2017 3th International Conference on Web Research (ICWR).

A Proposed Representation Framework for Semantic Science

Soroush Mobasheri, Mehrnoush Shamsfard*
Faculty of Computer Science and Engineering,
Shahid Beheshti University,
Tehran, Iran
s_mobasheri@sbu.ac.ir, m-shams@sbu.ac.ir

Abstract—**This work addresses the problem of enabling machines to perform scientific tasks, e.g. reasoning based on scientific laws and definitions, recognizing inter-dependence of scientific domains, and answering queries about science corpus. The building blocks of science, such as scientific terms, laws, problems, solutions, theories and disciplines are traditionally represented as single, atomic nodes in scientific ontologies. This makes it difficult to distinguish those constituents and use them properly in the automation of scientific activities.**

We support the idea of adding structure to the representation of different constituents of science corpus. The structure of a scientific law, for instance, would be different from that of a solution to a given scientific problem. It is shown through examples that considering those different structures can help in reasoning about scientific knowledge. Moreover, the domain-independent aspects of different constituents of science have the potential to be factored out in a meta-ontology. This meta-science can also contain general reasoning machinery about science.

Keywords—knowledge representation; scientific ontology; meta-science; semantic science

I. INTRODUCTION

The corpus of science is large and complex. Components of science, e.g. scientific domains, theories, laws, definitions, methods, questions and observations constitute a heterogeneous, huge, yet growing collection. As the volume and complexity of science increases, it becomes more desirable to make use of computing capabilities in performing research tasks.

This work addresses the problem of enabling machines to perform a sort of scientific tasks. As there are quite a large variety of tasks in science, we focus on the ability of answering a number of selected tasks. The tasks are chosen from among daily activities of science practitioners. Toward this goal, we propose a representation scheme for several types of components of science corpus, namely, scientific definitions, laws of nature, and research articles.

A. Problem Statement

Here is a list of scholar tasks, supposed to be performed by machines. We will use them in section III as the competing questions to evaluate the representation framework proposed in section II.

1. Find the value of a variable under question, from the values of related variables.
2. Given a specific entity *e*, retrieve the part of scientific knowledge which models *e*.
3. Recognize the dependence of a given scientific term to others.
4. Likewise, recognize the dependence of a given scientific domain to others.
5. Answer queries about different constituents of the science corpus, e.g. given a scientific problem *P*, list the collection of proposed solutions to *P*.

Section II proposes exemplar segments of the representation of science, and shows how it may be used to fulfil the above-mentioned requirements.

B. Related Work

Great works have been performed in order to build large-scale ontologies of real entities, as well as existing relations among them, e.g. Linked Data [1] and YAGO [2]. Machinery has been developed [3]–[6] in order to extract information about entities and relations from the Web. We deal with structured entities of science and their inter-relations, which together constitute the corpus of scientific knowledge. Moreover, those entities (e.g. the Hooke's Law) are supposed to govern attributes and relations of real world entities. These sort of (scientific) entities are well expected to be mentioned in the outstanding ontologies mentioned above. However, as will be seen in section II, it is necessary that the building blocks of science be presented with corresponding structures so that they can be employed in scientific reasoning.

Scientific ontologies are the main candidates to contain scientific terms, laws and other constituents of scientific

978-1-5386-0420-5/17/$31.00 ©2017 IEEE

knowledge. Ontology of Physics for Biology (OPB) is an ontology of classical physics, applied to the dynamics of biological systems [7]. Physical laws are not identified in this ontology. However, as an exemplar physical term, we consider *velocity* which is defined as a subclass of *rate property*. Velocity is not formally related to the concept *position* in OPB. Therefore relying computational engines cannot know from the ontology, how to find the velocity of a particle, given its position as a function of time. The lack of computational relations among scientific terms prohibits the use of scientific definitions and laws in reasoning. We propose in Sections II.B and II.C that scientific definitions and laws of nature may be used to equip scientific ontologies with inference rules. Those inference rule can potentially be employed in scientific reasoning procedures.

This extension may be equally proposed for other scientific ontologies, such as the ontology of physics provided in [7], the well-known SUMO (Suggested Upper Merged Ontology) [8], the SWEET (Semantic Web for Earth and Environmental Terminologies) [9] developed in NASA, and the ontology of Earth crust fractures [10]. Although some of these ontologies (e.g. SUMO and SWEET) does identify scientific laws, those laws (as well as scientific terms and variables) are not distinguished by means of any specific structure or inference mechanism.

Wolfram Alpha website [11] makes it possible to ask for the consistent value of one, from among all the variables of a known scientific law, while values of other variables are given. For instance, given the value V of the electric voltage across a resistor with given resistance R, one may ask for the consistent value of electric current I through it, based on the Ohm's law. While the website provides a large collection of scientific equations, it is left to the user to choose the desired equation. This is because there is no relation between the scientific equations on the one side, and the part of world they model. We propose that in addition to the mathematical constraint among variables, it is necessary for a scientific law L to explicitly point to some part of world, i.e. the part of world which is modeled by L. This brings the knowledge required for selecting relevant set of laws, from among the big collection of all the known laws of nature, which in turn, potentially makes it possible to reason through all relevant laws.

