Knowledge on the Web
Overview Paper

Marc Vornetran `marc.vornetran@gmail.com`

[1] Universiteit van Amsterdam (11569565)
[2] Vrije Universiteit Amsterdam (2629672)

Abstract. This paper provides a comprehensive overview of representing and sharing knowledge on the web. It takes relevant literature in the field of knowledge representation and the Semantic Web into account. The overview itself is twofold and starts out by presenting the groundwork involved in representing knowledge in digital environments. Ontologies help us define reality while logic enables reasoning in this context. Building onto the examined concepts the second part focusses on the Semantic Web and its goal to reshape how humans and machines share knowledge on a global scale. While a lot of time and resources has been spent on developing the global web of data the full potential, as envisioned by Tim Berners-Lee [3], remains yet to be unleashed.

Keywords: Knowledge, Media, Semantic Web, Ontologies, Linked Data, Logic, Reasoning, Controlled Natural Language

1 Introduction

Long before the invention of computers humans were accustomed to express knowledge in physical media. Primitive examples include cave drawings as a mean to preserve knowledge but also allow others to interpret this representation. More sophisticated approaches are books which represent knowledge printed on paper and expressed in a specific natural language. Technical advancement allows for the rapid production of books and thereby an increase in available knowledge. Having a growing amount of information at hand also requires ways to organise and structure this pool. Otherwise accessibility would diminish under an expanding mass of unstructured information. Libraries can be seen as a solution to organise books into an intentionally structured collection [9].

With the invention of computers the era of digitalisation began. Knowledge representation is no longer restricted to physical records but more dominantly focuses on the digital context. However, a true revolution of knowledge sharing happened due to the widespread adoption and success of the *World Wide Web* (WWW). It allows humans to contribute information to a global web while at the same time having this massive information source right at the fingertips at any time. Humans are easily able to extract meaningful information from the web by integrating many sources. Unfortunately, machines are unable to take the same advantage as loosely coupled, ambiguous and unstructured information

is a challenge to process automatically. Thus, the Semantic Web sets out to fix this issue. It envisions a global web of data which is accessible for both humans and machines.

> "Properly designed, the Semantic Web can assist the evolution of human knowledge as a whole." [3]

This paper covers the groundwork to represent knowledge and subsequently explains how these concepts play fundamental roles in the development of the Semantic Web. The first part, section 2, explores the concepts connected to representing knowledge using logic, ontologies and relationships. These components are picked up by section 3 to examine the building blocks of the Semantic Web: RDF, OWL and Linked Data.

2 Knowledge Organisation & Representation

2.1 Knowledge Organisation

As the amount of information only continues to grow exponentially the organisation of knowledge has become a central task. Whereas knowledge was traditionally primarily stored in books and other physical records the invention and widespread adoption of the internet has sparked an era of digitalisation. Nowadays both physical and digital resources need to be organised and made accessible.

To name a specific example, non-profit organisations often find themselves on a middle ground between using paper records and digital information management systems [27]. They often resort to so called *homebrew databases*[27] to meet their resource constraints, requirements and workflows. Despite having substantial information management needs because of the large amount of volunteers these organisations usually lack the technical expertise or resources for complex and sophisticated systems. On one hand this highlights the importance of adequate, efficient knowledge organisation systems but on the other hand also indicates room for improvement in this area.

Another example of this ongoing digital transformation is the *Netherlands Institute for Sound and Vision* in Hilversum. It is an archive for over *70%* of Dutch audiovisual content from 1899 to today.[3] Their main goal is not merely the preservation of cultural heritage but also organising, categorising and opening up access to the vast catalogue of media content. Access is not restricted to a specific target group but rather aimed at a broad range of fields: Education, science, individuals, creative professionals and media or business organisations. Furthermore, the institution, as one of the largest in Europe[4], also founded the

---

[3] `http://www.beeldengeluid.nl/en/about`, Accessed: 27-10-2017

[4] `https://en.wikipedia.org/wiki/Netherlands_Institute_for_Sound_and_Vision`, Accessed: 27-10-2017

Open Images[5] initiative to create a platform for others to consume but also to contribute their own material.

