

A Trilogy of Webs for Machines

Alexander Korth¹, Benjamin Hirsch², Till Plumbaum², and
Andreas Nürnberger¹

¹ Otto-von-Guericke Universität Magdeburg, Data & Knowledge Engineering Group,
Magdeburg, Germany

² Technische Universität Berlin, DAI-Labor, Berlin, Germany

Abstract. In the coming years we will see a revolution of machine knowledge and abilities which will emerge from different activities and trends in three distinct areas connected to the Internet: The emerging *Web of Data*, the *Web of Services*, and a new area, the *Web of Identities*. These areas, the webs, are about creating accessible and processable semantic knowledge about data, functions and individuals, respectively. They form a knowledge layer for a new era of applications. In this work, we will provide universal definitions for these webs, with a focus on the new introduced Web of Identities, and show how they are related. Furthermore, we outline current activities and trends, and point out the impact of the webs on the future of the Internet based on scenarios exploiting the blended knowledge of these activities.

1 Introduction

The amount of information is growing exponentially. Sources are developed or made accessible, content is produced en masse through the paradigm shift of consumers becoming producers in the Web 2.0 [14]. Now, users create data about themselves and their social connections and more and more companies are unlocking services and data.

Every day it becomes more complex and difficult to make this information accessible and usable. In the near future, a search engine's one dimensional list of ten results returned on the first page for a query will not be sufficient to cover the user needs. Users will demand to ask more sophisticated, natural queries and they assume that the system knows and understands the context in which they are asking their query. To answer those, today's search engines' primitive semantic understanding of the content they index is not sufficient as a knowledge base.

It is indispensable that machines are taken to another level of understanding. Machines have to understand the sense of the entered query, what a term means and what concepts are linked to it. Moreover, machines have to understand such a query to find a suitable service to answer that request. Therefore, they have to understand existing services, detecting new services, understanding what they do, and invoking them to generate further (or to process existing) information. Finally, all these actions are conducted because of a user query and thus, machines have to understand what a user wants.

In this work, we sketch an answer to the challenge of providing tomorrow's machines with a toolbox to find and interpret knowledge, to discover, orchestrate, and invoke services with the knowledge gained to solve highly complex tasks.

In the next section, we discuss selected activities in the areas of the Web of Data, the Web of Services and the Web of Identities and illustrate how they interrelate and form an emerging big picture of a Trilogy of Webs for Machines.

2 The Webs

Current research activities focus on making accessible semantic knowledge from open data sets (*Web of Data*) and semantically annotated services (*Web of Services*) to machines. Emerging within the Social Web, we want to introduce a new area: the *Web of Identities*. This web is about semantically interlinked users, their assets, facts, preferences, social graphs, etc. This content is highly privacy-sensitive, volatile and valuable. Compared to the Web of Data, user-centric access control mechanisms for privacy preservation are needed.

In the following, we want to introduce and explain all three Webs.

2.1 Web of Data

Definition 1 (Web of Data). *The Web of Data is a distributed web of interconnected data sets of semantically annotated data.*

Problem Motivation. The idea of the Web of Data originated within the Semantic Web [4]. The inability of machines to understand Web pages to a significant extent led to several initiatives to overcome this weakness. Initially, the aim of the Semantic Web was to invisibly annotate Web pages with a set of meta-attributes and categories in order to enable machines to interpret parts of the knowledge included in the text and to put it into some kind of context. This approach did not succeed since the annotation was rather complicated for humans with no technical background and it therefore got stuck in the bootstrapping process. Markup-based approaches like Microformats [12] and RDFa (Resource Description Framework in attributes) [1] follow a similar idea of annotating Web pages. On the other end of the technology spectrum we find full-blown ontologies that describe domain knowledge with the help of formal logic. This allows inference of new information from a set of facts, but is difficult to do with the right level of detail.

All these approaches have in common that they try to improve the machine-readability of Web pages that are designed for humans. But the horizon or depth of machine-readable knowledge that can be added to a page is limited: only the page itself and particular elements on it can be marked-up by applying these approaches.

This limitation and the fact that nowadays there are data sets containing lots of structured data about all kinds of information distributed over the world lead to the idea of creating a Web of Data: If these data sets are semantically described

and interconnected, a machine can traverse through this web to gather semantic, noise-free knowledge about arbitrary entities and domains, independent of the information contained in the original Web page.

State of the Art. A promising approach is the W3C SWEO Linking Open Data community project³ [5] (LOD). The project is uncoupled from the Web for humans and interconnects open data sets. The data sets contribute by granting access to their semantically linked knowledge and by linking to items of the same or other data sets. This way, the project follows basic design principles of the World Wide Web [3], e.g. simplicity, tolerance, modular design and decentralization. The LOD project currently counts more than 6.4 billion RDF triples⁴.

The LOD data sets can be accessed in heterogenous ways, e.g., through a Semantic Web browser or crawled by a spider of a semantic search engine.

Conclusion. With every fact and link added to the Web of Data, more general and specific knowledge is made accessible to machines. The Web of Data will enable a whole new generation of services. Through the semantic structuring of the data within the data sets and the interconnection of lots of different data sets, highly sophisticated queries become machine-processable and can be answered through a next generation of search services. Querying languages like SPARQL [15] and RQL [11] are already available.