In [12] an ontology of scientific experiments is provided. This ontology is later employed in the automation of parameter setting of a specific experiment in bioinformatics [13]. The Executable Paper Grand Challenge [14], [15] is a challenge asking for technology enabling validation of data and code, and decreasing the reviewer's workload. Both projects addressed the overall goal of automating scholar tasks. While targeting the same goal, the present work deals with scientific definitions, laws of nature, research articles and problems, as well as some general, domain-neutral aspects of scientific knowledge which we call meta-science.

It is proposed in [16] to equip conference papers with rich, open-standard structures. It is defended that this will help in enabling machines to mine scientific knowledge. While supporting this, we propose to spread this idea yet outside the articles, to the relations among research articles, (identified) research problems and solutions.

II. Proposed Representation Scheme

This section presents our proposal to achieve the goals mentioned in Section I.A. We begin by a brief description of the approach. Then we give a more detailed picture, for some types of the components of science.

A. Approach

Automation of any scientific task, requires the representation of the related components of science. For instance, let us consider the task of measuring the performance of a given algorithm, mentioned in Section I.A above. Suppose that an algorithm is proposed in a research article to solve a well-known problem. In this case, it is expected that the algorithm is applied against one, or a number of known datasets and the performance of the algorithm is measured and reported. Ideally, this measurement of performance may be expected to be performed automatically.

Let p be a problem, dataset d contain input data of an instance of p, and m be a performance indicator for any potential solution to p. Moreover, suppose that an algorithm a is proposed to solve p. Clearly, if p, a, and m are not formal, it may not be expected that the performance indicator be measured by the machine. On the other hand, if the problem p, the dataset d and the performance indicator m are formalized, and the algorithm a is provided as an executable code so that it can be applied against d to solve an example of p, then m may be expected to be measured by the machine. This example raises the idea that representing different constituents of science corpus can be helpful in achieving the goals mentioned in Section I.A.

This suggests that it would be helpful to take into account, different structures of the constituent parts of science corpus: a law of nature, for instance, differs from a research problem. Building blocks of science have different collections of characteristics based on their types. As a consequence, it will be convenient to represent relevant characteristics of each component. The following subsections propose the characteristics of scientific definitions, scientific laws, and research articles.

The knowledge of what the collection of characteristics of different parts of the science should be and how they are interrelated, is a knowledge about science, rather than being itself part of the science. We call this knowledge as meta-science and propose to be made explicit in a separate ontology. This ontology of meta-science (equivalently, meta-ontology of science) is briefly introduced in subsection E.

Nearly every single scientific document is already well structured. Traditionally, there are two major partitioning. One consists of divisions like chapters, sections and subsections. The other is made of parts of content such as definitions, theorems, research problems, methods and observations.

The first sort of parts are traditionally well identified. It is therefore quite normal to point to, say, a specific section of a given article. On the other hand, it is not common, if any, to formalize a part of the second kind. This lack of identification causes missing of a great deal of useful information. For instance, two distinct research articles A_1, A_2 may address the same research problem p, while there is no representative of p in the world-wide web, to which, A_1 and A_2 can point. Similarly, it is not a common practice to ask, as an instance, for the set of all research problems dealt with, during the past five years, in a given journal, nor is it possible to ask for the collection of different methods through which, researchers have approached a given goal. In order to enable machines to answer queries like these, it is necessary that scientific documents become inter-related based on their components, rather than solely based on their keywords and citations. Those components are, among others, definitions, theorems, research problems, methods, datasets, performance indicators and examples. This shows that the corpus of scientific literature has yet a good potential to get organized.

To put it briefly, we propose to meet the following considerations in the representation of science:

- Have different types of representatives for different types of constituents of scientific knowledge.
- Make explicit the domain-independent aspects of science, in a distinct ontology.

We explain, in the following subsections, the above mentioned idea in more detail. We do this by passing through some examples of several types of the constituents of science corpus, namely scientific terms, scientific laws and original research articles.

B. *Scientific Definitions*

In the present section we propose the idea that defining a scientific term in a formal manner can lead to at least one inference rule. This tends to equip the scientific glossary with a rich collection of inference rules.

We formalize a scientific term either by introducing the term as being primitive, or by defining the term based on other terms (avoiding cycles). For instance in physics (kinematics, to be exact) we may introduce the term *position* and the term *time* to be primitive, while *velocity* is defined to be the derivative of *position* with respect to *time*. *Velocity* is, therefore, based on two terms, namely *position* and *time* (see Fig. 1). We do not trace the dependency of terms to mathematics. This is why velocity is not considered to depend on the term *derivative*.

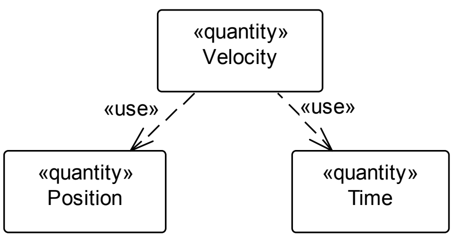

Fig. 1. Dependency among terms. The term *velocity* is defined based on *position* and *time*.