These examples are only a small subset of many organisation systems. They feature the main responsibilities of *Knowledge Organisation Systems (KOS)*: (1) Imposing order and structure on a collection of resources and (2) providing a set of supported interactions [9]. Resources hereby can be distinguished into four categories. First, a resource is either tangible (physical) or intangible (digital). The second division depends on whether a resource is the object itself (primary) or if it holds information about another object (description, meta-data). Most of the time, the value of primary resources is enhanced by providing secondary meta-data about them. This allows humans and machines to better understand a primary resource and make efficient use of it.

2.2 Knowledge Representations

Machines require *Knowledge Representations (KR)* as substitutes for real world information in order to be able to solve tasks. There are many different forms of these representations: Natural language text (Section 3.3), logic (Section 2.4), visual graphics or entries in a database. Unstructured information like text and graphics are usually perfectly fit for humans to understand. However, computers are better equipped to work with structured information like logic and relational databases. Especially logic is the fundamental building block for *reasoning* (Section 2.6) in the field of *Artificial Intelligence (AI)*.

According to Davis et al. [6] knowledge representations take upon *five* distinct, crucial roles:

1. It is a *surrogate* (substitute) for objects of the real world, thereby not taking direct action but rather using *reasoning* to draw conclusions.
2. As a *set of ontological commitments* determines how to think about the world.
3. It is a *fragmentary theory of intelligent reasoning* based on a fundamental mental conception. This reasoning produces sets of inferences which it either recommends or sanctions.
4. The role of a *medium for pragmatically efficient computation* organising information to produce recommended inferences.
5. It is a *medium of human expression* or a language in which humans describe the world.

Inference, in a broader sense, refers to a logical tool to generate new conclusions based on existing premises, expressions and facts. It is important to keep in mind that knowledge representations are inherently imperfect. They are merely a substitute for the entity (tangible, intangible) of the real world. The high fidelity of reality cannot be fully captured or even assumptions about reality can be wrong from the beginning. This inaccuracy of knowledge representations can lead to erroneous conclusion despite being based on correct reasoning [6].

---

[5] `https://openimages.eu`, Accessed: 28-10-2017

Although machine learning and statistic techniques have taken a central role in artificial intelligence software Shoham [25] stresses the importance of knowledge representations. Applying knowledge representation and philosophy to a problem reveals valuable insights about the problem space and generates a conceptual vocabulary to argue about it. Creating a good user experience is not possible or more difficult without a solid internal structure. Once this is set, machine learning and statistic algorithms are able to build on top to design a richer user experience [25].

Furthermore, knowledge representations are heavily used in the process of *Knowledge Modeling*. It aims to produce knowledge models which are interpretable by computers. Knowledge acquisition could then be deemed to be the extraction and transfer of knowledge from a source, e.g. the mind of a domain expert, to a digital representation. However, this is somewhat misleading as internal knowledge cannot be simply accessed and mutated into a different representation; knowledge is not a storable matter [5]. The term rather describes the process of developing completely new computer models in a domain which is a creative and difficult task.

In addition to that, the development of computational knowledge representations is further complicated by the *semantic gap*. Hein's definition describes it as "the difference in meaning between constructs formed within different representation systems" [15]. Representation systems, for example, refer to the translation between natural language (high-level) and formal language (low-level). Formalisms in logic (Section 2.4), ontologies (Section 2.5) and controlled natural language (Section 3.3) can be regarded as means to overcome the semantic gap.

2.3 Relationships

Hierarchies usually build on *relationships* between the entities within. Relations are useful to describe the connections between the parts of an object or larger complex systems. There are many different kinds of hierarchical relations, although two of them are commonly used: Aggregation (part of whole) and generalisation (superclass and subclasses).

