

Semantic web: Need, opportunities, and challenges in the modern day world

¹ Herald Noronha, ² Dr. Vijay Prakash Agrawal

¹ Research Scholar, Singhania University, Pacheri Bari, Jhunjhunu, Rajasthan, India

² Associate Professor, Department of Math, BSA, PG College, Mathura, Uttar Pradesh, India

Abstract

Information on the web is increasing in every single second and redundancy in information is growing rapidly. The Semantic Web is an extension of the current Web. The original idea of the Semantic Web was to bring machine-readable descriptions to the data and documents already on the Web, in order to improve search and data usage. The Web was, and in most cases it still is, a vast set of static and dynamically generated Web pages linked together.

Today, the Semantic Web is not only about increasing the expressiveness of Web information to enable the automatic or semiautomatic processing of Web resources and Web pages. Academia and industry have realized that the Semantic Web can facilitate the integration and interoperability of intra- and inter-business processes and systems, as well as enable the creation of global infrastructures for sharing documents and data, make searching and reusing information easier.

In this present paper a theoretical attempt has been made to discuss the real world need, applications, opportunities, and challenges of Semantic Web.

Keywords: Semantic Web, W3C, RDF, MAS, HTML

Introduction

The Semantic Web is an extension of web through standards by the World Wide Web Consortium (W3C). In other words, the Semantic Web is a Web 3.0 web technology- a way of linking data between systems or entities that allows for rich, self-describing interrelations of data available across the globe on the web. On the Semantic web information is described using a new W3C standard called the Resource Description Framework (RDF).

Agent and agent oriented technology is considered to be the one of the rapid growing research area. There is a widespread and ever growing interest in the potential of agent based technologies and develops the complex and advanced next generation software system using agents. Agent encompasses various diverse fields of computer engineering viz. Artificial intelligence, software engineering, internet applications, manufacturing process and distributed computing etc.

World Wide Web (WWW) - The Concept

In 1990, Tim Berners-Lee developed the first version of his World Wide Web program at CERN. The concept behind Berners-Lee's invention was to use hypertext as a means of organizing a distributed document system. Hypertext refers to a collection of documents with cross-references (also known as links) between them that enable readers to peruse the text in a non-sequential manner. In order to make the Web work on the Internet, Berners-Lee had to develop a mechanism for addressing documents on different machines, a protocol that allowed computers to request documents, and a simple language to describe the documents.

Despite its popularity, HTML suffered from two problems.

- First, whenever someone felt that HTML was insufficient for their needs, they would simply add additional tags to their documents, resulting in a number of non-standard variants.

- Second, because HTML was mostly designed for presentation to humans, it was difficult for machines to extract content and perform automated processing on the documents.

To solve these problems, the World Wide Web Consortium (W3C) developed the Extensible Markup Language (XML). XML's syntax will seem familiar to users of HTML. This is not surprising, since HTML is an application of SGML, XML's parent language. Like HTML (and SGML), XML allows angle-bracketed tags to be embedded in a text data stream, and these tags provide additional information about the text. However, unlike HTML, XML does not provide any meaning for these tags.

Semantic Web

As discussed earlier, the term "Semantic Web" was coined by Tim Berners-Lee, the inventor of World Wide Web and the director of the World Wide Web Consortium (W3C), which oversees the development of proposed Semantic Web standards. The Semantic Web is changing the way how scientific data are collected, deposited, and analyzed. It is regarded as an integrator across different content, information applications and systems. It has applications in publishing, blogging, & many other areas.

Semantic Web is about providing meaning to the data from different kinds of web resources to allow the machine to interpret and understand these enriched data to precisely answer and satisfy the Web users' requests. As a part of second generation web it gives users of web the ability to share their data beyond all the hidden barriers and the limitations of programs and websites using the meaning of the web.

Semantic web technology is built in a layered manner, i.e. it is processed in steps, each step built on top of another. The pragmatic justification of it is that it is easier to achieve

consensus on small steps, whereas it is much harder to get everyone on board if too much is attempted. For Semantic Web, Hyper-Text Web Technologies are used. These are the bottom layer technologies that are well known in the hypertext web domain. These technologies are used to implement semantic web application.

The layers of Semantic Web are depicted in figure 1 for better understanding and illustration.

