Opportunistic Reasoning for the Semantic Web: Adapting Reasoning to the Environment

Carlos Pedrinaci[1], Tim Smithers[2], and Amaia Bernaras[3]

[1] Knowledge Media Institute, The Open University, Milton Keynes, UK. c.pedrinaci@open.ac.uk
[2] Fatronik, San Sebastián, Spain. tsmithers@fatronik.com
[3] Idom, San Sebastián, Spain. abernaras@idom.com

Abstract. Despite the efforts devoted so far, the Semantic Web vision appears to be an eluding target. We propose a paradigm shift for the Semantic Web centred around the pragmatics of developing Semantic Web applications in order to overcome the bootstrapping problem it suffers from. This paradigm is based on the vision of the Semantic Web as the result emerging from the integration and collaboration of a plethora of Semantic Web applications, rather that as a global entity. On the basis of this assumption we describe and propose Opportunistic Reasoning as a general purpose reasoning model suitable for the development of reasonably scalable Semantic Web applications.

1 Introduction

The Semantic Web aims to extend the current Web with well-defined semantics, represented in a machine interpretable form, to support better and greater automation of tasks over the Web. Prototypes have been developed aiming at exploring the capabilities of ideas and technologies the Semantic Web builds upon and latest results exhibit a certain degree of maturity. Still, achieving the results initially envisioned in [1] seems to be a somehow eluding target. The Semantic Web suffers from a *bootstrapping problem* that is limiting its success despite the amount of effort and the number of tools that have been produced so far. This had led researchers to seek other forms of reasoning that could eventually circumvent the current limitations focussing on scalable reasoning as the major barrier to surmount [4].

We propose a paradigm shift for the Semantic Web centred around the pragmatics of developing Semantic Web applications. This paradigm shift is based on a subtle distinction on how we foresee the eventual emergence of the Semantic Web with respect to what appears to be the underlying and implicit assumption within the Semantic Web research community. Typically, the Semantic Web is considered as a whole, as an entity composed of semantic annotations over the existing Web. Thus, most efforts have been devoted to creating systems that generate annotations (semi)automatically and semantic search engines that can use these annotations to increase search accuracy. As a consequence, scalability is regarded as a major issue and it is limiting to some extent the uptake of Semantic Web technologies. These applications are necessary elements to obtain a Web with annotations but we are somehow missing the intelligent applications that make use of existing annotations to better support or to enable the automation of (new) tasks over the Web. Semantic search engines and information aggregators fall into this kind of intelligent applications but these are certainly not the unique or the most appealing usage we can make of annotations.

We consider the Semantic Web rather as the result emerging from having intelligent agents that make use of semantic data on the Web, to perform new tasks and eventually contribute to the overall Semantic Web with more semantic data. Indeed, this view of the Semantic Web is perfectly compatible with current Semantic Web research. It requires and makes use of the languages, and tools developed within the Semantic Web community, but it shifts the current focus towards the creation of Semantic Web applications. The main fundamental distinction this vision implies is the fact that we don't see the Semantic Web as a Knowledge Base but instead as a source of knowledge for the intelligent applications. This carries an important set of consequences with respect to reasoning in the Semantic Web which we review in detail in Section 2 taking [4] as a starting point.

A straightforward implication is that the uptake of the Semantic Web technologies and ideas does not depend so heavily on the availability of Web scale reasoners. After all, the Web was there much before we had access to high-speed networks, ultra efficient search engines or online stores. Instead the evolution of the Semantic Web becomes an incremental process–much like the Web itself–where the industry may progressively find commercial interest. Motivated by this subtle change of viewpoint we propose in Section 3 Opportunistic Reasoning as a general purpose reasoning model suitable for the development of reasonably scalable Semantic Web applications. Finally in Section 4 we conclude and introduce some future considerations.

