario. The experiments are designed to answer real world medical questions, retrieved from BioASQ challenges [6], using large scale data repositories, including DrugBank [7], Disease Ontology [8] and Semantic MEDLINE database [9].

This paper is organized as follows: Section 2 introduces plausible reasoning and the proposed plausible OWL extension. The architecture of SeDan, the plausible reasoner and the proposed query rewriting algorithm is discussed in Section 3. Section 4 elaborates on the design of the experiment, including the materials and methodology. The experimental results are provided in Section 5, followed by discussion and conclusion.

2 Plausible Reasoning

Plausible Reasoning (PR) is the manifestation of the "plasticity" element [10] of human capability to reason over incomplete data and discover unknown associations by leveraging semantics of concepts. PR is (a) non-demonstrative; capable of exploring new knowledge, (b) ampliative; generates knowledge beyond the captured knowledge, (c) non-monotonic; validity of the inferred knowledge depends on the available knowledge, and (d) subjective: depends on the individual perspective [3], [11].

PR performs inferencing by using a set of frequently recurring patterns that do not occur in formal logic [12]. Table 1 presents the 6 plausible patterns. [4] classifies these plausible patterns into 3 groups: hierarchy-based patterns, order-based and hybrid.

Table 1. Plausible Patterns [3]

Plausible Pattern	Description
Generalization [a]	Passing from a given set of objects to a larger set that contains the given set.
Specialization [a]	Passing from a given set of objects to a smaller set that is contained in the given one.
Interpolation [b]	Creating a new relation from observation space X to conclusion space Y, where $x_i \in X$ is not mapped to any $y \in Y$ (unknown relation), but other relations from $x_h, x_j (\neq x_i)$ to Y and $x_h < x_i < x_j$ are known.
A Fortiori [b]	An inference from a proposition with high degree of confidence to a less confident proposition that is not clearly specified but is implicit in the first one.
Similarity/ Dissimilarity [c]	Moving between any two comparable nodes (siblings) in the concept hierarchy.

[a] Hierarchy-based patterns, [b] Order-based patterns, [c] Hybrid patterns

Within the Semantic Web, hierarchical semantics (i.e., *rdfs:subClassOf*, *owl:instanceOf*) conduct generalization and specialization. However, there is no construct in RDF(S) or OWL that supports the partial order. Hence, [4] introduced a plausible extension to OWL to support the representation of and reasoning with order-based semantics within the SW framework. Code 1 shows a snapshot of the implementation of the proposed plausible extension to OWL (PLOWL): an *OrderedProperty* is a property to reflect partial order (i.e., *plowl:standsBefore*) of two entities w.r.t a measurable property (*plowl:Context*). Hence, the plausible reasoner will be able to conduct interpolation and a fortiori reasoning. Hybrid patterns can be performed using either hierarchical relations (*owl:sameAs*) or partial order of concepts; they probe hierarchy and move between any two comparable nodes, or consider the concepts that are (dis)similar regarding some measurable properties.

```
plowl:OrderedProperty
   a owl:Class ;
   rdfs:label "OrderedProperty" ;
   rdfs:comment "The class of (partial) ordered properties." ;
   rdfs:subClassOf owl:ObjectProperty.
plowl:standsAfter
   a owl:OrderedProperty;
   rdfs:range owl:Thing;
   rdfs:domain owl:Thing;
   rdfs:comment "This object property is used to model ordering relation to
   show which concept (subject) locates after another concept (object) regarding
   a specific context. The inverse property is standsBefore."^^xsd:string.
plowl:hasContext
   a owl:ObjectProperty;
   rdfs:range plowl:Context;
   rdfs:comment "This object property links an object property to the context
   nodes being applied to it."^^xsd:string.
```

Code 1. Implementation of the proposed plausible extension to OWL (PLOWL)

3 SeDan: Semantics-based Data Analytics Framework

The SeDan framework (Fig. 1) includes three main modules: plausible reasoner, knowledge sources, and user interface. The plausible reasoner (discussed more in the following section) develops plausible patterns by manipulating the underlying graph with SPARQL query rewriting using OWL DL constructs and domain ontology (discussed more in Section 3.1 Query Rewriting Algorithm). Knowledge sources provide semantics and ontological constructs to conduct the query rewriting, and assertional knowledge to be used to evaluate the expanded query. The system accepts the query with a list of desired plausible patterns via the user interface (similar to the interface provided in [3]), and in return, delivers the plausible answer(s) and their justifications.