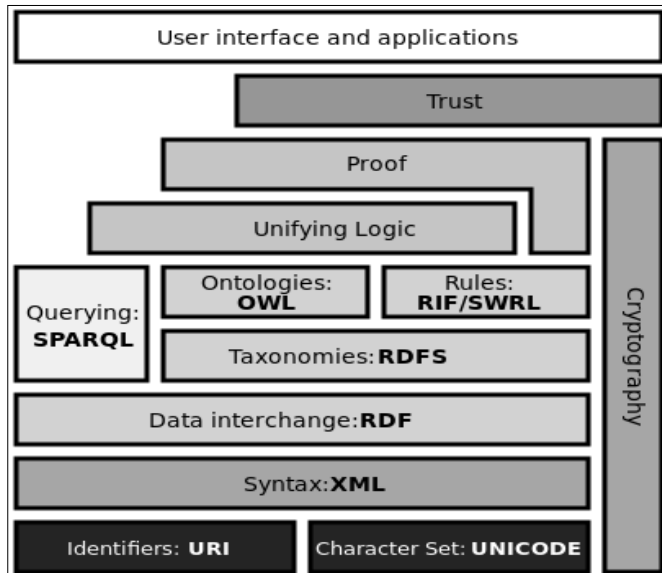


Fig 1: Layers of Semantic Web

These layers are described as follow:

1. **Uniform Resource Identifiers (URI):** URI is used to identify the semantic web resources. This unique identification is required so as to provide manipulation with the resources in the top of the layers. In other words, this layer makes sure of providing the uniform identifiers of lots of resources. It likes the strings starting with “http:” or “ftp:” that often find on the World Wide Web.
2. **Unicode:** This layer provides the uniform standards for all kinds of languages in the world to coding characters. It helps to represent and manipulate text in various languages, thus enabling a bringing of the gap between the human languages and semantic applications.
3. **Extensible Markup Language (XML):** XML is the markup language that is creates the semantic web documents in the form of structured data. XML replaces HTML for its individual advantages. In addition, it gives a definition on the methods describing data. These are the middle layer technologies, most of which have been standardized by W3Cm and can be used to create semantic web applications. All are standardized by the W3C except for RIF/SWRL.
4. **Namespace:** This layer provides many ways to differentiate names, so the resources which have same names and different means are still used.
5. **Resource Description Framework (RDF):** RDF is the framework that is used to express data in a meaningful way. It expresses data in the form of triples, which is easier to express information in the form of a graph. Here reserve specifications refer to model and syntax of RDF.
6. **Resource Description Framework Schema (RDFS):** RDF Schema (RDFS) provides the schema, i.e. the

vocabulary, for the RDF to maintain a proper structure of the document. It enables to maintain a proper hierarchy of classes and its properties. So, it is a kind of language for describing RDF vocabularies and has some basic elements, e.g. Resource, Class, Property, subclassOf, subPropertyOf, range, domain etc.

7. **Ontology: OWL- Web Ontology Language (OWL)** is used to add more meaning, constraints and restriction to the RDF representation. It expresses the semantic of the RDF statement. It is a formal, explicit specification of a shared conceptualization. It needs to make sure all knowledge can be known by people and computer.
8. **SPARQL: SPARQL Protocol and RDF Query Language (SPARQL)** is an RDF query language that is used for querying in the database that is represented by the RDF. Querying is done so as to retrieve information by the semantic applications. These are *Unrealized Semantic Web Technologies*. These are the top layer technologies that are yet to be standardized or are ideas that need to be implemented to completely create semantic web applications.
9. **Rule Interchange Format/ Semantic Web Rule Language (RIF/SWRL):** RIF/SWRL is used to add rules to the RDF data. This enables to represent information that cannot be directly expressed by the OWL. Rules are basically many principles or regulations between the upper and lower in the level structure.
10. **Logic Framework:** It provides logic inference ability on the knowledge of Ontology describing.
11. **Proof:** On being logic inference ability, give a proof on whether a statement is right or wrong.
12. **Trust:** Trust for statement support, means the premises come from trusted sources and relying on formal language to retrieve new information. In other words, detect whether the web information is trusted or not.
13. **Cryptography:** This is to ensure that the statements coming from semantic web applications are from proper sources and this can be implemented using digital signatures of RDF documents.
14. **User Interface:** This is the top most layers that will enable the humans to use the semantic web applications.