Odell identifies *six* different types of aggregation which is also referred to as *composition* [24]. Composition describes the links between parts of an object. All parts together form the object itself as a whole. The *six* kinds of composition according to Odell [24] are as follows:

1. *Component-integral object composition:* The configuration of the parts is essential for the function and structure of the object they support. Parts are separate objects of their own but they must be well arranged with each other to form the overall entity. *(Example: A display is part of a smartphone)*
2. *Material-object composition:* Invariant version of the previous composition. Parts of the object cannot be removed. They are what an object is made of. *(Example: Coca-Cola is partly water)*
3. *Portion-object composition:* Parts of the object are of the same kind as the composition itself. The chunks are homeomeric. *(Example: A drop of water is still water)*

4. *Place-area composition:* Invariant version of the previous composition. Homeomeric pieces of the object cannot be removed. *(Example: Amsterdam is part of The Netherlands)*
5. *Member-bunch composition:* Parts of the object belong to a collection. They bear no functional or structural relationship to each other. *(Example: A drop of water is part of rain)*
6. *Member-partnership composition:* Invariant version of the previous composition. Members of the partnership cannot be removed without terminating it. *(Example: Two partners form a marriage)*

Winston et al. [29] define *three* distinctive properties to determine the type of composition:

1. *Configuration* states whether parts of the object have a functional or structural meaning towards each other (Composition 1 - 4).
2. *Homeomerous* parts are of the same kind as the object they form (Composition 3 & 4).
3. *Invariance* declares that parts of the object cannot be removed without terminating the composition (Composition 2, 4 & 6).

Clearly defining the type of relationship between entities is especially useful for creating *Ontologies* (Section 2.5) and also relevant for *Linked Data* (Section 3.1 in the *Semantic Web* (Section 3).

2.4 Logic

Logic is concerned with the correct and valid principles of reasoning. The human psychology behind reasoning is a different field of study. It is based on premises, conclusions and inference. Inference, in a broader sense, is the step of drawing conclusions from a set of premises. The formal and structured nature of logic enables machines to automatically execute different types of inference tasks. This is a central responsibility of knowledge-based systems. They can be used to verify the consistency of argument chains, autonomously answer questions (See also Section 2.6) or conduct deductive reasoning (generation of new conclusions). Furthermore, formal logic definitions are a part of reusable and shareable ontology (Section 2.5) specifications.

A set of premises and a conclusion form an *argument*. The argument is considered to be *valid* if it is impossible for all premises to be *true* and the conclusion to be *false* [13]. It does not matter if the premises of an argument are true for it to be valid.

However, a *sound* argument is valid and also requires its premises to be *true*. Proving a sound argument involves checking all premises to be correct and whether the conclusion follows the premises.

Valid argument	**Sound argument**
All humans like dogs.	Breathing means you are alive.
Humans who like dogs are friendly.	You are breathing.
Therefore, all humans are friendly.	Therefore, you are alive.

The previous examples are written using quantified *syllogisms* (hence the "all" keyword). This simple logic notation dates back to the definition by Aristotle in around 350 B.C. There are more formalisms and notations which vary in their expressiveness, complexity, decidability and efficiency. Formalisms should be selected depending on their characteristics and the intended usage context. The following list is ordered going from very simple to complex:

– Propositional logic (inexpressive, simple, decidable & efficient)
– Syllogisms
– Description Logic, OWL (See also Section 3)
– First-order logic
– Higher-order logic (expressive, complex, undecidable & inefficient)

Nowadays *first-order logic* is the most commonly used formalism and acts as the core for modern logic [7].