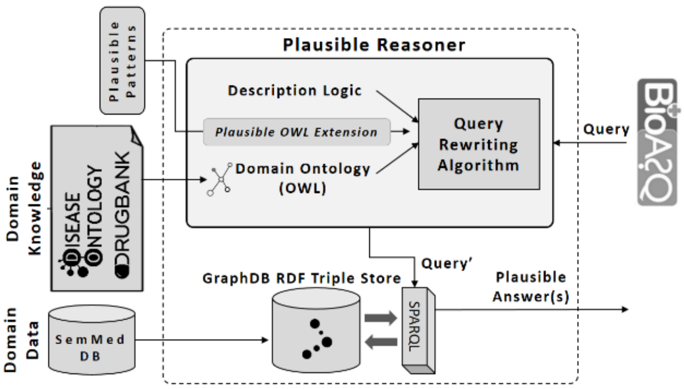

Fig. 1. SeDan framework

3.1 Query Rewriting Algorithm

Query Rewriting Algorithm is the core of the plausible reasoner in SeDan. Query Rewriting (QR) is an approach that uses ontological constructs to transform a given query to an expanded version that elicits both explicit (what a KB knows) and unknown (what it assumes) knowledge from the data [13], [14]. Within the SW framework, OWL 2 QL profile is supported by DL-Lite family. Data independence and support of other variants of DL-Lite have made QL a suitable approach to ontological knowledge discovery in large RDF stores with various levels of expressivity.

SeDan [4] leverages QR as a technique to implement plausible patterns and solve queries that there is no answer for them at the first place. Inspired by GCLRR algorithm [15], the plausible reasoner of SeDan starts the query rewriting (Algorithm 1) with the initial query, a set of preferred plausible patterns and an ontology based on $DL - Lite_T$ axioms that can be semantically enriched with introduced plausible OWL extension.

Starting with the initial query, the algorithm tries to replace the body atom of the query D (step 7), with new atom D'. The atom D' should be (i) semantically related to D ($\exists \alpha \in \mathcal{T}\ \alpha(D, D')$), and (ii) applicable to the preferred plausible patterns (step 6). The new conjunctive query, resulting from replacing an atom, will be added to R, the set of conjunctive queries. This algorithm keeps formulating new queries until there is no unique query to be added.

Algorithm 1. The proposed QR algorithm [4]

Input: A query in a triple format, a set of plausible patterns
$\pi \in \Pi: \{GEN, SPEC, SIM, DIS, FORT, INTP\}$, $DL - Lite_T$ TBOx $\mathcal{T}$ enriched with PL-OWL extension
Output: R, a set of rewriting queries.
1: $R = \{Q\}$;
2: **repeat**
3:   **foreach** *query* $Q \in R$ **do**
4:     **foreach** *atom* D *in* Q **do**
5:       **foreach** *axiom* $\alpha \in \mathcal{T}$ **do**
6:         **if** α is applicable to any $\pi \in \Pi$, w.r.t. D
7:           $Q' = \exists D'.Q(D \rightarrow D') \wedge \alpha(D, D')$;
8:           $R = R \cup \{Q'\}$;
13: **until** no unique query can be added to R;
14: **return** R;

4 Materials and Design of the Experiment

In the real medical world, doctors are required to answer intelligent medical questions over large health data, which is usually sparse, noisy, incomplete and uncertain. Hence, to evaluate the efficacy of SeDan, we try to simulate a real medical setting. In this order, we pose questions from BioASQ challenges over Semantic MEDLINE database [9]. DrugBank [7] and Disease Ontology [8] underpin the query rewriting algorithm to rewrite the initial queries with no answer. The sources, the data processing and the selection of medical questions are discussed as follow.

