A Conceptual Roadmap for Scalable Semantic Computing

Graham Hench[1], Elena Simperl[2], Alexander Wahler[1], Dieter Fensel[1]
[1]*STI International &* [2]*STI Innsbruck, University of Innsbruck, Austria*
[1]*{firstname.lastname}@sti2.org,* [2]*{firstname.lastname}@sti2.at*

Abstract

Semantics has recently become a promising candidate panacea to computing scalability problems. However, this rather ambiguous candidacy title does not specify how this solution should be fulfilled, i.e. what steps must still be taken and which objectives must be implemented to provide truly scalable semantic computing platforms. The purpose of this conceptual roadmap is to address the scalability challenges within specific areas of semantic computing, to provide achievable solutions based on ongoing research and development, and to present a coordinated overview of what must take place in the next five years in order to achieve worldwide scalability.

1. Introduction

In this paper we present a roadmap outlining the steps that need to be taken towards achieving scalable semantic computing. Given the current state of affairs - and presuming the fields of semantic computing continue to advance in the same direction as the predominant ICT R&D of the last decade - semantic computing still faces major challenges in bringing about scalable computing. Based upon these major challenges, we propose approachable solutions that can be realized over the next five years. Collectively these solutions properly integrate to provide a worldwide semantic network of services (semantic service web), the overall goal of this roadmap. If these scalability objectives are fulfilled, a Web of Services (engineered, composed, and provisioned solutions) will evolve as the new prime showcase of scalable semantic computing.

The challenges and solutions addressed in this conceptual roadmap do not presume to represent the entirety of scalability limiting factors; nor do they presume that the defined solutions seamlessly (and sufficiently) integrate with one another. These challenges have been chosen for two main reasons: 1) they are indeed prevalently crucial to the progression of semantic computing; and 2) their solutions require coordinating activities and collaborative objectives to ensure their integration with one another and current standards.

The first section discusses the start-of-the-art and provincial challenges of achieving scalability in semantic computing. Section 3 provides solutions to these challenges and lays out the necessary steps to be achieved within each of these areas over the next five years. The subsections of Section 3 also provide references to ongoing R&D projects that are currently working on implementing the solutions proposed in this roadmap. Finally, Section 4 concludes by further emphasizing the importance of a sustainable integration framework of these solutions: the end product of the roadmap for scalable semantic computing.

2. Challenges of Scalable Semantic Computing

Scalability, though a typical design requirement for most powerful high-level architectures, is often progressively suppressed throughout the development and realization of the initial architectural plans of a singular research project. Similar to the security requirements and directives of the Semantic Web vision, scalability, though initially the outright goal of the foundations of semantic technologies (particularly with their application to the World Wide Web), still remains a prominent challenge in the advancement of modern computing. It is clear that scalability will remain a prominent challenge for current ambitious research objectives which maintain the optimistic goals of a:

- Worldwide network of solutions
- Worldwide operational platform for these solutions

The addition of semantics is of course a step in the right direction (if for nothing else, then the semi-automation of the infinite number of manually dependent IT processes required in achieving the two above mentioned goals). The addition of semantics is also, however, somewhat of a burden since as with all extendible computing models, the addition of data, resources, and their accompanying tools and operational frameworks needed to utilize their contributive value increases complexity; resultantly, the prototypical solution must resolve such complexity challenges and this often (though unintentionally) leads to overly customized solutions. Thus scalability once again creeps back into the stack of challenging milestones in semantic computing.

In addition to the major scalability challenges specific to semantic computing, particularly those addressed throughout this roadmap, scalable solutions resulting from all R&D efforts (regardless of which scientific field) fall prey to two detrimental factors (though it should be noted, the application of semantics has already pacified the previous outrageous demand for sustainable integration frameworks):

Integration. Though the intentions of most large-scale integrated projects are to provide modular technologies that are compatible with the current solutions, it is often the case that the end results are customized towards particular use cases and problems which inherently leads away from potential scalability. Furthermore, regardless of conceptual proof or technical prowess, R&D results are only as powerful as their subsequent dissemination. Dissemination of scalable technologies depends entirely upon their ability to 1) integrate with both current (traditional) standard procedures (specifically, business ICT solutions) and state-of-art (new) technologies (often resulting from parallel R&D efforts); and 2) handle an infinitely growing network of information.