2.5 Ontologies

The preceding section introduces logic formalisms as a way to reason about the world. While this enables reasoning in software it lacks a formal specification to share and reuse knowledge in a domain. *Ontologies* are designed to be a solution for this problem by defining a collection of terms and their relationships [3]. Philosophy describes an ontology as the theory of what exists in reality. Information technology uses ontologies as specification documents for a hierarchy of terms. Usually ontologies contain a taxonomy including inference rules. The taxonomy defines classes of objects in a hierarchy using relations (See also Section 2.3). Classes have a set of properties associated with them. Their corresponding subclasses inherit these properties and may extend the superclass [3].

Designing ontologies is a creative process usually performed by humans. In order to create ontologies for knowledge sharing and interoperability design decisions should be based on objective criteria and not subjectiveness [10]. Gruber [10] proposes *five* principles for designing ontologies:

1. *Clarity:* Definitions of terms should be objective and as complete as possible. If possible a representation in logic formalisms should be preferred. Documentation for definitions should be provided in natural language.
2. *Coherence:* Inferences should be consistent with the definitions, as well as the underlying axioms. Coherence also applies to documentation in natural language.
3. *Extendibility:* An ontology should anticipate the use of shared vocabulary. It should be possible to extend and specialise an ontology without altering the existing definitions.
4. *Minimal encoding bias:* The dependency on specific symbol-level encoding should be avoided. Design choices should not be based on notation convenience or ease of implementation.

5. *Minimal ontological commitment:* Ontological commitment (See also Section 2.2) should be kept at a minimum. The amount of claims about the world being modeled should be the lowest number still sufficient for support. In other words, the ontology should contain only essential assumptions about the world.

The provided principles act as a foundation for designing shareable and reusable ontologies in knowledge-based systems. Nevertheless, Gruber [10] points out that the nature of design processes involves making compromises between criteria. However, not all goals are mutually exclusive and can still be useful to guide the definition of ontologies.

2.6 Knowledge Generation

Knowledge Generation can be divided into formal research and informal activities or interactions between humans [23]. Both types involve external acquisition and internal creation of knowledge[6].

The following sections focus on *two* specific applications of knowledge generation. First, *Reasoning* (Section 2.6) is heavily used in the *DeepQA* (QA = Question Answering) architecture designed by researchers at IBM. Second, *Crowdsourcing* (Section 2.6) is an alternative approach to question answering. It utilises wisdom of crowds for reducing errors in decision making.

Reasoning Ferrucci et al. [8] describe how they, as researchers at IBM, set out to develop an architecture for *Question Answering (QA)*. The overall architecture is named *DeepQA* and the concrete implementation gained popularity as *Watson*. Waton's goal is to compete in the U.S. quiz show *"Jeopardy!"* as a real-time, digital contestant. Participants of the TV show have to answer questions in broad general knowledge while also responding fast. As the contestants are usually humans questions are formulated in natural language (Section 3.3) which is difficult to understand for machines.

The paper *Building Watson* [8] states that combining various techniques, algorithms and implementations for tackling the challenges led to better performance than committing to a single, specific solution. Challenges include: Natural language processing, source identification, hypothesis generation, ranking and finally scoring. Deploying large-scale parallelism allows the system to evaluate multiple interpretations and hypotheses in a short amount of time. One of the most important design decisions is the use of probabilities and confidence. Watson evaluates a wide range of question interpretations and a large number of hypotheses. It bases its evaluation on shallow and deep knowledge; not restricting itself to a subset. Components of the system do not commit to answers which could lead to missed correct evaluation paths. Confidence scores are attached to question interpretation and follow up answers.

---

[6] `https://www.igi-global.com/dictionary/knowledge-generation/45440`, Accessed: 29-10-2017

After *three* years of research and development Watson is able to compete at human expert level [8]. Rapid experimental prototyping and constant, thorough evaluation of the components performance is a key strategy leading to the success of the platform.

Crowdsourcing utilises *human computing* to answer questions and solve tasks which machines may be unable to. The term *human computation* was coined by Luis von Ahn [20] and regards human brains as processors in a distributed system. Each brain solves small parts of a larger computational task at hand. It is a useful technique for decision making by using the majority answer in crowds to reduce the chance of errors.