2.2 Web of Services

Definition 2 (Web of Services). *The Web of Services is a distributed web of semantically annotated services.*

Problem Motivation. The Services sector has become the world's biggest business sector forming 64% of the world-wide Gross Domestic Product (GDP) [10]. This sector has a pressure to make their services easier and more widely accessible, as well as to adapt to ever faster changes in the market environment.

The Service Oriented Architecture (SOA) paradigm has become the predominant approach to (enterprise) software engineering, to streamline the IT infrastructure within an organization as well as to interact with external entities. Its principles [8] call for services that have their (formally described) interface decoupled from its functionality and described in an abstract fashion. While SOA can be implemented with a host of different technologies, Web Services have become the technology of choice.

³ esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData, accessed Oct 2010

⁴ esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData, accessed Oct 2010

However, it should be noted here that while services are all the rage, there is no clear definition as to what constitutes a service, neither on a technological nor conceptual level. In the former case, Web Services offer a quasi-standard, but the subset of standards that is agreed upon is neither powerful nor expressive enough to handle the possible applications of services. The two main technologies for webservices, WSDL and REST, implement two different views to services. In the former case, services are thought of as coarse grained blocks of functionality that are composed to higher level services, mostly in the context of enterprise applications and SOA, while the latter is used by many websites to provide grammatical access to their service. For both technologies, the standards so far lack any semantic description, making its use within automated scenarios and higher-level services impossible because machines can not understand what a service does.

Having said all that, today there are already all kinds of services with all levels of complexity on the Web and the number of them is expected to grow exponentially. The services follow different standards and a lot of them are proprietary, uni-directional and designed to be used by humans to mash-up something new. There are editorial catalogs, e.g. ProgrammableWeb⁵, designed for humans searching for a particular service. A lot of Web 2.0 services provide services to read existing or create new data exposing almost all of their functionality through their Application Programming Interface (API)⁶. As mentioned above, Web Services follow an agreed-upon standard dealing with the service definition but still lack a semantic description. While there are a number of different approaches to adding a semantic description to Web Services, such as OWL-S [13], WSMO [9], or WSDL-S [2], none has so far managed to break out of its academic confinement.

Once services are annotated semantically, they can be accessed by machines automating service discovery, execution, billing or revenue sharing, orchestration, replacement on failure based on experience (Quality of Service) etc. These Web Services will be brought together in a Web of Services according to Web principles.

State of the Art. Many works deal with the topics Internet of Services and SOA in general, in research as well as industry. We will focus here on the larger research projects. Closest to our idea is the SOA4All project⁷. It addresses the motivated issues through four cornerstones [7]: Firstly, Web principles and Web technology as the underlying infrastructure are used in the Web of Services. Secondly, they plan to implement user participation in terms of, e.g., ranking of services. Thirdly, they want to facilitate Semantic Web technology to abstract from syntax to semantics to grant machines knowledge about the services. Last

⁵ www.programmableweb.com, accessed Oct 2010

⁶ In 2007, Twitter counted ten times more traffic on their API than on their website (www.readwritetalk.com/2007/09/05/biz-stone-co-founder-twitter, accessed Dec 2008)

⁷ www.soa4all.eu, accessed Oct 2010

but not least, they plan to implement a context management to enable processing of user requirements when it comes to service contracting or orchestration.

The TripCom project⁸ concerns itself with the design and implementation of an architecture for application integration based on the combination of Semantic Web, Web Services, and tuple spaces, called the triple space service technology. There, services can persistently publish semantically annotated data in order to facilitate orchestration and choreography of services.

The SHAPE project⁹ provides a unified approach to the definition of semantically enhanced SOA. The focus lies on the integration of model-driven approaches with semantics and SOA.

Somewhat orthogonal to the development of services is the large research area of multi agent systems [17]. There, similar to SOA, distributed agents communicate and cooperate to achieve some goal. Where services are generally considered passive however, agents are autonomous and proactive. Given some semantically described goal, an agent tries to bring about a situation where the goal holds true. He does this by interacting with other agents, cooperating with them to change the state of the world. In the context of the Web of Services, agents play a vital role in that they, at least in the realm of academia, already created a Web of Services, where machines, i.e. agents, autonomously searched for functionality and used different services based on their semantically described capabilities. The Agentcities [6] project tried to create a global, open, heterogeneous network of agent platforms and services to which any agent researcher could connect their agents. Services could automatically be offered and used.

Conclusion. The Web of Services will enable machines to work with a huge toolbox of functionalities. Services might answer queries (from humans or other services), or create further knowledge which could also flow back to the Web of Data. Automated service orchestration and service chaining will be an important tool to quicken innovation cycles.