Sustainability. Regardless of standardization efforts, it is inevitable that a certain amount of experience based *know-how,* or *lessons learned,* or even acquired scientific research methodologies (that which is gained from informal discussions and debatable forum topics), is lost when a R&D project comes to an end. Essentially, the provision and maintenance of scalable technologies is not possible given the small scope of singular R&D projects and the rapid increase in the application of semantics to large-scale systems. True scalabilty can only be verified via practial, real-word applcations of tools and methodolgies, of which the semantic community is lacking. Even though most R&D projects invest a significant amount of resources in evaluation, feedback from a *real* user community cannot be imitated. Sustaining scalability not only requires the integration of traditional and state-of-the-art methodologies, but ongoing and future R&D efforts must additionally be collectively steered in order to avoid the further production of disparate, heterogeneous solutions (thus, inherently closed and non-scalable), as is currently the case.

The integration and sustainability challenges have been confronted by semantics. However, since semantics is often used as an ambiguous panacea, the purpose of this roadmap is to address specific areas of semantic computing that must continue to advance over the next five years in order to achieve true worldwide scalability.

The five major challenges in achieving scalability in semantic computing are:

Reasoning **-** Although scalability is a feature that demands advancements from all areas of semantic computing, the heart of the problem lies within current reasoning systems which are only built for small, closed, trustworthy, consistent, and static domains. The most advanced use cases for semantic computing today require reasoning about 10 billion RDF triples in less than 100 ms. This figure can only be expected to increase over the next five years. Current logical inference systems do not scale to the amount of information which supports large distributed networks such as the Semantic Web.

SOA - Currently, SOA is primarily used for internal integration and much less for external integration. Furthermore, very few companies offer their (internal) services to others, e.g. customers of their own. These factors alone already confine the scalabity potential of SOA. SOA must be engineered to provide support beyond internal enterprises; the real challenge is to provide an architecture which is able to integrate a network of billions of services (e.g. a semantically-enabled SOA). An integral part of a scalable SOA is a scalable middleware (although much of the literature uses these terms synonymously, we view them as two separate developments in the roadmap towards scalable semantic computing).

Middleware. Contrary to the asynchronous undirected communication model of the Web, current middleware technologies provide a framework for directed message exchange based communication. This is a detrimental contradiction as it restricts scalable advancement (i.e. the difficult combination of conflicting communication models). Scalability has been a steadily progressing research topic as the Web has evolved into the immense collection of information we know today. Middleware technologies will not meet the large-scale requirements unless they too evolve, transitioning from a notion of *messaging* to a notion of *publishing.*

Identity and Reference. Semantics has allowed for a framework that is no longer limited to a domain of digital data, but extended to include objects in other realms like products, organizations, associations, countries, events, publications, hotels or people. The solutions based upon the identifying links and references between entities (such as semantically enabled SOA or semantic computing middleware) must extend beyond hyperlinks to include virtually any type of relation.

Applications. The Semantic Web is indeed the most developed semantic oriented activity, and the results do serve as an appropriate showcase for the application of semantic computing technologies. However, scalable semantic computing is not just about supporting an infinitely growing network of information (size); the modularity and flexable application (interoperabilty) of semantic solutions to other ICT dependent systems, such as business process management, is also necearry in confirming scalabilty in semantic computing.

Again, these major challenges are not a complete summary of necessary technological advancements towards scalable semantic computing. Several other areas of semantics (e.g. ontology engineering) must also advance at the same pace.

3. A Conceptual Roadmap for Scalable Semantic Computing

The solutions to these major challenges will require a significant investment of R&D efforts over the next couple of years. For brevity, the solutions proposed in this roadmap are simply approach outlines and conceptual directives. Specific implementation plans and future technical benchmarks are beyond the scope of this work, however the reader is encouraged to consult the references (specifically those referring to ongoing R&D projects). Below are the technical advances that must take place over the next five years in order to resolve the major scalability challenges of semantic computing.

Scalable Reasoning. Research efforts must be focused upon fusing logic reasoning systems with search (in the sense of information retrieval), and seriously considering the notion of limited rationality inherent to reasoning over distributed open domains with no defined boundaries, such as the Semantic Web. *Completeness,* in such a context, is then an unrealistic requirement for an inference procedure. Thus the first step is to develop novel approaches to massive distributed *incomplete* reasoning that will remove these scalability barriers.

Scalable SOA. Over the next five years, the fundamental principles and compatible technologies of the Web, Web 2.0, Semantic Web and Context Management must be combined in order to provide a scalable SOA, one that provides on a worldwide service platform rather than one that merely supports internal integration.

Scalable Middleware. In order for semantic middleware technologies to scale as space based middleware must be developed which facilitates persistent publishing, essentially shared virtual storage, in order to transition from directed to undirected communication middleware models.