The meaning of the term velocity is derived from that of *position*, and that of *derivative with respect to time*, which is often shortened as the *rate of change*. This leads to the inference rules stated in TABLE I.

TABLE I. INFERENCE RULES CORRESPONDING TO THE DEFINITION OF VELOCITY.

About. x: particle.
(1) If $\mathbf{r}(t)$ is the position of x, then its velocity is $D[\mathbf{r}(t)]$ where $D[.]$ denotes derivative operator.
(2) If $\mathbf{v}(t)$ is the velocity of x and $\mathbf{r}_0$ its position at instance t_0 of time, then its position is $\mathbf{r}_0 + D^{-1}[\mathbf{v}(t)]$ where $D^{-1}[.]$ denotes antiderivative operator.

That is, given the position of a particle as a function **r** of time, then the velocity of the particle is found through the application of differentiation operator on **r**, provided that **r** is differentiable.

This is how the meaning of *velocity* is obtained through the terms *position* and *time* and the differentiation operator. The differentiation operator comes from mathematics and has indeed a formal meaning. The meaning, however, is part of mathematics rather than physics and therefore, is not considered in the ontology of physics.

It is notable that the dependency of the term velocity to differentiation operator, induces the dependency of physics, as a scientific domain, to mathematics. This point itself, is not part of physics or mathematics. It is rather a bit of knowledge *about* science. We consider this rule as a part of, what we call meta-science. More on this in Section II.E.

Theoretical terms of physics can be easily formalized, since physics is already well formal. In order to deal with less formal terms, we now consider the term *rock*, as defined in geology. Here is the definition:

> "(1) A consolidated or unconsolidated aggregate of mineral grains consisting of one or more mineral species and having some degree of chemical and mineralogic constancy. (2) In the popular sense, a hard, compact material with some coherence, derived from the earth."[17]
>
> "Rocks are aggregates of many different mineral grains, which are fused, cemented, or bound together."[18]
>
> "Rocks are aggregates of minerals – usually several, but sometimes only one or two."[18]

It is found from these quotations that rocks are aggregates of minerals. As a result, the term *rock* is based on the term *mineral* (Fig. 2). Let us now focus on this term. We begin by its definition in the literature.

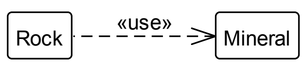

Fig. 2. The term *rock* is defined based on the term *mineral*.

> "With a few notable exceptions (water, mercury, opal), minerals are solid, inorganic elements or elemental compounds. They have definite atomic structures and chemical compositions which vary within fixed limits. Each and every quartz crystal, whether crystallized in a sandstone

vein, or in volcanic lava, possesses the same chemical and physical properties."[18]

Minerals are identified through their chemical composition and physical characteristics. Chemical compositions are given as chemical formulas. This is how *mineral*, as a scientific term, is based on *chemical composition*. The mineral named dolomite, for instance, has chemical formula $CaMg(CO_3)_2$.

A mineral is described through a set of characteristics, namely its *crystal system*, *habit*, *twining*, *cleavage*, *fracture*, *hardness*, *specific gravity*, *color*, *streak*, *transparency*, and *lustre* [18]. Each characteristic has a definite domain of values. The crystal system, for instance, receives one of the possible values *cubic*, *tetragonal*, *orthorhombic*, *monoclinic*, *triclinic*, and *hexagonal*/*trigonal*.

The meaning of the term *mineral* may be obtained from the meanings of chemical formulas, as well as that of physical characteristics mentioned above. The chemical formula brings its meaning from chemistry, as it indicates chemical atoms the mineral is composed of. Each physical characteristic has a specific meaning. The hardness, for instance have the following meaning, which can be formulated in an inference rule, as stated in TABLE II.

"Minerals with higher degree of hardness will scratch those lower in the scale, but not those higher in scale."[18]

TABLE II. THE INFERENCE RULE CORRESPONDING TO THE DEFINITION OF HARDNESS OF MINERALS

About. *x*,*y*: minerals.
Variables. h_x, h_y: hardnesses of *x*, *y* respectively
Inference rules.
$h_x > h_y$, if and only if *x* does scratch *y* and *y* does not scratch *x*.

Examples presented in this subsection show that formal representation of scientific terms gives rise to the ability of inference. Thus, the definition of every new term will reinforce, as expected, the inference machinery of the represented science. The next subsection shows that this is also true for scientific laws.

C. Laws of Nature

Laws of nature, also known as scientific laws, announce constraints or limitations among variables of some part of world. Newton's second law, for instance, puts a specific mathematical relation between two variables, i.e. the *net force* acting on, and *momentum* of any point mass. These can be represented as inference rules.

Therefore we may distinguish a law of nature through specifying (1) the specific part of world that the law models, (2) the variables it contains, and (3) possible inferences through its application.

In what follows, we consider a number of well-known laws of nature and reformulate them as inference rules. Let us consider, as exemplar laws of nature, Newton's laws of motion.

Newton's first law of motion: "The momentum of a point mass is constant when it is free of external forces."[19]

We recall that the momentum of a point mass is, by definition, the multiplication of its mass and velocity. Newton's first law of motion may therefore be characterized as in TABLE III. The part of world which is modeled by this law and the variables contained in the law are shown in Fig. 3.