CrowdTruth[7] is a framework for crowdsourcing ground truth data with a focus on *gold standard* data in medical applications. Gold standard refers to the best available data under given conditions [26]. The project is developed in collaboration between researchers from the Network Institute[8] of "Vrije Universiteit (VU)" Amsterdam and IBM Netherlands[9].

Two researchers of the projects team, Lora Aroyo and Chris Welty, identity and debunk *seven* myths in the field of crowdsourcing in their paper *Truth is a Lie* [1]. They reject the fallacy of an ideal single truth. As a result *crowd truth* is proposed which captures human subjectiveness and a range of valid interpretations.

3 Semantic Web

The world wide web has revolutionised the ability of humans to access and share knowledge on a global scale. However, computers are only able to parse the content of web pages on a technical level. While humans can answer questions using single queries or combining multiple ones machines get stuck when trying to retrieve meaning from web pages. The *Semantic Web* aims to evolve the current web into a state where both humans and machines can autonomously carry out sophisticated tasks [3]. It is therefore not a separate web but rather the next evolutionary step which allows to embed and express meaning. This enables the potential for a shift in responsibility from the user to intelligent software agents. These agents would be able to carry out complex queries on their own and presenting meaningful information to users.

Horrocks [16] also criticises this deficit in the current web. Despite all the necessary information being available it requires integrating multiple sources. Current solutions are usually specific software solutions customised for a problem. He identifies the focus of content presentation as one of the main obstacles for the Semantic Web to overcome. Most information is unstructured and not explicitly linked to other resources.

---

[7] `http://crowdtruth.org`, Accessed: 29-10-2017
[8] `http://www.networkinstitute.org`, Accessed: 29-10-2017
[9] `https://www.ibm.com/planetwide/nl/`, Accessed: 29-10-2017

The *World Wide Web Consortium* (W3C)[10] is responsible for developing open web standards. They introduced the *Resource Description Framework* (RDF)[11] as a simple yet flexible language for describing resources and their relationships on the web [16]. One of its key feature is the usage of *Internationalized Resource Identifiers* (IRI)[12] to reference other resources. This mechanism is related to hyperlinks, also called *Uniform Resource Locators* (URL), which allow the navigation between different web pages. RDF itself is a labeled, directed graph built using triples [16]. Triples consist of *three* components: (1) subject, (2) predicate and (3) object. The predicate describes a binary relationship between the subject and object.

While RDF is a rather simple ontology language it is inherently limited in its capabilities by design. The Semantic Web community recognised the need for more expressive ontologies and thus the *Web Ontology Language* (OWL)[13] was created. OWL adds richer semantics to RDF resources using schemas. It allows detailed description of classes and their corresponding properties. Being based on Description Logic (Section 2.4) it also includes mechanisms to add restrictions to classes, properties and relationships.

Berners-Lee [3] as well as Heath and Bizer [14] argue that *Linked Data* (Section 3.1) is the true strength of the Semantic Web. Providing structured data, opening access to it and linking it to other data resources transforms the web into a global web of data. With an emphasis on being usable not only by humans but also by intelligent software agents in a meaningful way.

3.1 Linked Data

In a consolidated effort by Google, Microsoft, Yahoo and Yandex the platform *Schema.org*[14] was founded. The goal is to define, maintain and promote reusable vocabularies for structured data on the web. Guha et al. [11] analyse the history of structured data and identify *three* goals for future standardisation efforts:

1. *Easy participation:* Publishers and developers should be able to participate without compromising existing tools and workflows. Complexity should stay with the smallest amount of users.
2. *Avoid long specifications:* Good documentation is essential for widespread adoption. Especially examples can be useful as recipes to start out.
3. *Incremental complexity:* Complexity should increase progressively alongside adoption trends. If the need arises more complex layers can be added over time.