BioASQ Medical Questions. *BioASQ* challenges [16] are a series of competitions (2013-2017) on large-scale biomedical semantic indexing and question answering. The purpose of the challenges is to assess the capability of machines to semantically index very large numbers of healthcare and life science publications and ontologies to compose brief and easy to understand answers to real-life biomedical questions.

BioASQ questions are formulated by European biomedical experts, reflecting variety of real-life inquiries. The questions belong to 4 distinct categories: yes or no, factoid, list, and summary questions. Summary questions ask for textual answers, which require Natural Language Processing techniques that is not within the scope of this study. Also, some of the questions include qualitative terms (i.e., the most known bacterium) that is not simply possible to transform them into SPARQL queries. Moreover, SeDan targets healthcare applications. While, the BioASQ questions belong to a variety of context; biology, pharmacology, etc. In this regard, for this study, we focused only on those questions of BioASQ that (i) ask for *yes/no* or *factoid* answers; and (ii) ask about *treatment* or *diagnoses*, which comprise massive portion of the questions that doctors confront with. Section 5, Experimental Results, elaborates more on the retrieved questions from BioASQ Task 5, the latest challenge of the series.

Semantic MEDLINE Database. *Semantic MEDLINE database* (SemMedDB) [9] is a significant endeavor to facilitate healthcare and life science studies by providing a comprehensive resource of structured semantic predications. SemMedDB is a database of over 26 million biomedical publications. SemMedDB currently contains over 89 million records (as subject-predicate-object triples) extracted from PubMed citations.

In SemMedDB, the concepts (subjects and objects of the predications) belong to about 120 Unified Medical Language System (UMLS) semantic types (i.e., activity, vitamin, etc.) that could be grouped into 11 National Library of Medicine (NLM) semantic groups (i.e., physiology, disorders, etc.). Predicates are distributed among 34 relations (i.e., causes, occurs in, etc.) and additional 27 negation relations. The predicates also can be classified by plausible semantics to hierarchical relations (i.e., *part of*), order-based relations (i.e., *precedes*) and hybrid (i.e., *same as*).

In this experiment, regarding the nature of the questions selected from BioASQ challenges, we use only 3 semantic groups: disorders (DISO), chemicals & drugs (CHEM), and genes & molecular sequences (GENE). Hence, any combination of these semantic groups, including 6 types of predications; DISO-DISO, DISO-CHEM, DISO-GENE, CHEM-CHEM, CHEM-GENE and GENE-GENE, were extracted from SemMedDB [17]. The resulting RDF repository contains over 11 million semantic predications.

DrugBank. *DrugBank* [7] is a comprehensive database of biochemical and pharmacological information about drugs and drug targets. Each drug entry includes extensive information on properties, structure, and biology of the drugs. In the current setting of SeDan, we are exploiting *DrugBank* version 4.5.0, which in total the RDF representation of it contains over 3.8 million predications.

Disease Ontology. The *Human Disease Ontology* (DO) [8] is a standardized ontology for both human common and rare diseases. The *Disease Ontology* semantically integrates disease and medical vocabularies across disparate biomedical resources; MeSH, ICD, NCI's thesaurus, SNOMED and OMIM. The most up to date version of the DO contains 203,125 semantic predications.

GraphDB. The aforesaid materials and sources are stored to and queried via *GraphDB* RDF triple store (http://graphdb.ontotext.com/). GraphDB is a graph database with RDF and SPARQL support. The capabilities of semantic inferencing, efficient handling of massive volumes of data, real-time inferencing and support of quadruples make *GraphDB* an appropriate tool for SPARQL endpoint in SeDan architecture.

5 Experimental Results

As explained earlier, we focused only on those questions of BioASQ that (i) ask about *treatment* or *diagnoses*; and (ii) ask for *yes/no* or *factoid* answers. As a result, after exploring both the training and test dataset of Task 5-PhaseB of BioASQ series, we retrieved 44 questions (Available here: https://goo.gl/UvP9Sc), including 18 questions asking about *causes* of diseases and 26 questions asking about *treatments* (Table 2).