TABLE III. CHARACTERISTICS OF NEWTON'S FIRST LAW OF MOTION

About. *x*: point mass.
Variables. $\mathbf{p}(t)$: momentum of *x*.
Inference rules.
x is free of external forces, if and only if $D[\mathbf{p}(t)] = \mathbf{0}$.

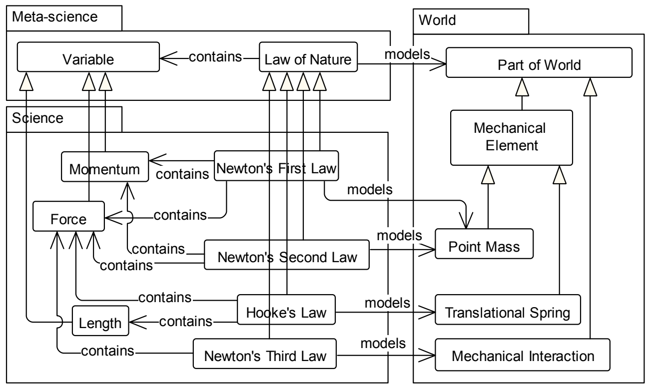

Fig. 3. Examples of scientific laws, parts of world modeled by those laws, and variables contained in laws. It is specified that a law of nature models part of world, and contains a number of variables. This regulation is part of metascience, while scientific laws and variables are part of science.

Newton's second law of motion: "The time rate of change of the momentum of a point mass m is equal to the net external force acting upon it."[19]

Newton's second law of motion: "The time rate of change of the momentum of a point mass m is equal to the net external force acting upon it."[19]

This law can be formalized easily. If we represent the momentum of the point mass by **p**, Newton's second law of motion can be formulated as $D[\mathbf{p}(t)]=\mathbf{F}(t)$, where $\mathbf{F}(t)$ represents the net external force acting upon the point mass (see TABLE IV. And Fig. 3). If, moreover, it is agreed upon that the net external force acting on a free point mass is zero, then the above-mentioned mathematical formula could also represent Newton's first law as well.

TABLE IV. CHARACTERISTICS OF NEWTON'S SECOND LAW OF MOTION

About. *x*: point mass.
Variables. $\mathbf{p}(t)$: momentum of *x*, $\mathbf{F}(t)$: net external force acting on *x*.
Inference rules.
(1) If $\mathbf{p}(t)$ is known to be the function $\mathbf{h}(t)$, then $\mathbf{F}(t) = D[\mathbf{h}(t)]$.
(2) If $\mathbf{F}(t)$ is known to be the function $\mathbf{k}(t)$, then $\mathbf{p}(t) = \mathbf{p}(t_0) + D^{-1}[\mathbf{k}(t)]$.

Newton's third law should be handled with care since it is customary to write it in a compact and somehow cryptic phrase:

Newton's third law of motion: "actio = reactio."[19]

The detailed form of this law would be like the following.

If an object O_1 applies force $\mathbf{F}_{12}$ to another object O_2, then O_2 applies force $\mathbf{F}_{21}$ to O_1 and $\mathbf{F}_{12}+\mathbf{F}_{21}=\mathbf{0}$. This is characterized in TABLE V. and Fig. 3.

TABLE V. CHARACTERISTICS OF NEWTON'S THIRD LAW OF MOTION

About. x: Mechanical interaction between objects O_1 and O_2. *Variables*. $\mathbf{F}_{12}(t)$: force applied by O_1 to O_2, $\mathbf{F}_{21}(t)$: force applied by O_2 to O_1. Inference rules. (1) If $\mathbf{F}_{12}(t)$ is known to be the function $\mathbf{h}(t)$, then $\mathbf{F}_{21}(t) = -\mathbf{h}(t)$. (1) If $\mathbf{F}_{21}(t)$ is known to be the function $\mathbf{k}(t)$, then $\mathbf{F}_{12}(t) = -\mathbf{k}(t)$.

Turning toward geology, here is the essential law used in the so-called *relative dating*, that is, "placing rocks in their proper sequence of formation"[20].

> The law of superposition: "This basic rule applies to materials that were originally deposited at Earth's surface, such as layers of sedimentary rock and volcanic lava flows. The law simply states that the youngest layer is on top, and the oldest layer is on the bottom (assuming that nothing has turned the layers upside down, which sometimes happens). Stated another way, a layer is older than the ones above it and younger than the ones below."[20]

Layers are widespread phenomena and are essentially identifiable through a wide area, sometimes even comparable to continents. Although they may be turned upside down, this can occur only locally. Besides, this is a situation which can be distinguished easily. As a result, this phenomenon can lead to the understanding that crustal disturbances have indeed caused the situation. Therefore, we could make the inference mentioned in TABLE VI. based on the evidence.

TABLE VI. CHARACTERISTICS OF THE LAW OF SUPERPOSITION IN GEOSCIENCES

About. L_1: layer, L_2: layer. *Variables*. a_i: age of L_i, d_i: depth of L_i (i=1,2). Inference rules. (1) If $d_1 < d_2$ and L_1 and L_2 are in order, then $a_1 < a_2$. (2) If $d_1 < d_2$ and $a_1 < a_2$, then L_1 and L_2 are in order. (3) If $d_1 < d_2$ and $a_1 < a_2$, then L_1 and L_2 are upside down.