Wikipedia[15] is easily one of the most popular knowledge bases on the web. It makes extensive use of crowdsourcing (Section 2.6) to gather knowledge spanning

---

[10] `https://www.w3.org`, Accessed: 30-10-2017

[11] `https://www.w3.org/standards/techs/rdf#w3c_all`, Accessed: 30-10-2017

[12] `https://www.ietf.org/rfc/rfc3987.txt`, Accessed: 30-10-2017

[13] `https://www.w3.org/standards/techs/owl#w3c_all`, Accessed: 30-10-2017

[14] `http://schema.org`, Accessed: 30-10-2017

[15] `https://www.wikipedia.org`, Accessed: 30-10-2017

all topics in various languages. While it may not be visible on the surface *two* projects aim to enable the potential of Wikipedia as a source for Linked Data [28].

Wikidata[16] acts as the central storage of structured data for Wikipedia. The projects purpose is to extract and store structure data from Wikipedia which is otherwise scattered in multiple articles and duplicated in many languages [28]. In return Wikipedia can access Wikidata as a repository to display information. Furthermore, this platform also allows machines to make efficient use of its knowledge.

Another effort to build the open web of data is *DBpedia*[17]. Similar to Wikidata it is a crowdsourced attempt to extract structured data from Wikipedia and making it available on the web [4]. All entries can be dereferenced using their globally unique identifier. This results in a detailed RDF representation which applications utilise to provide meaningful information to their users.

Despite strong community efforts to build an open web of linked data Beek et al. [17] state that the current Semantic Web falls short to meet the initial goals. It is not yet machine-processable nor are applications able to autonomously traverse it and access different sources. They propose a fundamentally different approach for realising these goals called *LOD Laundromat*[18]. In contrast to the Semantic Web's nature it acts as a centralised hub for querying cleaned, structured data.

3.2 User Requirements

Benson and Karger [2] conducted research on how end-users publish data on the web. They differentiate between needs of expert and novice users. Current standards, like RDF and OWL (Section 3), are of technical nature and better suited for professional developers and engineers. Looking back at the goals layed out by Guha et al. [11] (Section 3.1) these tools do not necessarily meet the requirements of easy participation and incremental complexity. Their research finds that novice data authors are stuck with either publishing their data as unstructured text (e.g. blogs, wikis) or handing it over to big companies (e.g. Epicurious, Facebook) which transform it into a rich browsing experience [2]. Furthermore, the study explores the JavaScript[19] framework *Exhibit*[20] as a tool to publish interactive, structure data visualisations on the web. Exhibit's focus lays on allowing all end-users, regardless of their technical background, to contribute their data to the Semantic Web.

---

[16] `https://www.wikidata.org/wiki/Wikidata:Main_Page`, Accessed: 30-10-2017

[17] `http://wiki.dbpedia.org`, Accessed: 30-10-2017

[18] `http://lodlaundromat.org`, Accessed: 30-10-2017

[19] `https://developer.mozilla.org/en-US/docs/Web/JavaScript`, Accessed: 30-10-2017

[20] `http://www.simile-widgets.org/exhibit/`, Accessed: 30-10-2017

3.3 Natural Language

The majority of the web's presentational content is written in natural language as it is intended to be consumed by humans. Computers are currently unable to easily extract meaning from content in natural language because of the complexity and ambiguity involved. However, automated semantic analysis is possible using *Controlled Natural Languages* (CNL). A single, succinct definition for CNL does not exist. Kuhn [19] describes CNL as a constructed language based on *one* natural language. It is more restrictive in terms of the lexicon, syntax and semantics. Most natural properties are preserved to strike a middle ground in usability for humans and machines.