Scalable Identity and Reference. The scalability of SOA and middleware will be limited by their ability to identify and refer to information. Semantics have already enabled machines to refer to more than just data. In less than a decade, identity and reference management technologies must evolve from supporting a "Web of Documents" to a "Web of Entities." A framework must be established which implements the reuse of global identifiers for entities.

Scalable Applications. Current R&D projects place heavy emphasis on the application of semantic technologies to large-scale, real-world problems. The Semantic Web, however, is but one of many applications of semantic computing. In the next five years, efforts must be made to ensure the further dissemination and evaluation of scalable semantic computing technologies in other fields dependent upon large-scale ICT systems (e.g. business process management).

The following subsections further elaborate upon these proposed solutions, highlighting the efforts of specific ongoing R&D projects geared towards these areas of semantic computing.

3.1. Scalable Reasoning

The first step towards scalability in semantic computing is new innovative improvements in current reasoning methodologies. After all, in the narrower sense scalability can be reduced to a problem of computational logic. However herein lies a faulty presumption: that reasoning infrastructures must be founded upon paradigms that are *strictly* based in logic.

In order for reasoning to scale, a new distributed reasoning framework must be established, one which places what is traditionally referred to as *reasoning* amongst four other steps, resulting in a massive inference framework:

1. *Retrieve* raw content and assertions that may contribute to a solution.
2. *Abstract* that information into the forms needed by its heterogeneous reasoning methods.
3. *Select* the most promising approaches to try first.
4. *Reason* using multiple deductive, inductive, abductive, and probabilistic means to move closer to a solution given the selected methods and data.
5. *Decide* whether sufficiently many, sufficiently accurate and precise solutions have been found, and, if not, whether it's worth trying harder.

These sequential steps then push semantic reasoning beyond the scalability limitations of current logic based reasoning systems by incorporating methodologies and techniques from a variety of research areas such as information retrieval, cognition, ontology engineering, statistics, machine learning, probabilistic inference, economics, computational theory, decision theory, etc. A proposal for achieving these developments (i.e. providing a platform for massive distributed incomplete reasoning) over the next five years is laid out in The Large Knowledge Collider (LarKC) project [2]. While LarKC R&D objectives propose technical solutions for beyond just innovative reasoning techniques, its greatest contribution towards realizing the roadmap will be providing an reasoning framework which can perform on a scale which is well beyond what is currently possible. If such an infrastructure is not built, scalable "meaning-based computing" will never happen.

3.2. Scalable SOA

While reasoning maybe the first essential step towards scalability, current SOA must take the most ambitious step from handling in-house enterprise solutions to handling billions of distributed services. Semantics is the clear answer to bringing SOA to scale in the next five years, but this will require the combination and further development of four recent technological advances. The following steps must be taken:

1. Further development of a very open and loosely coupled architecture based upon successful *Web* principles, namely a standardized communication protocol suite.
2. Incorporation of efficient and cost-effective *Web 2.0* structured human-machine cooperation techniques.
3. Adoption of *Semantic Web* technology as a means to abstract from syntax to semantics as is required for meaningful service discovery.
4. Integration of *Context Management* as a way to process, in a machine understandable way, user needs that facilitates the customization of existing services for the needs of use.

SOA has already reached a mature level. Scalability is its final challenge. SOA4All, a large-scale EU-funded integrated project, takes the current SOA state-of-the-art in concert with a implementation plan focusing on the four steps listed above, in order to bring SOA to a worldwide scale [3].

3.3. Scalable Middleware

A necessary paradigm shift from directed to undirected communication must take place in order for semantic middleware technologies to scale. Web services middleware platform, for example, mistakenly provide Web services with a directed communication (message based) apparatus. In this regard, *Web* services would perhaps be more appropriately named as *Email* (*Message* or *SOAP)* services. Alternatively, the Web relies on an asynchronous undirected communication model: persistent *publishing* rather than *messaging*. Suddenly, limitations of directed communication are lifted and the remaining scalability middleware challenges are no worse than the challenges faced by the Semantic Web. The intended meaning, as in *Web-enabled* Web service communication, will then be reinstated.

The key to scalable semantic middleware is the development of an efficient space based model: a virtual storage and access space that allows for persistent publishing and removes the scalability limitations of message based models. The development of such a semantic space consists of the combination of three crucial technologies: tuple spaces, the Semantic Web, and Web services.

Triple Spaces, the result of the ongoing R&D project TripCom (Triple Space Communication), combines these three technologies to provide persistent asynchronous communication for Web Services. Over the next five years, a Triple Space-like model must become standardized and properly disseminated in order to overcome the scalability challenge of incorrectly modeled semantic middleware [7].