The examples of laws of nature, presented in the current subsection, defend the idea that a law of nature would provide one or more inference rule. As a result, the represented collection of laws of nature provide a (potentially large) set of inference rules which can be employed as building blocks of scientific reasoning.

D. Original Research Articles

It is evident that clearly structured articles are more easily understood. Significant effort has been spent, to define suitable standards for the structure of research articles in different scientific disciplines. The well-known guideline IMRAD (Introduction, Methods, Results And Discussion) is worthy to mention. According to IMRAD, an original research article would mention the *research problem*, the *importance* of the considered problem, *related work*, the *method* used in the research, the *results* obtained by doing the research and *outcomes* of those results [21], [22].

The fact that the majority of research articles are organized in a standard structure, provides an opportunity, which is mentioned in what follows.

We propose to consider some parts of research articles, as identified entities. These parts include, at least, the research problem, the method used in the research, and the results. In the following we take a closer look.

There are a number of known research problems, stated in any research domain. Very often, each research problem is addressed in a large number of articles. As an instance, consider a sub-domain of artificial intelligence, called *pattern analysis and computer vision* in the literature. To name just several, out of many research problems in this domain we mention: *image classification* (e.g. addressed in [23], [24], *object detection* ([25], [26]), *object recognition* ([27], [28]) and *object segmentation* ([25], [29]).

If, as we propose here, one considers each research problem as an identified entity, and expect each research article to *point* to its addressed research problem, then each problem will be linked to all those research articles that address it. Fig. 4 shows this, as well as several other exemplar entities. In the upper part of the figure, Pattern analysis and computer vision (PACV) and representation and reasoning are shown to be sub-domains of artificial intelligence, which is itself a sub-domain of computer science. Four exemplar research problems stated in the domain PACV are shown below the domains. Below those problems, two well-known scientific journals are associated with their corresponding publishers, as well as their scientific domains, that is PACV. Finally, there are research articles aggregated in their corresponding journals and each associated with the research problem(s) it addresses.

We now focus on one of these articles, namely Toshev et al. (2012) [25]. The article addresses two research problems, *object detection* and *object segmentation*. Both problems belong to the scientific domain PACV as mentioned earlier and shown in Fig. 4.

In order to setup a method of object detection, a novel shape descriptor, called *chordiogram* is introduced in the article. An algorithm, named *boundary structure segmentation* (BoSS) is also provided as a solution to object segmentation.

What we propose here, is that the novel products of research, reported in research articles, be considered as identified objects. For instance in [25], the novelties are the *chordiogram* and the *BoSS* (Fig. 5).

Identifying research novelties as independent objects, can give rise to an enormous, yet structured collection of typed novelties. The *chordiogram* [25], for instance, is a statistical shape descriptor. Therefore, it may be considered as an instance of the class *shape descriptor*. Other shape descriptors, e.g. in [30] have also been introduced in the literature, making other instances of the same class. Moreover, other types of descriptors such as image descriptors [31], video descriptors [32], gradient field descriptors [33] have been introduced through other researches. Fig. 6 shows this exemplar part of the (potentially large) taxonomy of descriptors.