Table 2. Statistics on the retrieved questions, their resolvability before and after employing the plausible reasoner

Questions asking about	Quantity	Initial Queries		Plausibly Expanded Queries	
		Answered	Not Answered	Answered	Not Answered
Causes	18	6	12	6	6
Treatments	26	7	19	2	17
Total	**44**	**13**	**31**	**8**	**23**

To be able to evaluate the competence of SeDan in expanding the coverage of the knowledge base, firstly, we posed the original questions (as initial SPARQL queries without any plausible manipulation) over the RDF triple store. As Table 2 shows, only 13 questions (%30) were answered using only existing triples stored in the knowledge base – deductive inference. Twelve out of 18 (%67) *causes* questions and 19 out of 26 (%73) *treatments* questions were not resolvable.

In the second step of the experiment, the unresolvable queries from the first step were asked again, but this time by leveraging the full capacity of the proposed plausible reasoner. Table 2 shows semantics-driven query rewriting using plausible patterns can provide answer for half of the initially unanswered *causes* questions and 2 of the unanswered *treatments* questions. Table 3 provides the details of the 8 plausibly answered queries (6 *causes* and 2 *treatments* questions), engaged plausible patterns and the semantics that conducted those patters. It also identifies the database or ontology that the semantics are retrieved from.

For example, query number 2 asks if *statins* cause *diabetes*. Using the SemMedDB there is no explicit information that can justify the initial query of the question − *ASK{statins cause diabetes}*. But, using the semantics in *DrugBank*, we can find *Pravastatin* as a *subStructure* of *statins*. On the other hand, using the triples in the *SemMedDB*, we know *Pancreatitis precedes* (as a *plausible ordered property* - Code 1) *Diabetes* and *Diabetes precedes Myocardial Infarction*, and *Pravastatin causes* both. In this regard, based on the rationale behind the *interpolation pattern* that *"if something is true about two stages of a phenomena, then it might be true for any stages in between",* then *Pancreatitis causes Diabetes.* Likewise, with the logic of specialization pattern that *"when something is true about a class/entity, it might be true about its super class (parent) as well"*, then we can *plausibly* infer *statins cause Diabetes*. Similarly, all the other plausible answers in Table 3 can be justified and explained.

Table 3. details of the plausibly answered queries, engaged plausible patterns and the semantics that conducted those patters

	Plausible Pattern	Ontology/DB	Semantics conducting the Query Expansion
Questions asking about *CAUSES*			
1. What causes Katayama Fever?	{SIM}	Disease Ontology	intestinal_schistosomiasis *hasExactSynonym* Katayama_fever.
2. Do statins cause diabetes?	{SPEC, INTP}	DrugBank and SemMedDB	Pravastatin *subStructure* statins. Pancreatitis *PRECEDES* Diabetes. Diabetes *PRECEDES* Myocardial_Infarction.
3. Can levothyroxine sodium cause insomnia?	{AFORT, SPEC}	SemMedDB	Shock *PRECEDES* Psychiatric_problem. Psychophysiological_Insomnia *subClassOf* Insomnia.
4. Which enzyme deficiency can cause GM1 gangliosidoses?	{SIM}	SemMedDB	Gangliosidosis_GM1 *ISA* Gangliosidoses. Gangliosidosis_GM2 *ISA* Gangliosidoses.
5. Which antibodies cause Riedel thyroiditis?	{GEN}	SemMedDB	Riedel's_thyroiditis *ISA* Thyroid_Diseases.
6. What is the cause of episodic ataxia type 6?	{GEN, SIM}	Disease Ontology	episodic_ataxia_type_6 subClassOf episodic ataxia. episodic_ataxia *hasExactSynonym* Isaacs syndrome.
Questions asking about *TREATMENTS*			
7. Does Herceptin treat prostate cancer?	{SPEC, AFORT}	SemMedDB	Carcinoma *PRECEDES* Malignant_Neoplasms. Malignant_Neoplasms *ISA* Prostate Cancer.
8. What is the treatment of acute myocarditis?	{SIM}	Disease Ontology	septic myocarditis *subClassOf* Myocarditis. acute myocarditis *subClassOf* Myocarditis.