2 The Semantic Web: So far, So good

After approximately ten years of research on integrating Artificial Intelligence techniques with the Web, we are still seeking for the "*killer application*". So far, research on the Semantic Web has focussed on defining ontology languages, on creating tools that manipulate them, on automating the generation of metadata, on integrating search engines with ontology reasoners, etc. There are indeed many interesting and promising results which have even been applied in the industry. However, the Semantic Web vision as it was presented in [1] is yet to become a reality. The Semantic Web is certainly an appealing goal and it clearly requires applying Artificial Intelligence techniques. What is not so clear is why it is still facing this *bootstrapping problem*.

In [4] the authors identify some of the reasons for this, focussing mainly on the lack of approaches that can scale to reasoning on the Semantic Web. They revisit the underlying assumptions in state-of-the-art solutions and argue for the fusion of reasoning with search and the application of the principle of limited rationality as a solution for overcoming these limitations. Scalability is indeed a key aspect when it comes to the Web, and the ideas exposed in [4] or in [10] appear to be promising approaches to overcoming or at least minimizing scalability issues. However, scalability is not all there needs to be. In the remainder of this section we revisit these very same assumptions [4] and expand on them trying to identify the main features Semantic Web applications should provide.

The first assumptions which are questioned are the number of axioms and facts contemplated which are too low to be realistic. Indeed, it is perfectly plausible to expect a huge set of facts and axioms in an environment like the Semantic Web. This has crucial implications in terms of scalability but we would like to put more emphasis here on the consequences from a Knowledge Engineering perspective. Classical Knowledge Based Systems were based on a rather *naïve* view that complex systems could be built upon a more or less extensive set of rules. The numerous systems developed under such an assumption proved that it had some important weaknesses [3], like the complexity of the Knowledge Bases or the fact that the control of the application of the knowledge was implicit in the ordering of the rules. As a consequence the development of complex Knowledge-Based Systems was a hard and tedious task, requiring the creation of complete systems mostly from scratch with very little possibility for knowledge or software reuse (other than the inference engines themselves). Similarly, maintenance of these systems was often an even harder task because of the inherent inter-dependencies existing between the many rules.

Ontologies provide for some modularity and reusability but, ontology languages defined for the Web have (purposely) limited expressivity in order to maintain computationally attractive characteristics. And the development of Knowledge-Based Systems that can reason with the information on the Web does not, indeed cannot, uniquely rely on the manipulation of ontologies [7]. In fact for problem-solving beyond classification we need *dynamic knowledge* accompanying the *static knowledge* encapsulated in the ontologies [15]. This need is recognised within the Semantic Web research community as can be distilled for example from the research in Semantic Web Services [9, 6] and ongoing standardization efforts as pursued by the Rule Interchange Format within the W3C. Unfortunately, adding further reasoning capabilities carries computational drawbacks which are to be added to the previously mentioned scalability issues [2]. Leaving aside the well-known tradeoff between expressivity and tractability, the bottom-line conclusion one can draw is the need for embracing Knowledge Engineering principles if we want Semantic Web applications to scale. Among these principles the main one is perhaps modularity and in this sense research in Problem-Solving Methods is probably of the biggest relevance [17, 15, 7].

The third assumption which is questioned is the completeness of the inference rules, which given the characteristics of the Web, does not seem to be a reasonable requirement. Completeness is indeed a desirable feature but it is often not realistic (especially in the Semantic Web) and most importantly, it is not even necessary in many occasions. Humans exhibit a remarkable ability with respect to reasoning and we definitely do not take every single fact, and possible implication into account before taking decisions. Many knowledge intensive tasks, such as Design, are carried with a characteristic opportunism with respect to the exploration of the problem to be solved [16]. For instance when an architect is designing a building the fact that a room does not have an entrance is not particularly relevant at early stages of the design where designing the spaces is definitely more important. Later on, the designer will detect the situation and will perform the appropriate adaptations. The lack of awareness with respect to this fact does not prevent the architect from working on the design. Instead an overload of information could perfectly be cumbersome and distracting. In other words, many decisions can be, and usually are, adopted without knowing every single fact that could be derived from the existing facts. Actually, this simplification is an essential feature of Knowledge-Based Systems, for "the very act of preparing knowledge to support reasoning requires that we leave many facts unknown, unsaid, or crudely summarized" [10]. Therefore rather than complete reasoning, what it is required is an appropriate means for dynamically guiding the reasoning process in an opportunistic and incremental manner, deriving knowledge or solving problems as the need arises.