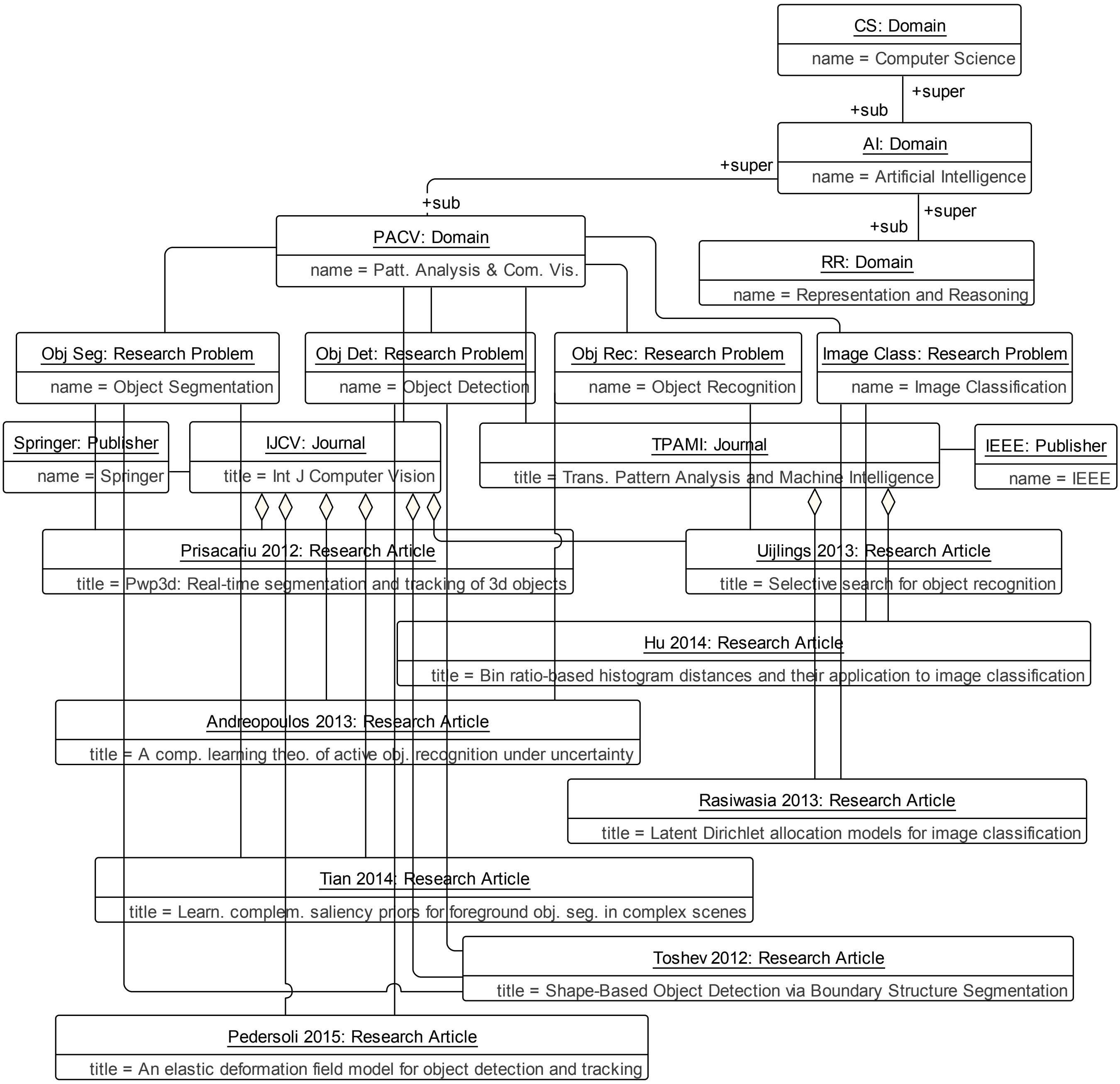

Fig. 4. An exemplar segment of the graph of science. The segment includes some research domains, research problems, scientific journals, and several research articles addressing the mentioned problems.

The standard of the structure of research articles, should not be considered as part of the science, rather, it is part of the meta-science. This standard may well differ in different scientific disciplines. In some research disciplines, such as computer science, there are research articles that present a computer algorithm as a solution to a research problem. That is, the algorithm is the main outcome of the performed research. The standard may expect that the algorithm proposed in the article be indeed executable as mentioned in section I.B about the Executable Paper Grand Challenge. This, together with other suitably stated requirements, can make it possible to test the algorithm proposed in the paper in order to reduce the needed effort to review the article, while making the provided knowledge more repeatable. Some of the other requirements to make this possible are, identification of typed datasets that scientists use in order to test their proposed algorithms, as well as formulation and identification of well-known performance indicators for the solutions of the addressed research problem.

The presented example shows that if the literature of science is organized so that different components of science such as research problems, research domains, proposed solutions, methods and datasets are identified and interrelated, then it will be possible to get answers to queries like “list all proposed solutions to the specific research problem p”.

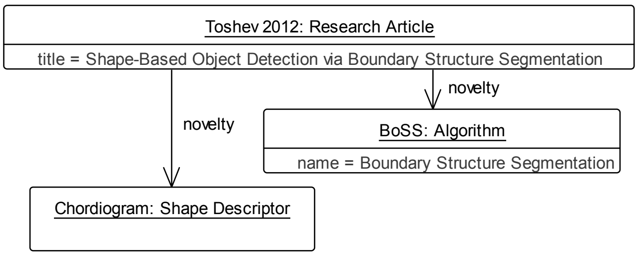

Fig. 5. The novelties reported in [25] are a shape descriptor called chordiogram, and an algorithm called BoSS.

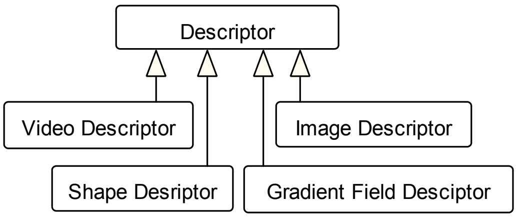

Fig. 6. Part of taxonomy of descriptors

E. Meta-science

Traditionally, scientific documents such as textbooks and research articles are inter-related by citation. Identification of different components of scientific knowledge (Section II.*D*) will produce a new-generation graph, namely, the graph of inter-related nodes of science corpus. These nodes will be of different sorts such as (but not limited to) *research problems*, *research methods*, *solutions*, *algorithms*, *observations* and *datasets*, and have contents in different formats like text, figures, images, data, mathematical or chemical formulas. We call this *the graph of science*.

Another sort of knowledge, the *meta-science*, is the knowledge *about* science. Meta-science contains domain-independent rules of scientific inference. In the following, an exemplar rule which can be properly included in the meta-science is presented.

Suppose that a scientific term t_1 in scientific domain d_1 is defined based on the term t_2 of another domain d_2 of scientific knowledge. This dependency induces the dependency of their corresponding domains (TABLE VII.). Note that the rule is about entities of science, rather than entities in the outside world, as is the case in natural sciences. This rule, as a bit of knowledge about science, is part of meta-science. We used it in Section II.B to deduce that *geology*, as a scientific domain, uses *chemistry*.