Characteristics of the Modern Day World

We live in a world where information and knowledge are considered to be key enablers of business and economic performance and critical pillars of sustainable development. At global level the following are some of the characteristics of the new world context:

- **Globalization:** Creation and consumption of knowledge and information are made in the global context. From this perspective the elimination of local boundaries and the exploitation of synergies and capacities beyond local boundaries require advanced adoption mechanisms that permit realization of opportunities, deep understanding of threats and strategic fit to human and social networks towards new levels of performance.
- **Networking:** In our era, business and economic activities as well as competition require new models of business networking. Within this context, advanced documentation of skills, competencies, business models and context based collaboration define new demands for advanced business and social networking at global level.

- **Shared Models:** A global consensus towards peace, development, prosperity and a better world needs to be based on shared conceptual models that define the average common understanding of human societies for the “issues” that matters at global scale. And while this can be perceived as a “too optimistic” scenario or it can be characterized as wishful thinking case, in the Global information Landscape shared models are required for interoperability, exploitation of synergies and definition of new milestones for collective intelligence.
- **Collective Intelligence:** The increased capacities of networking as a result of globalization and widespread adoption of share models has resulted to the development of a global trend to applying collective intelligence filters or collaborative filtering in the context of the global information world. Such development challenges many traditional models of business performance, marketing and profitability.
- **Open Paradigm:** This is a key new characteristic of our world. Open paradigm relates with several complementary movements, like the ones of open source software paradigm, open content, open access, open knowledge, open research, open culture. The underlying idea has an amazing capacity to support new business models and several applications in the short and long term horizon.

Applications of Semantic Web

1. E- Business

According to Wikipedia, e-business may be defined as “*any business process that relies on an automated information system*”. E-business software solution allows the integration of intra- and inter-firm business processes and can be conducted using the Web, the Internet, extranet, etc. The application of Semantic Web in the field of e-business is wide and significant in terms of exchanging information between different business groups for mutual or collective purposes. It is believed to provide good semantic solutions when information exchange is concerned.

Advantages of E-business

Semantic Web has found a prominent place in e-business in terms of

- searching of relevant data,
- exchange of information between different agents,
- filtering of relevant information useful for finding good business sites or analyzing new market trends,
- online advertisements,
- composition and integration of complex systems,
- multimedia collection,
- exchange of machine dialogue across the domains,
- virtual community and vocabulary flexibility, and
- Standardization.

Limitations of E-business

One of the major limitations or drawbacks in e-business is the problem of interoperability between systems of two or more business partners (business-to-business). Interoperability is to be ensured for the business exchange to be effective and efficient between business companies. Extensible Markup Language (XML) has been used to provide the Web interoperability for the past few years. XML is capable of providing only syntactic interpretation, and not semantic

interpretation. It does not understand the content and meaning of the messages being exchanged among different systems.

Semantic Web solves the problem of interoperability using Web Ontology Language (OWL). OWL is a popular language used for representing ontologies on the Web. It is a World Wide Consortium (W3C) Standard and provides a strong ontology that is used as a standard. All the services on the Internet share the same standard for the interpretation of terms being exchanged between them. OWL facilitates greater interoperability of Web content.

Moreover, the software used earlier for business purposes were hard-coded and always required direct human intervention by changing the code and then running the script. But with the use of Semantic Web tools and software, the pages can be dynamically changed and highly tailored and cost-effective results can be produced.

2. Social Networking

Social networking has become an important part of the modern society and puts a strong impact on social, political, educational, professional, personal, and business life.

Advantages of Social Networking

- It connects people across the world through social networking sites like Facebook, Orkut, MySpace, LinkedIn, etc.
- It allows information sharing on twitter, messaging through Yahoo Messenger, Google talk, content and ideas sharing through blogs, discussion forums, uploading and downloading of media, tagging, through wikis and podcasts, etc.
- The social networking have attracted millions of users across the globe and has become the most popular, convenient and cheaper mode of communication.
- A lot of social networking sites are coming into the business because of wealth being generated by these across the world.

Limitations of Social Networking

Social networking sites or services offer the basic features, but there are certain limitations in terms of connecting people and content in a meaningful way.