AceWiki[21] is a semantic wiki prototype built around the use of CNL. Contributors are able to formulate statements in *Attempto Controlled English* (ACE) [18]. This subset of natural English is automatically processable by the system but provides a shallow learning curve for users. As a result AceWiki allows the creation of a semantic knowledge base without requiring deep expertise in the Semantic Web.

Another approach to solve the challenge of natural language for machines is *WordNet*[22]. It is a lexical database for English collecting words and their associated meaning [21]. Sets of synonyms are linked through semantic relations thereby determining word definitions. In its core WordNet aims to solve the challenge of *polysemous* words for systems processing natural language [21]. Polysemous words have more than *one* meaning (Poly: Greek "many"[23]) which requires computers to understand the author's intended meaning.

3.4 Provenance

Facilitating shared knowledge through the Semantic Web has resulted in a need for *Provenance*. The paper by Moreau and Groth [22] examines *four* use cases: (1) Quality assessment, (2) compliance, (3) cataloguing and (4) replay.

First, *quality assessment*, as the name implies, is concerned with measuring content quality according to a range of criteria: Trustworthiness, careful research, use of expert knowledge and whether it is based on correct information. Second, *compliance* deals with the acquisition and production of data following a set of rules and guidelines. Third, *cataloguing* refers to the organisation of information to make it both accessible and reusable. Finally, *replay* describes the usefulness of having an organisational memory. This would allow the reproduction of data processes to find flaws or errors.

3.5 Web Search

Searching on the web is commonly solved using a keyword-based input and presenting a list of results to the user. For a long time the results were displayed

---

[21] http://attempto.ifi.uzh.ch/acewiki/, Accessed: 30-10-2017

[22] https://wordnet.princeton.edu, Accessed: 30-10-2017

[23] http://www.dictionary.com/browse/poly-, Accessed: 30-10-2017

as a primitive collection of links. The collection is ordered according to internal relevance algorithms of search engines. In some cases (e.g. Google[24]) links from advertisers take precedence to generate revenue.

Cross-company research by Haas et al. [12] from Microsoft, Yahoo and Facebook shows that users prefer enhanced search results and even attribute more relevance to these elements. Enhanced hereby refers to the integration of images, videos, further links, key-value pairs of data and other interactive elements into search results. This is possible by performing information extraction using Semantic Web techniques. Their research also indicates that efficient generation of enhanced search results is achievable while also having a meaningful impact in terms of user impressions. Despite structured information being scarcely scattered across the web the majority of user queries focusses on search results with a high potential for enhancement [12].

4   Conclusion

The Semantic Web builds on concepts from knowledge representation which have been researched even before Tim Berners-Lee [3] envisioned it. Both RDF and OWL are languages to create and share ontologies (Section 2.5) on the web. OWL is meant to overcome the limitations of RDF having its basis in description logic (Section 2.4) and forming relationship hierarchies (Section 2.3).

A lot of time and resources have been spent by communities, researchers and companies to achieve the goals of the Semantic Web. Projects like Schema.org [11] aim to provide consolidated vocabularies for providing structured data. Wikidata [28] and DBpedia [4], as large pillars of the linked open data cloud, reuse these vocabularies to allow intelligent software agents to extract meaning from the global web of data.

However, the true potential of the Semantic Web has not been unlocked yet. Different user requirements between novice and expert target groups pose a challenge. Large parts of the web are still dominated by natural language and unstructured information which machines naturally struggle with. It is rather difficult to foresee what kind of applications and new possibilities would result out of the existence of a global, structured web. Early signs, like enhanced search results [12] and Watson of the DeepQA platform [8], showcase the power behind semantic technologies.

A truly Semantic Web where both humans and machines are able to carry out sophisticated tasks with minimal effort would benefit the human race itself. Building a global knowledge web, which does not discriminate between its users, should be the primary rationale to advance the state of the world.

---

[24] `https://google.com`, Accessed: 30-10-2017

References

[1] Aroyo, L. and Welty, C. (2015). Truth Is a Lie: Crowd Truth and the Seven Myths of Human Annotation. *AI Magazine*, 36(1):15–24.