TABLE VII. DEPENDENCY AMONG SCIENTIFIC DOMAINS

If d_1, d_2 are scientific domains, t_1, t_2 are scientific terms, t_i is defined in d_i, (i=1,2) and the definition of t_1 uses t_2, then the domain d_1 uses the domain d_2.

We now present another bit of meta-science. The standards of the structure and content of a scientific document, an example of which being IMRAD mentioned in Section II.D is part of the knowledge about science, that is, the meta-science.

As yet another segment of meta-science, consider *scientific laws*. Scientific laws are knowledge about the world. They are parts of the body of scientific knowledge. In contrast, the considerations about scientific laws, are knowledge about science and therefore takes place in meta-science. In the following, we propose a minimal standard governing laws of nature.

1. A law of nature models a part of world. Examples: Newton’s second law of motion is about point masses. Newton’s third law of motion points to mechanical connections between two (or more) entities.
2. A scientific law contains a definite collection of variables (e.g. force and translational momentum in Newton’s second law and forces in Newton’s third law of motion).
3. A scientific law makes one or more inferences when applied (Examples: the equalities in Newton’s second and third laws of motion).

The requirements 1 and 2 above are shown in Fig. 3. Note that it is contained in what we call *meta-science* (equivalently, *meta-ontology of science*). The structure of TABLE III. TABLE IV. TABLE V. and TABLE VI. comply with the above mentioned standard.

III. Evaluation

The competence questions stated in section I.A can be answered using the proposed representation framework. Answer to question 1 is clearly supported by the representation of scientific definitions of variables (section II.B) and that of laws of nature (section II.II.C).

The point that a law of nature *should* explicitly target a part of world, answers the second competing question mentioned in section I.A,

Our proposal explicitly models the fact that a given variable v_1 uses another variable v_2 to be formally defined. This supports question 3 of section I.A. TABLE VII shows how a reasoner relying on the meta-ontology of science may conclude that a scientific domain uses another scientific domain to define its variables (question 4 in section I.A). Other inference rules may be developed by using the same idea. For instance, if the domain d_1 contains a method m_1, which makes use of a method or a variable provided by another domain d_2, then d_1 relies on d_2.

Section II.II.D showed that identification of scientific research problems and proposed solutions brings the ability of answering queries like the one stated in competing question 5. While it is presented in [16] for different parts of a research article, we defend the idea for components of scientific knowledge outside articles as well.

IV. Conclusion

The examples presented in Section II make it evident that the representation scheme proposed here makes improvements as follows:

1. The ontology of science is equipped with a rich collection of inference rules, which supply the building blocks of any potential reasoning about scientific knowledge. Those rules come from scientific definitions and scientific laws, among others.
2. The above mentioned inference rules point to their corresponding relevant parts of world, making it easier to properly employ them in automation of scientific reasoning.
3. Identification and adding structure to the building blocks of scientific knowledge as well as scientific problems and research empowers the semantic web with the ability of answering queries about different constituents of science corpus.
4. Meta-science factors out the common aspects of scientific knowledge and reasoning.

The proposed structure of laws of nature provides an important possibility. In the real world, elements may participate in objective relations to constitute complexes. Let us consider elements $e_1,\ldots,e_n$ have participated in a relation r. This makes a constraint among their variables $v_1,\ldots,v_n$ respectively. This constraint is modeled through a dedicated law of nature. As an instance, when two or more point objects participate in a mechanical connection, each object receives a force. The third Newton's law of motion states that the forces received by different individual objects add up to zero. That a law of nature targets a part of world, provides the possibility for a scientific reasoner to combine laws of nature governing individual elements with those governing objective relations in order to develop models of complex entities. This potential goes far beyond the computations provided by a single law of nature, as served in Wolfram Alpha website [11] mentioned in section I.B. Further work is required to develop scientific models of a given complex entity with known structure, using laws of nature targeting constituent elements and their inter-relations.

Meta-science includes domain-independent aspects of scientific knowledge. It is well expected that scientific reasoning may be decomposed into a domain-specific part which raises from individual, numerous, professional scientific laws and definitions, and a domain-neutral component that reflects the rational, strategic aspects of reasoning which is more or less common among scientific domains, but yet highlights the main differences among methodologies of major disciplines e.g. natural versus engineering sciences.

References

[1] T. Heath and C. Bizer, "Linked data: Evolving the web into a global data space," Synth. Lect. Semant. web theory Technol., vol. 1, no. 1, pp. 1–136, 2011.

[2] F. M. Suchanek, G. Kasneci, and G. Weikum, "Yago: a core of semantic knowledge," in Proceedings of the 16th international conference on World Wide Web, 2007, pp. 697–706.

[3] M. Banko, M. J. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni, "Open Information Extraction from the Web.," in IJCAI, 2007, vol. 7, pp. 2670–2676.

[4] G. Weikum and M. Theobald, "From information to knowledge: harvesting entities and relationships from web sources," in Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, 2010, pp. 65–76.