A fourth aspect which is challenged in [4] is the need for consistency and the trustworthiness and correctness of inference rules. The authors rightly point out that it does not make sense to assume or enforce this since the Web is populated with contradictory information. Dropping this assumption allows to apply probabilistic approaches to reasoning, whereby decisions can be performed over incomplete knowledge. This is a common approach to dealing with realistic domains that are characterised by their complexity [10]. This approach opens up interesting opportunities towards scalable solutions as required for the Semantic Web. However, a possible consequence is that, reasoning over incomplete knowledge leads to a plausible yet unfortunately incorrect decision. For example it could be possible, although highly unlikely that throwing a dice 1000 times we never obtain a 6. With respect to reasoning this implies that future facts observed could invalidate previously assumed ones, which requires applying non-monotonic reasoning. Making use of probabilistic reasoning is perhaps an appropriate way for reasoning over huge knowledge bases like the Semantic Web. However, it is worth noting that using this technique within some knowledge intensive task will eventually require the integration of some truth maintenance mechanism.

A final issue raised by the authors is that the Web is essentially dynamic and as a consequence the previous arguments are simply reinforced. The number of facts and axioms not only will be huge but it will also tend to increase. And enforcing or assuming the completeness and correctness of the inferences does not seem to be justified since new facts are continuously added, modified or retracted. This reinforces the fact that the Semantic Web as a whole should be considered as an open world. However taking into account the previous discussions, it also leads towards the fact that applications that want to reason about the information on the Semantic Web should rather consider their workspace as a closed world where new facts can be continuously added based on changes on the Web which could possibly invalidate previous inferences or assumptions. Again, it turns out that opportunism is a desirable behaviour, if not a required one for Semantic Web applications.

3 Opportunistic Reasoning for the Semantic Web

In any kind of reasoning, a central question is: what to do next? When it comes to automated reasoning this involves determining which pieces of knowledge to apply and when to do it. This is what is usually referred to as the "*control problem*". The Oxford English Dictionary defines opportunism as "the practice or policy of exploiting circumstances or opportunities to gain immediate advantage, rather than following a predetermined plan". Opportunistic Reasoning was defined in [3] as "the ability of a system to exploit its best data and most promising methods". Opportunistic Reasoning establishes that reasoning should respond opportunistically to changes in the current state of the problem solving process. The Blackboard Model is the Opportunistic Reasoning Model per excellence and we shall use it in this paper to illustrate its main characteristics taking into previous use cases and analysis performed. Finally, we review the general applicability of this reasoning model and contrast it with the essential features of the Web.

3.1 The Blackboard Model

During the 1970's and 1980's the "*Blackboard Model*" was proposed, developed, and applied in an important number of applications as a way to surmount the inconveniences of the classic Knowledge Based Systems approach [3]. The Blackboard Model is often presented using the analogy of a group of persons in front of a blackboard trying to put together a jigsaw puzzle. Each of them looks at his or her pieces and sees if any of them fit, in which case they update the solutions. New updates cause other pieces to fall into place and the whole puzzle can be solved in complete silence. The solution is built incrementally and opportunistically as opposed to, say, starting systematically from the top left corner and trying each piece.

In the blackboard literature, the group of people are referred to as *Knowledge Sources* (KSs) to better account for the role they play [3]. The fundamental philosophy of this problem-solving model establishes that KSs do not communicate with each other directly, instead all the interactions strictly happen through modifications on a shared workspace, i.e., the blackboard. Experts of particular aspects of the problem contribute to the overall problem-solving activity in an incremental and opportunistic way. The Blackboard Model as described by the metaphor, is a conceptual definition of a reasoning behaviour and does not prescribe any particular implementation detail. It is therefore important not to take the Blackboard Model as a computational specification, but rather as a conceptual guideline about how to perform problem-solving reasoning [3]. Even though the Blackboard Model is relatively easy to understand, it encapsulates an important number of characteristics which are not that obvious at first glance. We shall review them main ones briefly in the remainder of this section, the reader is referred to [3] and [11] for further details.