* For the sake of clarity, namespaces (i.e., odo, do, etc) are removed from the triples.
** GEN: Generalization, SPEC: Specialization, SIM: Similarity, AFORT: A Fortiori, INTP: Interpolation

6 Discussion

The experiment illustrates how a plausible pattern, alone or in combination with other pattern(s), can provide plausible answer(s) and extends the coverage of a knowledge base. In this experiment, the plausible reasoner of SeDan expanded the query answering coverage of SemMedDB by %34 in *causes* questions, %12 in *treatments* questions and %18 in total. Although SeDan demonstrates good competency in answering *causes* questions, it only solved few unresolvable *treatments* questions.

Further investigations showed all the queries, which SeDan fails to answer, ask about a concept (i.e., drug or disease) that neither of the ontology or databases include it. Thus, no related semantics or associations is available to conduct query rewriting algorithm. As an example, one of the questions asks about the cancers that can be treated with *Delamanid*, which *DrugBank* and *SemMedDB* do not include any information in that regard. Similarly, another question asks the treatment options for *anxiety in autism spectrum disorder*, which cannot be found in any of *Disease Ontology* or *SemMedDB*.

As mentioned, plausible reasoning is a data-driven inference approach that leverages semantics between concepts to derive new associations. Thus, success and failure of SeDan depends on the richness of the captured domain knowledge. To address this issue, linking to more variety of open data will enrich the semantics supporting the plausible patterns and strengthen the discovery of hidden relationships among data. Moreover, the purpose of BioASQ competitions is to challenge the participants by asking new intelligent questions, that unlike the routine medical questions, are not supposed to be easily answered. So, it was expected that questions about ongoing research, drugs that are not yet approved, or not yet documented diseases would arise. Considering that, we expect SeDan to accomplish better performance in daily routine settings.

In addition to expansion of the knowledge base coverage, the experimental results also indicate the plausible reasoner can provide *complementary* answers for those queries that are already resolved deductively – without plausible reasoning. As an example, SemMedDB returns 7 answers for the question asking for the treatment of *gastric lymphoma*. However, rewriting the query using *interpolation* pattern provides 48 answers (including that initial 7 answers). Hence, SeDan can be considered as a framework that expands the query answering coverage of knowledge base (by exploring the plausible closure) and enriches the solution closure (by fortifying the deductive closure).

As Table 3 demonstrates, SeDan supplements final answers with the plausible patterns and the semantics conducting them, which makes the reasoning process of SeDan clear, transparent and justifiable for the user. Especially in medical scenarios, and life science in general, experts need to understand how an answer/decision is obtained and why it is true. So, the necessity of justification and explanation of answers is inevitable.

7 Conclusion and Future Work

This real word experiment proved that even a (very) large knowledge base (in our case SemMedDB with over 85 million records) suffers from incompleteness and may not be able to answer all the intelligent questions. In this regard, innovative knowledge engineering approaches is required to address the gap. Fuzzy or Bayesian models for a best guess estimation [18] addressed the incompleteness within the knowledge bases. However, these approaches need expert's input, calculating statistical associations, or require probability distribution that may not be always available. Inspired by human thinking process in inferring new knowledge and leveraging the semantics available in ontologies, [3] demonstrated the utility of two well-formalized plausible reasoning approaches, inductive and analogical reasoning, which rely only on hierarchical relations to conduct plausible inference and do not take ordered-based relations into account.

SEmantics-based Data ANalytics (SeDan) framework introduced a plausible reasoner that includes a semantics-driven query rewriting algorithm that is conducted by plausible patterns, including both hierarchical and ordered-based. It also developed a plausible extension to OWL to support (partial) order-based patterns. We evaluated the efficiency of SeDan in a real medical setting. The results showed it expanded the KB coverage by resolving %50, %11 and %26 of initially unanswered questions asking about causes, treatments of diseases and all the unanswered questions, respectively.

In this paper, we focused on functionality of SeDan. Investigating the performance of the query rewriting algorithm and possible improvement is the future work. Also, studying the efficiency of SeDan in answering more routine medical questions will give us a better understating of its competence in daily diagnoses and treatments. We also consider incorporating analogical rules reflecting the clinicians' knowledge to conduct the reasoning in the cases that plausible patterns fail. Currently, only *DrugBank* and *Disease Ontology* are incorporated into the domain knowledge. We need to enrich the domain semantics by linking the framework to other open data.