[5] M. Schmitz, R. Bart, S. Soderland, O. Etzioni, and others, "Open language learning for information extraction," in Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012, pp. 523–534.

[6] F. M. Suchanek, M. Sozio, and G. Weikum, "SOFIE: a self-organizing framework for information extraction," in Proceedings of the 18th international conference on World wide web, 2009, pp. 631–640.

[7] D. L. Cook, M. L. Neal, F. L. Bookstein, and J. H. Gennari, "Ontology of physics for biology: representing physical dependencies as a basis for biological processes.," J. Biomed. Semant., vol. 4, p. 41, 2013.

[8] I. Niles and A. Pease, "Towards a standard upper ontology," in Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001, 2001, pp. 2–9.

[9] R. G. Raskin and M. J. Pan, "Knowledge representation in the semantic web for Earth and environmental terminology (SWEET)," Comput. Geosci., vol. 31, no. 9, pp. 1119–1125, 2005.

[10] J. Zhong, A. Aydina, and D. L. McGuinness, "Ontology of fractures," J. Struct. Geol., vol. 31, no. 3, pp. 251–259, 2009.

[11] "Wolfram Alpha," WolframAlpha. 2016.

[12] L. N. Soldatova and R. D. King, "An ontology of scientific experiments," J. R. Soc. Interface, vol. 3, no. 11, pp. 795–803, 2006.

[13] L. N. Soldatova, A. Clare, A. Sparkes, and R. D. King, "An ontology for a Robot Scientist," Bioinformatics, vol. 22, no. 14, pp. e464–e471, 2006.

[14] "Executable Paper Grand Challenge." 2011.

[15] P. Van Gorp and S. Mazanek, "SHARE: a web portal for creating and sharing executable research papers," Procedia Comput. Sci., vol. 4, pp. 589–597, 2011.

[16] A. de Waard and G. Tel, "The ABCDE Format Enabling Semantic Conference Proceedings.," in SemWiki, 2006.

[17] M. G. H. Education, Dictionary of Geology & Mineralogy. McGraw-Hill Education, 2003.

[18] C. Pellant, Rocks and Minerals, Doling Kindersly Handbooks. Dorling Kindersley Ltd., London, 2000.

[19] D. Gross, W. Hauger, J. Schröder, W. A. Wall, and S. Govindjee, Engineering Mechanics 3: Dynamics, vol. 3. Springer Science & Business Media, 2014.

[20] F. K. Lutgens, E. J. Tarbuck, and D. Tasa, Essentials of Geology. Prentice Hall, 2012.

[21] P. K. R. Nair and V. D. Nair, "Organization of a Research Paper: The IMRAD Format," in Scientific Writing and Communication in Agriculture and Natural Resources, Springer, 2014, pp. 13–25.

[22] G. S. Patience, D. C. Boffito, and P. Patience, Communicate Science Papers, Presentations, and Posters Effectively. Academic Press, 2015.

[23] W. Hu, N. Xie, R. Hu, H. Ling, Q. Chen, S. Yan, and S. Maybank, "Bin Ratio-Based Histogram Distances and Their Application to Image Classification," Pattern Anal. Mach. Intell. IEEE Trans., vol. 36, no. 12, pp. 2338–2352, 2014.

[24] N. Rasiwasia and N. Vasconcelos, "Latent dirichlet allocation models for image classification," Pattern Anal. Mach. Intell. IEEE Trans., vol. 35, no. 11, pp. 2665–2679, 2013.

[25] A. Toshev, B. Taskar, and K. Daniilidis, "Shape-based object detection via boundary structure segmentation," Int. J. Comput. Vis., vol. 99, no. 2, pp. 123–146, 2012.

[26] M. Pedersoli, R. Timofte, T. Tuytelaars, and L. Van Gool, "An elastic

deformation field model for object detection and tracking," Int. J. Comput. Vis., vol. 111, no. 2, pp. 137–152, 2015.

[27] A. Andreopoulos and J. K. Tsotsos, "A computational learning theory of active object recognition under uncertainty," Int. J. Comput. Vis., vol. 101, no. 1, pp. 95–142, 2013.

[28] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, 2013.

[29] V. A. Prisacariu and I. D. Reid, "Pwp3d: Real-time segmentation and tracking of 3d objects," Int. J. Comput. Vis., vol. 98, no. 3, pp. 335–354, 2012.

[30] V. P. D. Kumar and T. Tessamma, "Performance study of an improved Legendre Moment Descriptor as region-based shape descriptor," Pattern Recognit. Image Anal., vol. 18, no. 1, pp. 23–29, 2008.

[31] D. V Sorokin and A. S. Krylov, "A projection local image descriptor," Pattern Recognit. Image Anal., vol. 22, no. 2, pp. 380–385, 2012.

[32] B. Solmaz, S. M. Assari, and M. Shah, "Classifying web videos using a global video descriptor," Mach. Vis. Appl., vol. 24, no. 7, pp. 1473–1485, 2013.

[33] V. V Myasnikov, "A model-based gradient field descriptor as an efficient tool for recognizing and analyzing digital images," Pattern Recognit. Image Anal., vol. 25, no. 1, pp. 60–67, 2015.