The Blackboard Model stems from a direct application of the "*divide and conquer*" principle. The rationale underlying such a principle is to decompose complex problems into more manageable and presumably simpler ones. Solutions to a problem can be obtained by bringing together solutions to sub-problems. This is a well-known, widely applied, and often successful approach to dealing with complex or large problems which makes it particularly useful for an environment like the Web. In addition to simplifying problem-solving, the divide and conquer approach leads to partitioning the expertise required for the reasoning activity. The divide and conquer approach underpining the Blackboard Model thus, relies in partitioning the problem-solving expertise that propagates over the whole application development cycle. The modularity brought by the independence of expertise supports the incremental development, extension and replacement of Knowledge Sources in a relatively simple way and supports the integration different reasoning mechanisms to overcome the limitations of a unique technique. This allows for a rather straightforward application of Knowledge Engineering principles as presented in [15, 17].

The blackboard metaphor, establishes that all the interactions among the knowledge sources must take place indirectly via the shared blackboard. As a consequence, communication among experts does not rely on any agreed interaction protocol. Instead, it requires a common representation language for the information to be understood by various Knowledge Sources; ontologies being a rather obvious candidate to fulfil this role. A natural requirement, if we want communication to take place only via the shared blackboard, with no direct communication among the Knowledge Sources, is that KSs have to be "self-activating". Hence, this reasoning model relies on event-based activations of the Knowledge Sources, who—much like the people solving the jigsaw puzzle—are watching the blackboard to react to changes made by other KSs. The event-based activation, enables applying this reasoning model to particularly dynamic domains like the Web.

As promoted by the blackboard metaphor, the Blackboard Model is an incremental reasoning model where reasoning progresses step-by-step. Consequently, solutions are generated incrementally guided by the different and collaborating sources of expertise. This allows for a guided exploration of the possible solutions, potentially reducing the amount of calculations to be performed in order to find a solution (if any). Moreover, thanks to the application of several sources of knowledge focussing on different aspects of the problem, this reasoning model represents a good heuristic that avoids some typical problems suffered by traditional algorithms whose behaviour is pre-defined, such as Hill-Climbing, to cite a well-known example. The behaviour exhibited by the Blackboard Model is characterised by its ever changing focus of attention, or, in Feigenbaum's words, the "island-driven exploration of the solution space" [3], which confers on this reasoning model the outstanding ability to deal with uncertainty and smoothly adapt to changing conditions.

The distinctive property of the Blackboard Model, which stems from the previous characteristics, is what is usually referred to as Opportunistic reasoning. Opportunistic reasoning is a flexible knowledge application strategy as opposed to fixed algorithms whose behaviour is pre-established and cannot therefore be adapted and reoriented facing particular situations, at least, not easily. Hence, this expression accounts for the prominent ability of the Blackboard Model to seamlessly integrate a collection of collaborating Knowledge Sources into the overall reasoning activity, applying their expertise at the most opportune time. Perhaps, the most prominent example of the opportunistic reasoning capabilities of the Blackboard Model, is illustrated by the BB1 blackboard control architecture [3]. BB1 was a task-independent implementation of the blackboard control architecture whose goal was to provide a solution to the control problem (i.e., deciding which of the potential actions the system should perform next). Through this work, the authors explored the ability of the Blackboard Model for meta-reasoning and conscious reasoning when applying the Blackboard Problem-Solving Model. Their resulting framework, and conclusions regarding the general application of the Blackboard Model for this particular problem-solving task (i.e., solving the control problem), accounted for the outstanding capabilities to seamlessly, flexibly and dynamically coordinate the various collaborating experts towards achieving a solution.