[2] Benson, E. and Karger, D. R. (2014). End-users publishing structured information on the web. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 1265–1274.

[3] Berners-lee, T., Hendler, J., and Lassila, O. (2002). The Semantic Web. 21:1–18.

[4] Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009). DBpedia - A crystallization point for the Web of Data. *Journal of Web Semantics*, 7(3):154–165.

[5] Clancey, W. J. (1989). The Knowledge Level Reinterpreted: Modelling How Systems Interact. *Machine Learning*, 4(3/4):285–291.

[6] Davis, R., Shrobe, H., and Szolovits, P. (1993). What Is a Knowledge Representation? *AI Magazine*, 14(1):17.

[7] Ferreirós, J. (2001). The Road to Modern Logic-An Interpretation. *The Bulletin of Symbolic Logic*, 7(4):441–484.

[8] Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. a., Lally, A., Murdock, J. W., Nyberg, E., Prager, J., Schlaefer, N., and Welty, C. (2010). Building Watson: An Overview of the DeepQA Project. *AI Magazine*, 31(3):59–79.

[9] Glushko, R. J. (2013). *The discipline of organizing*, volume 40.

[10] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5-6):907–928.

[11] Guha, R. V., Brickley, D., and MacBeth, S. (2015). Schema.org: Evolution of Structured Data on the Web. *Queue*, 13(9):10–37.

[12] Haas, K., Mika, P., Tarjan, P., and Blanco, R. (2011). Enhanced results for web search. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information - SIGIR '11*, page 725.

[13] Harry J. Gensler (1954). Introduction to Logic. *Philosophy and Phenomenological Research*, 14:560.

[14] Heath, T. and Bizer, C. (2011). Linked Data: Evolving the Web into a Global Data Space. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1):1–136.

[15] Hein, A. M. (2010). Identification and Bridging of Semantic Gaps in the Context of Multi-Domain Engineering. *Forum on Philosophy, Engineering & Technology*, pages 57–58.

[16] Horrocks, I. (2003). Ontologies and the Semantic Web.

[17] If, E., Don, W., Beek, W., Rietveld, L., and Schlobach, S. (2016). Why the Semantic Web Needs Centralization. *IEEE Internet Computing*, 16:78–81.

[18] Kuhn, T. (2009). AceWiki: A natural and expressive semantic wiki. *CEUR Workshop Proceedings*, 543.

[19] Kuhn, T. (2013). A Survey and Classification of Controlled Natural Languages. *Association for Computational Linguistics*, (March 2013).

[20] Law, E., von Ahn, L., and Ahn, L. V. (2005). Human Computation. *2008 Ieee 24th International Conference on Data Engineering, Vols 1-3*, 5(3):1–121.

[21] Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

[22] Moreau, L. and Groth, P. (2013). Provenance: An Introduction to PROV. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 3(4):1–129.

[23] National Research Council (2014). *Capturing change in science, technology, and innovation: Improving indicators to inform policy*, volume 51.

[24] Odell, J. J. (1994). Six different kinds of composition. *JOOP Journal of Object-Oriented Programming*, 5(8):10–15.

[25] Shoham, Y. (2015). Why knowledge representation matters. *Communications of the ACM*, 59(1):47–49.

[26] Versi, E. (1992). "Gold standard" is an appropriate term. *British Medical Journal*, 305(6846):187.

[27] Voida, A., Harmon, E., and Al-ani, B. (2011). Homebrew Databases : Complexities of Everyday Information Management in Nonprofit Organizations. *Human Factors*, pages 915–924.

[28] Vrandečić, D. and Krötzsch, M. (2014). Wikidata: A Free Collaborative Knowledgebase. *Communications of the ACM*, 57(10):78–85.

[29] Winston, M. E., Chaffin, R., and Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive Science*, 11(4):417–444.