- One of the limitations is the lack of interoperability among different social networking sites. Suppose a person having a profile account in one of the site wants to reopen the new profile account in some other site, and wants the same information in his previous account to be migrated into his new account, will not be able to do so because the provision is not available in social networking services. In such a scenario, the user has to re-enter the entire information, and update its information at two different sites, making the process more cumbersome.
- Second limitation is the lack of privacy. In case of certain centralized sites, the user does not have complete control over the information they provide. Thirdly, social networking sites do not provide relationships or connections between common profiles of people sharing common features liking visiting the same sites, or same likes, and hobbies.

Semantic Web solves the problem of interoperability by providing globally accepted semantics to share information about people, their profiles, contents and connections through

which they are interconnected. Security and privacy is being preserved as data is machine-readable.

3. Knowledge Management

Knowledge management system (KMS) as a whole describes the creation of knowledge repositories, a method for knowledge access and sharing as well as communication through collaboration and also enhancing the knowledge environment and managing knowledge as an asset for an organization.

Development of a good KMS is basically a collaborative effort. The traditional collaboration of KMS has undergone a revolution with the advent of WWW. The WWW or Internet provides a knowledge repository with variety of information from various sources and also from geographically distant corners of the world. This makes the repository information rich like never before but this also brings the real challenges.

These challenges, like overload of information, keyword searching being not appropriate, integration of information from heterogeneous sources and geographically-distributed Intranet problems, have been triggered by the Web. In fact, this enormous amount of data has made it increasingly difficult to search, access, present, and maintain the information required by a wide variety of users. This is because information content is mainly presented in a natural language. Thus, a wide gap has occurred between the information available for tools aimed for knowledge extraction and the information maintained in human-readable form.

The most critical issue in intelligent knowledge management is how to represent and extract the semantic meaning from information contents. Researchers have tried to address this issue through various research areas including artificial intelligence, information retrieval, natural language processing, multimedia, knowledge management, etc. All these methods ask for a smarter Web to assist in knowledge acquisition, knowledge representation, and knowledge sharing and distribution of human knowledge through the Web. This requirement has been answered by Semantic Web.

Tim Berners Lee has referred to Semantic Web as an extended Web of machine-readable information and automated services that extends far beyond current capabilities. Semantic Web is useful for KMS because this technology helps people to find and correlate the information they require irrespective of their sources and types. Semantic Web is relevant to knowledge management because it has the capacity to enhance the speed with which information can be synthesized manifolds. This is achieved by automating its aggregation and analysis. Most of the time information on the Web is presented in HTML format but the problem with it is that this format does not provide structure or metadata useful for effective management. Without structure, elements of content cannot be related to each other, and without metadata, the nature of the elements themselves cannot be known. Semantic Web is designed to provide these missing components. It can provide structure, (through the use of XML tags); metadata descriptors, (through RDF) and relationships (through Web Ontology Language). 'Ontology' is the key enabling power in realizing the full potential of Semantic Web technology. Ontology is not knowledge or information. It is in fact, meta-information which is information about information.

In the context of the Semantic Web, using a special ontology language, the relationships between the various terms within

the information can be encoded. Ontologies provide background information, which strengthens the description of the data and which helps in making the context of the information more explicit. Since ontologies are shared specifications, the same ontologies can be used for the annotation of multiple data sources, which include webpages, collections of XML documents, relational databases, etc.

The use of such shared terminologies enables interoperability between these data sources up to a certain extent. But, this does not solve the integration problem completely, because it is not possible for all individuals and organizations on the Semantic Web to use one common terminology or ontology. It is very much possible that different ontologies will appear and, to enable interoperation, mediation is required between these ontologies.

Ontology mediation is necessary in semantic knowledge management for enabling sharing of data between heterogeneous knowledge bases and also to allow applications to reuse data from different knowledge bases. Another utility of ontology mediation appears in Semantic Web Services. In general, it is not necessary that the requester and the provider of a service use the same terminology in their communication, and thus, mediation is required to facilitate communication between the knowledge seekers and the knowledge providers.