3.2 On the Pragmatics of the Blackboard Model

So far we have focussed on the main characteristics of the Blackboard Model as the Opportunistic Reasoning Model per excellence. In this section we focus on the applicability of this reasoning model in general but also in particular to the Web. Some authors, have already realised the importance of better establishing the conditions under which the Blackboard Model is appropriate. However, these are not much more than useful guidelines. For instance, [8] argues that the Blackboard approach is generally suitable for solving ill-structured problems (i.e., a problem with poorly defined goals and where no predefined algorithm exists that leads to a solution) and complex problems (i.e., one made up of a large number of parts that interact in a nonsimple way). [8] also gives some properties that often characterise the problems that were successfully solved through the Blackboard Model.

Taking into account the current state-of-the-art in blackboard systems, we have extended the set of properties to include those we have noticed are recurrent [11]. These characteristics have been identified from a large set of applications for performing diverse knowledge-intensive tasks within diverse domains. These include voice recognition, military situations monitoring and assessment, signal processing, drug design support, events design support, military planning, process scheduling systems, etc [3, 14, 13, 11]. In general the occurrence of some combination of the following problem characteristics can serve as a good indication of the appropriateness of the blackboard approach[1]:

1. A large solution space;
2. Noisy or unreliable problem data;
3. A continuous data flow;
4. The need to integrate diverse and heterogeneous data;
5. The need to integrate different sources of knowledge;
6. The need to apply several reasoning methods;
7. The need to develop various lines of reasoning;
8. The need for incremental reasoning;
9. The need for an opportunistic control of the reasoning process;
10. The need for an event-based activation of the reasoning;
11. High complexity of the task;
12. The need for mixed initiative control. That is, the need for a collaborative framework where computer and users can interchangeably take the initiative;
13. Meta-reasoning or conscious reasoning;
14. The need for psychologically reasonable implementations. This stands for the ability to map human problem-solving into the automated implementation as well as for the capacity to provide explanations of the reasoning steps performed.

As can be distilled from the previous characteristics, the Blackboard Model is a general and versatile reasoning model, particularly well suited for supporting reasoning processes over the Web. It provides an outstanding support for reasoning in highly dynamic environments. It supports adapting the reasoning process to the very typical and diverse events of the Web, such as remote execution exceptions, a continuous data flow, connectivity problems, etc. In fact, the event-based activation of the expertise paves the way for the seamless, effective and Knowledge-Based choreography of remote (knowledge-based) services distributed over the Web, as shown for instance in [12, 13].

This reasoning has been shown to be particularly well suited for dealing with noisy, unreliable, heterogeneous and massive data. This is, as discussed in [4] and in Section 2, of crucial importance for reasoning over the Web, where its extension, the lack of central control and the principle of tolerance have led to a highly noisy, often contradictory, unreliable and increasingly expanding source of information. In fact, the island-driven exploration of the solution space that characterises Opportunistic Reasoning together with its prominent capacity for integrating diverse reasoning techniques, provides an excellent support for reasoning over the Web.

[1] The existence of some of these characteristics does not make of the Blackboard Model an appropriate reasoning model. However, the more of these characteristics that appear in a problem, the more likely the blackboard model would be appropriate.

4 Conclusions and Outlook

The Web opens up a plethora of possibilities which have already been demonstrated in a variety of different scenarios. Reasoning over the Web is part of the great potentials that can be brought to reality with some efforts and discipline. To do so, we must however, assume the very nature of the Web and adapt our techniques and technologies to it. Any kind of application conceived for the Web must be prepared for this particular environment. This involves being ready to deal with an extremely large and distributed environment, populated by uncertain, dynamic and to some extent autonomous information, served by machines which are prone to errors.

We have proposed viewing the Semantic Web as the emerging result from having intelligent agents consuming and generating semantic data on the Web, as a means to surmount the so-called bootstrapping problem. The incremental nature of this approach provides the simplifications required for creating useful reasoning systems out of the existing semantic data on the Web while it allows the creation of more complex system on the basis of existing ones which are considered as knowledge services providers. Finally, we have presented Opportunistic Reasoning as a general purpose reasoning that can appropriately be applied to the developing applications that reason over the Web, therefore contributing to the overall evolution of the Semantic Web.