3.4. Scalable Identity and Reference

With the further development of semantics and the current overwhelming amount of publishable information, an immediate challenge facing scalability is to develop a system of uniquely identifing an infinite amount of objects (entitites). More complex methodologies are demanded of the URI in order to provide a "natural" approach to accessing information (entities and objects, rather than just data and resources). Identity and reference management must accommodate the decentralized production of interlinked content and its rapid integration in order to support scalable semantic SOA and middleware platforms. In order to achieve this over the next five years, quite a few steps must take place:

1. Issuing of globally unique, rigid identifiers for entities.
2. Enabling retrieval and reuse of entities in order to ensure unique reference to objects,

thereby allowing for seemless (correct and precise) information integration.
3. Referencing external information about entities.

OKKAM, an ongoing R&D project, focuses upon enabling such a network of uniquely identifiable of entities (Web of Entities). Over the next five years, in order to bring identity and reference to the same scalable level as other semantic technologies, OKKAM will provide:

1. A suitable infrastructure which can support the open and sustainable growth of the Web of Entities.
2. A critical mass of new entity-aware content and data accessible to a very large number of users in a relatively short time.
3. A collection of exemplary and high impact applications [1].

3.5. Scalable Applications

Scalability challenges rise to the surface in two main scenarios: increase in the size of the problem and increase in the diversity of the problem. The Semantic Web caters more heavily to resolving the former, however effort from other IT fields is lacking in support for the latter. In order to ensure the development of scalable semantic applications, semantics must be applied to other academic and industrial scenarios. If semantic solutions are not applied to diverse problems (derived from multiple fields of research and industry), then scalability is not possible.

The failure of business process reengineering and the emergence of business process management (BPM) with the primary objective of shifting the control from the IT experts to the business experts inspired the semantic community to contribute. Since 2006, the SUPER (Semantics Utilized for Process Management within and between Enterprises) project has been developing a technological platform constituting BPM enriched with machine readable semantics, which not only allows business experts to manage business processes on their own, but also provides means to further automate the business process life cycle. The challenge in combining Semantic Web Services and BPM are manifold: the heterogeneity of information sources, the complexity of the processes to be managed and their inter-dependencies, as well as the high robustness and scalability requirements at intra- and inter-enterprise level [6].

Over the next five years, steps must be taken to ensure that the results of such projects are properly evaluated, brought to a mature level, and effectively standardized and disseminated; if semantics cannot scale beyond their favorite Semantic Web showcase, then this roadmap will not be achieved.

4. Web of Services

If this roadmap is fulfilled, then the ideal showcase of semantic computing scalability will be a Web of Services. In this context, *services* no longer carries the same connotations as in Web *services*; instead, services should be understood as a generic provision which can be customized to meet a set of needs, whereby the recipient (be it user/agent) is not at all aware of which application/process/service is actually being provided. Software, as well as Web Services and Semantic Web Services, are encompassed into the term service. The Web then provides the communication protocols necessary to deliver the service. And though the dichotomy was never so discrete in the first place, at least we stumble upon a point in computing history where things seem to have come full circle: we return to a time when business provisions are again reduced to simple goods and services.

A Web of Services represents a network of maximum scalability since it provides the infrastructure for the largest knowledge set imaginable (the future Internet). It would be naïve to propose that such a Web of Services could be achieved by resolving the five major scalability challenges mentioned in this roadmap alone; however confronting these conceptual problems at an early stage is an important progressive step.

Second to the development of these solutions is their simultaneous and subsequent coordination (or integration and sustainability as mentioned in Section 2). Service Web 3.0 is an EU-funded coordinated support action which addresses these emerging developments, contributes to the implementation of framework programs, and their projects, and supports the preparation of future community research and technological development [4]. Under normal circumstances, much of the work done in Service Web 3.0 would seem quite frivolous given it’s short duration (2 years) and rather ambitious tasks. However, the objectives of Service Web 3.0 directly coincide with the purpose behind Semantic Technology Institute International (STI2), a worldwide organization that coordinates, integrates, and consolidates the semantic-related activities of interested academic, industrial, and governmental organizations [5]. In order to ensure that the accomplishments in this roadmap will be coordinated with one another, and paralleled with the directives of leading standardization consortiums such as W3C and OASIS, organizations like STI2 will continue collaborative research after the R&D projects have come to a close.