Challenges for Semantic Web Applications

Following are a selective list of challenges for Semantic Web applications in close relevance to the previous discussion:

- **Definition of new modes for human, knowledge and business networking beyond local boundaries:** Traditional business and knowledge networking emphasized on a narrow perspective for the ultimate objective of networking. Semantic Web through Ontologies and Social networks anchors networking to well-defined conceptual models that match information sources and human services. By providing an infrastructure of shared semantics and Ontologies where reasoning and trust are "process and service oriented", we have a great opportunity at business level.
- **Globalizing information and definition of new Contexts for value exploitation:** The provision of local information assets at global level and the design of new contexts for exploitation are for the Semantic Web two of the key value proposition. The design of multiple reference levels to the same set of information and knowledge delivers a new level for dynamic, personalized systems.
- **Delivering and integrating quality to information:** One of the main obstacles in the current web relates with a very limited performance on the quality assessment of content. It seems that we suffer from an enormous explosion of content diffusion and a very poor performance on capacities to explore qualitative information. And while information quality is a very subjective concept, the same moment businesses and people as customers, citizens, patients, learners, professionals, etc require systems and infrastructures that deliver assessment models of information quality.
- **Integration of isolated information assets:** In any context, personal, organizational or global the integration of isolated information is a key challenge. The "value" related with integration is always related with the inquiry. In simple words integration has always a very concrete

“gap” component. Individuals, Organizations, society require integration for addressing specific performance gaps that relate with limited capacity to build more meaningful services.

- **Support of business value and co-located/distributed business models:** It is obvious that Semantic Web evolution requires the adoption from industry. This critical milestone implies that Semantic Web technologies can be integrated to the business models of modern organizations and businesses. As always such a requirement challenges the strategic fit of technologies to business perspectives. From a business strategy point of view there is a key demand to “translate” the key aspects of Semantic Web technologies to business terminology. Semantics, ontologies, resource description frameworks, etc, means nothing to business people that have an absolute different way to interpret business requirements.
- **Promotion of a critical shift in humans understanding and interacting with digital world:** Semantic Web needs to respond to the great demand of humans to explore new modes of interactivity with the digital world. And it is obvious that people prefer to behave with similar conscious and intelligent mechanisms. The sooner the Semantic Web will prove its capacity to provide these intelligent mechanisms the greater its adoption and support at global level.

Conclusion

To conclude, it can be said that Semantic Web has become the most important part of the web without which modern web is unimaginable. But, it has still to overcome from its shortcomings and face the challenges. Once it has overcome the prevailing challenges, it will definitely set a milestone in the history of WWW which will open the new pave for ontologies.