The approach presented in this paper introduces pragmatic ideas that could alleviate to an important extent the bootstrapping problem within the Semantic Web. This research does not however intend to be a solution to the development of a Web scale semantic search engine. Still, there are many ideas which could be reused and in fact there are currently other researchers applying similar approaches in order to develop Web scale reasoning systems [5]. In this sense we expect that the integration of probabilistic reasoning as proposed in [4] and [10] together with the outstanding qualities of the Opportunistic Reasoning model could yield promising results.

References

1. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, (5):34–43, May 2001.
2. R. Brachman and H. Levesque. *Knowledge Representation and Reasoning*. Morgan Kaufmann, June 2004.
3. R. S. Engelmore and A. J. Morgan. *Blackboard Systems*. The Insight Series in Aritificial Intelligence. Addison-Wesley, 1988. ISBN: 0-201-17431-6.
4. Dieter Fensel and Frank Van Harmelen. Unifying reasoning and search to web scale. *IEEE Internet Computing*, 11(2):94–96, March/April 2007.
5. Dieter Fensel, Reto Krummenacher, Omair Shafiq, Eva Kuehn, Johannes Riemer, Ying Ding, and Bernd Draxler. Tsc - triple space computing. *Journal of electronics & information technology (e&i Elektrotechnik & Informationstechnik), special issue on ICT research in Austria*, 2 2007.

6. Dieter Fensel, Holger Lausen, Axel Polleres, Jos de Bruijn, Michael Stollberg, Dumitru Roman, and John Domingue. *Enabling Semantic Web Services: The Web Service Modeling Ontology*. Springer, 2007.
7. M. A. Musen. Ontologies: Necessary–Indeed Essential–but Not Sufficient. *IEEE Intelligent Systems*, 19(1):77–79, January/February 2004.
8. H. P. Nii. Blackboard application systems and a knowledge engineering perspective. *AI Magazine*, 7(3):82–106, Summer 1986.
9. OWL Services Coalition. Owl-s: Semantic markup for web services, 2003. http://www.daml.org/services/owl-s/1.0/.
10. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.
11. C. Pedrinaci. *Knowledge-Based Reasoning over the Web*. PhD thesis, Universidad del País Vasco/Euskal Herriko Unibertsitatea, November 2005. Available at: http://kmi.open.ac.uk/people/carlos/publications/Thesis-Final.pdf.
12. C. Pedrinaci, Z. Baida, H. Akkermans, A. Bernaras, J. Gordijn, and T. Smithers. Music rights clearance business analysis and delivery. In K. Bauknecht, B. Pröll, and H. Werthner, editors, *6th International Conference on Electronic Commerce and Web Technologies*, volume 3590 of *Lecture Notes in Computer Science*, pages 198–207. Springer-Verlag, August 2005.
13. C. Pedrinaci, T. Smithers, and A. Bernaras. Technical issues in the development of knowledge-based services for the semantic web. In *SWAP: Semantic Web Applications and Perspectives. 2nd Italian Semantic Web Workshop.*, Trento, Faculty of Economics, December 2005.
14. N. Sadeh, D. Hildum, T. Laliberty, J. McA'Nulty, D. Kjenstad, and A. Tseng. A blackboard architecture for integrating process planning and production scheduling. *Concurrent Engineering: Research & Applications*, 6(2), June 1998.
15. Guus Schreiber, Hans Akkermans, Anjo Anjewierden, Robert de Hoog, Nigel Shadbolt, Walter Van de Velde, and Bob Wielinga. *Knowledge Engineering and Management: The CommonKADS Methodology*. MIT Press, 1999.
16. T. Smithers. Towards a knowledge level theory of design process. In J. S. Gero and F. Sudweeks, editors, *Artificial Intelligence in Design '98*, pages 3–21. Kluwer Academic Publishers, 1998.
17. Rudi Studer, Richard Benjamins, and Dieter Fensel. Knowledge engineering: Principles and methods. *Data Knowledge Engineering*, 25(1-2):161–197, 1998.