Acknowledgment: This research is supported by an NSERC Discovery Grant.

References

[1] L. Hood and S. H. Friend, “Predictive, personalized, preventive, participatory (P4) cancer medicine,” *Nat. Rev. Clin. Oncol.*, vol. 8, no. 3, pp. 184–187, Mar. 2011.

[2] D. Windridge and M. Bober, “A kernel-based framework for medical big-data analytics,” *Interact. Knowl. Discov. Data Min. Biomed. Informatics*, pp. 197–208, 2014.

[3] H. Mohammadhassanzadeh, W. Van Woensel, S. R. Abidi, and S. S. R. Abidi, “Semantics-based plausible reasoning to extend the knowledge coverage of medical knowledge bases for improved clinical decision support,” *BioData Min.*, vol. 10, no. 1, p. 7, 2017.

[4] H. Mohammadhassanzadeh, S. Raza Abidi, M. S. Shah, M. Karamollahi, and S. S. R. Abidi, “SeDAn: a Plausible Reasoning Approach for Semantics-based Data Analytics in Healthcare,” in *Workshop on Artificial Intelligence with Application in Health (WAIAH 2017), held within the AI*IA 2017 Conference*, 2017.

[5] D. Ausín, D. López-de-Ipina, and F. Castanedo, “A probabilistic OWL reasoner for intelligent environments,” in *Proceedings of the 10th International Conference on Uncertainty Reasoning for the Semantic Web*, 2014, vol. 1259, pp. 1–12.

[6] “The BioASQ Challenge.” [Online]. Available: http://www.bioasq.org/.

[7] V. Law, C. Knox, Y. Djoumbou, and T. Jewison, “DrugBank 4.0: shedding new light on drug metabolism,” *Nucleic Acids Res.*, vol. 42, no. D1, pp. D1091–D1097, 2013.

[8] W. Kibbe, C. Arze, V. Felix, and E. Mitraka, “Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data,” *Nucleic Acids Res.*, vol. 43, no. D1, pp. D1071–D1078, 2014.

[9] T. Rindflesch, H. Kilicoglu, and M. Fiszman, “Semantic MEDLINE: An advanced information management application for biomedicine,” *Inf. Serv. Use*, vol. 31, no. 1–2, pp. 15–21, 2011.

[10] P. Naur, “Computing versus human thinking,” *Commun. ACM*, vol. 50, no. 1, pp. 85–94, Jan. 2007.

[11] G. Pólya, *Mathematics and Plausible Reasoning (vol. 2): Patterns of plausible inference*. Princeton University Press, 1954.

[12] M. Virvou and K. Kabassi, “Adapting the human plausible reasoning theory to a graphical user interface,” *IEEE Trans. Syst. Man, Cybern. A Syst. Humans*, vol. 34, no. 4, pp. 546–563, 2004.

[13] H. Pérez-Urbina and E. Rodrıguez-Dıaz, “Evaluation of query rewriting approaches for OWL 2,” *Proc. of SSWS+ HPCSW*, 2012.

[14] S. Grimm and B. Motik, “Closed World Reasoning in the Semantic Web through Epistemic Operators,” *OWLED*, 2005.

[15] H. Pérez-Urbina, B. Motik, and I. Horrocks, “A Comparison of Query Rewriting Techniques for DL-lite,” *Descr. Logics*, 2009.

[16] G. Tsatsaronis *et al.*, “An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition,” *BMC Bioinformatics*, vol. 16, no. 1, p. 138, Dec. 2015.

[17] C. Tao, Y. Zhang, G. Jiang, M.-M. Bouamrane, and C. G. Chute, “Optimizing semantic MEDLINE for translational science studies using semantic web technologies,” in *Proceedings of the 2nd international workshop on Managing interoperability and compleXity in health systems*, 2012, pp. 53–58.

[18] R. Almeida, U. Kaymak, and J. Sousa, “A new approach to dealing with missing values in data-driven fuzzy modeling,” *Fuzzy Syst. (FUZZ), 2010 IEEE Int. Conf.*, 2010.