The following figure (Fg. 1) shows an estimated five-year plan representing the challenges and the soltuions mentioned in this roadmap, in accord to termination dates of particular ongoing R&D projects with related research objectives.

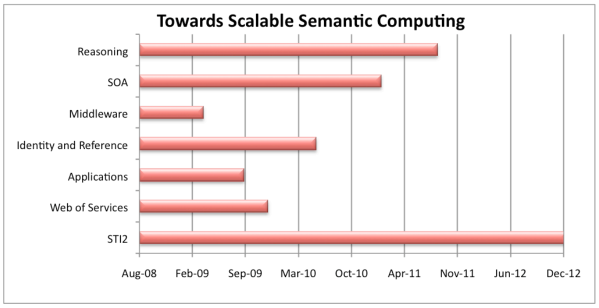

Figure 1. Five-year R&D timeline

The projected timeline shows two significant factors reflective of the survey of the R&D projects discussed in this roadmap. For one, over the next year more effort must be allocated towards the further dissemination of the outcomes from terminating projects, namely those focused on the development of scalable middleware and applications; additionally, the initial steps to acquire funding to continue research in these areas should be taken.

The second significant factor is the burden placed upon organizations such as STI2. Ambitious goals such as promoting the standardization and commercialization of semantic technologies require the infrastructure and guidance provided by STI2; for these goals are far less achievable for single project consortiums or institutions, especially considering the limited timeframe of average R&D projects. Hopefully the timeline serves to promote investment in not only further technical research, but further supportive and coordinated actions (e.g. Service Web 3.0) as well.

The timeline does not mean to correlate the successful integration of these technologies. Rather, the final year will be dedicated towards harmonizing these research efforts. While this is normally quite an ambitious task, projects such as Service Web 3.0 and organizations like STI2, promote collaborative agreement wherever cross-project synergies exist (from the conceptual planning stages to final technical implementation) in order to reduce the subsequent integration effort required to provide scalable solutions.

5. Conclusion

The wheels have been set in motion and semantic technologies are slowly emerging to a level of maturity. It is an opportunity to reflect upon how far we've come, and to focus on what the next steps must be. Most importantly we must remember not to hinder ourselves by worsening the problem we originally set out to solve. If true worldwide scalability is to be achieved, research directives and roadmaps such as this one must be set in place. We are already progressing to a point where the semantic prototypes and beta applications are starting to enumerate; a coordinating infrastructure is required to collaboratively drive the fields of semantic research; particularly, but not limited to, resolving the following scalability challenges: reasoning, SOA, middleware, identity and reference, and applications. This brief roadmap has not tried to cover all areas of semantic computing that focus upon scalability, only those which we view as the most challenging and achievable major milestones in the next five years of research and development. The result of this five-year plan will be a culmination of technological advances in these areas: a Web of Services.

We must learn from our own mistakes. Do not set out to solve scalability with disparate R&D objectives, and then entirely rely upon proclaimed pluggable interfaces to support integration and sustainability. From the initial stages to final implementation, collaboration is paramount in creating a worldwide solution.

Acknowledgement. This work is based upon the collaboration between several European R&D projects. It is supported by the European Commission through the work of the following projects: OKKAM (ICT-2007.4.2, Ref. 215032), LarKC (ICT-2007.4.2, Ref. 215535), SOA4All (ICT-2007.1.2 Ref. 215219), Service Web 3.0 (ICT-2007.1.2, Ref. 216937), SUPER (FP6-026850), and TripCom (IST-4-027324-STP). Special thanks also goes to those project representatives who actively contribute to STI International.

5. References

[1] Enabling the Web of Entities. A scalable and sustainable solution for systematic and global identifier reuse in decentralized information environments (OKKAM). FWP 7 Collective project (ICT-2007.4.2), Ref. 215032. www.okkam.org.

[2] Large scale semantic computing semantic Web technologies distributed reasoning probabilistic reasoning web-scale inference information retrieval (LarKC). FWP 7

Collaborative project (ICT-2007.4.2), Ref. 215535. www.larkc.eu.

[3] Service oriented architectures for all (SOA4All). FWP 7 Collaborative project (ICT-2007.1.2), Ref. 215219. www.soa4all.org.

[4] Service Web 3.0. FWP 7 Coordination and support action (ICT-2007.1.2), Ref. 216937. www.serviceweb30.eu.

[5] STI International, www.sti2.org.

[6] Semantics Utilized for Process Management within and between Enterprises (SUPER). FP6-026850. www.ip-super.org.

[7] Triple Space Communication (TripCom). FWP 6 (IST-4-027324-STP). www.tripcom.org.