References

1. Aberer K. *et al.* Gridvine: Building internet-scale semantic overlay networks. Number 3298 in LNCS. 2004, 107-121.
2. Arumugam M, *et al.* Towards peer-to-peer semantic web: A Distributed Environment for sharing Semantic knowledge on the web.
3. Bajwa IS, Chaudhary MA. A Language Engineering System for Graphical User Interface Design (LESGUID): A rule based approach: Information and Communication Technologies, 2006, 2:3582-3586.
4. Bellifemine F, *et al.* JADE a white paper: EXP in search of innovation. 2003, 3.
5. Benjamins R, Decker S, Fensel D, Gomez-Perez A. Building Ontologies for the Internet: A Mid Term Report, International Journal of Human Computer Studies (IJHCS), September. 687-712.
6. Bigus JP. *et al.* ABLE: A toolkit for building multiagent autonomic systems. IBM systems journal. 2002, 41.
7. Castano S. *et al.* Ontology-addressable contents in p2p networks. In Proc. Of WWW03, 14th Sem PGRID Workshop. 2003.
8. Decker S, Erdmann M. Ontology Based Access to Distributed and Semi-Structured Information”, Proceedings of DS-8. Kluwer Academic Publisher, Boston. 1999, 351-369.
9. Doane SM, Lemke AC. Using cognitive simulation to develop user interface design principles, System Sciences, Proceedings of the Twenty-Third Annual Hawaii International Conference on. 1990; 2:547-554.
10. Doorenbos RB, Etzioni O, Weld DS. A scalable comparison shopping agent for the world-wide web. In Johnson, W. L. and Hayes-Roth, B., editors, Proceedings of the First International Conference on Autonomous Agents (Agents’97), Marina del Rey, CA, USA. ACM Press. 1997, 39-48.
11. Ehrig M, Staab S. QOM quick ontology mapping. Number 3298 in LNCS. 2004, 683-697.
12. Eichmann D. Supporting multiple domains in a single reuse repository. InProc. Fourth International Conference of Software Engineering and Knowledge Engineering (SEKE’92). 1992, 164-169.
13. Fensel D, Ding Y, Omelayenko B, Schulten E, Botquin G, Brown M. *et al.* Product Data Integration in B2B E-commerce, IEEE Intelligent Systems (Special Issue on Intelligent E-Business). 2001; 16(4):54-59.
14. Fernández M, Gómez-Pérez A, Juristo N. Methontology: From Ontological Art Towards Ontological Engineering. Symposium on Ontological Engineering of AAI. Stanford (California). 1997.
15. Garland A, Alterman R. Preparation of multi-agent knowledge for reuse. In Aha, D. W. and Ram, A., editors, Working Notes for the AAI Symposium on Adaptation of Knowledge for Reuse, Cambridge, MA. AAI. 1995.
16. Guarino N. Formal Ontology and Information Systems. In Proceedings of Formal Ontology in Information Systems. 1998. Netherlands: IOS Press.
17. Halevy A. *et al.* Schema mediation in peer data management systems. In: Proceedings of the International Conference on Data Engineering (ICDE03), Bangalore, India. 2003.
18. Joachims T, Freitag D, Mitchell TM. Web watcher: A tour guide for the world wide web. In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97). 1997, 770-777.
19. Jorge Cardoso, Martin Hepp, Miltiadis Lytras. Real World Applications of Semantic Web Technology and Ontologies. 2008, 1.
20. Kuhnel R. Reaching agreements through argumentation: a logical model and implementation. Artificial Intelligence. 1999; 104(1):1-69.
21. Lesser V, Hornling B, Klassner F, Raja A, Wagner T, Zhang S. Big: An agent for resource-bounded information gathering and decision making. Artificial Intelligence. 2000; 118(1-2):197-244.
22. Marshall CC, Shipman MFM. Which Semantic Web? In, Proceedings of the fourteenth ACM Conference on Hypertext and Hypermedia. Nottingham, UK, New York: ACM Press. 2003.
23. Nejdil W. *et al.* Edutella: A p2p networking infrastructure based on rdf. In: Proceedings of the WWW2002, Honolulu, Hawaii, USA. 2002, 604-615.
24. Quiroz J, Shankar A, Dascalu SM, Louis SJ. Software Environment for Research on Evolving User Interface Designs, Software Engineering Advances, 2007. ICSEA 2007. International Conference on. 2007; 25(31):84-84.
25. Schelfhout K, Coninx T, Helleboogh A, Holvoet T, Steegmans E, Weyns D. Agent Implementation Patterns,

- Proceedings of the OOPSLA 2002 Workshop on Agent-Oriented Methodologies (Debenham, J. and Henderson-Sellers, B. and Jennings, N. and Odell, J., eds.). 2002, 119-130.
26. Sichman JS, Demazeau Y. Exploiting social reasoning to deal with agency level inconsistency. In Proceedings of the First International Conference on MultiAgent Systems (ICMAS-95). AAAI Press/The MIT Press. 1995.
 27. Tamma V. *et al.* SERSE: searching for semantic web content. In: Proceedings of ECAI. 2004.
 28. Terry R, Payne, Rahul Singh, Katia Sycara. Calendar Agents on the Semantic Web, IEEE Intelligent Systems. 2002; 17(3):84-86.
 29. Tim Finin, Rich Fritzson, Don McKay. A Knowledge Query and Manipulation Language for Intelligent Agent Interoperability, Fourth National Symposium on Concurrent Engineering, CE & CALS Conference, Washington DC. 1992.
 30. Turner J, Jennings N. Improving the scalability of multi-agent systems. In Wagner, T. and Rana, O., editors, Infrastructure for Agents, Multi-Agent Systems, and Scalable Multi-Agent Systems, volume 1887 of Lecture Notes in Artificial Intelligence. 2000, 246-262. Springer Verlag.
 31. Wooldridge M, Bussmann S, Klosterberg M. Production sequencing as negotiation. In: Proceedings of the First International Conference on the Practical Application of Intelligent Agents and Multi-agent Systems, London. 1996
 32. Wooldridge M, Jennings N. Intelligent agents: Theory and practice. Knowledge engineering review. 1995; 10:115-152
 33. Zilberstein S, Russell S. Approximate reasoning using anytime algorithms. In Natarajan, S., editor, Imprecise and Approximate Computation. Kluwer Academic Publishers. 1995.