2.3 Web of Identities

Definition 3 (Web of Identities). *The Web of Identities is a distributed web about people: their personae, their social graphs and their assets. It provides privacy-preserving access to user profile information.*

Problem Motivation. Web 2.0 brought a paradigm shift in terms of user participation and contribution. The distinction between content consumers and content producers got blurred. The world-wide impact on user participation is immense: e.g., the world's biggest Social Network Service (SNS) Facebook¹⁰ currently counts more than 500 million active users sharing 30 billion pieces of

⁸ www.tripcom.org, accessed Oct 2010

⁹ www.shape-project.eu, access Oct 2010

¹⁰ www.facebook.com

content (web links, news stories, blog posts, notes, photo albums, etc.) every month¹¹.

The major problem of these SNSs is a conflict of interests in terms of data ownership and privacy: The application providers put all their efforts in the growing of the user-base and content-base and usually own all the data that is added to their site. As the user-base and content-base directly influences the enterprise valuation, the companies are not interested in disclosing user or content details¹². Currently, the users are not aware of these problems but as the number of services that the users access grows, the downsides get obvious: The user has to re-enter his personal information, re-enter his preferences, re-enter his relationships to other users time and again. No site provides sophisticated data synchronization features and the user never controls, possesses or owns his data.

To overcome these issues, we envision a solution which we call the Web of Identities. In it, an interconnected web of Identity Providers (IDP) take care of their customer's data. IDPs host all of the users' data, e.g. their identities, personae, social graphs, groups, messages, files, comments, and presence information like current geographic location, available device and connection for calling or messaging. From a features point of view, IDPs have to provide all needed management, privacy, access control, security, trust, authorization, authentication, and accounting (AAA) functionality. Given the permission of the user, third party services, e.g. an SNS, can read and write needed fragments of this data. The user is in full control of his data being hosted and exposed. In this scenario, third party services utilize the data of the IDP's database they are allowed to access on behalf of the users. The user profile, the social graph, etc. can be synchronized with the IDP. The third party service only has to implement the delta of data and functionality that is not provided by the IDPs. Each user of a particular third party service can have his identity information hosted by a different IDP of his choice.

As the Web of Data and the Web of Services, the Web of Identities should follow basic Web principles and a set of agreed standards. Currently, we see approaches being developed both, from scratch and by big platforms. Only if an agreement on standards can be achieved, the Web of Identities can emerge and flourish.

State of the Art. There are bottom-up as well as top-down approaches, some are driven by commercial interests and some are non-profit. All of them are to a certain extent coherent with our vision.

First of all, we want to name the non-profit OpenID [16], which initially focused on providing a distributed SSO(Single-Sign-On) solution for authentication. The OpenID framework is designed with respect to the Web design principles and is completely open. The framework will most certainly become the identity and authentication foundation for IDP implementations.

¹¹ www.facebook.com/press/info.php?statistics, accessed Oct 2010

¹² the companies' behavior is called lock-in, walled garden or data silo

A big picture for all bottom-up approaches is drawn by Marc Canter with his Open Mesh¹³. Canter outlines a vision of what building blocks are needed and how they could be put together in a common infrastructure.

The non-profit DataPortability¹⁴ group deals with the establishment of open standards and protocols for the exchange of data between applications and vendors. The protocols and standards are already widely agreed upon and now need to be further diffused and adopted. The development of open, non-proprietary specifications for Web technologies is also the dedication of the non-profit Open Web Foundation¹⁵. A distributed initiative of providing SNS building block functionality is taken by the non-profit DiSo Project¹⁶ (Distributed Social Networks). The team implements a WordPress plugin that implements some of the standards supported by the DataPortability group.

A more aggregated approach is the non-profit OpenSocial Foundation¹⁷. It creates specifications intending to spread social content and functionality across the Web following a centralistic approach.

The EU-funded PrimeLife¹⁸ project aims at bringing life-long privacy and user-control over personal information and autonomy to the Information Society. The project at its current state is promising to end up as an IDP implementation. For this to become reality, it is crucial that a set of standards and protocols for future interoperability within the Web of Identities are prevailing shortly because the project did already spend two thirds of its runtime.

Last but not least, some big players on the IDP market, e.g. Google, Yahoo, Facebook, are opening slowly. On the one hand, they do not want to expose data but on the other hand they want to extend their reach outside of the platform and to keep up the meaning they make to their users, so they open bit by bit. Features like Google's Friend Connect¹⁹ or Facebook's Connect²⁰ all aim at spreading fragments of the platforms' features to outside of the platform. That serves the users' needs of accessing locked-in data for a transitional time but does not solve the ownership issue.

Conclusion. We see the attention and need for this solution rising and all the named activities converging in the Web of Identities. We are sure the directions and corrections will be driven by user needs and the market. Research has to be done in the areas of empowering the user to take control of her data. Features like reach control, revokable access rights and the management of which third

¹³ blog.broadbandmechanics.com/2008/05/how-to-build-the-open-mesh, accessed Nov2010

¹⁴ www.dataportability.org, accessed Nov 2010

¹⁵ www.openwebfoundation.org, accessed Nov 2010

¹⁶ www.diso-project.org, accessed Oct 2010

¹⁷ www.opensocial.org, accessed Nov 2010

¹⁸ www.primelife.eu, accessed Dec 2008

¹⁹ www.google.com/friendconnect, accessed Dec 2008

²⁰ developers.facebook.com/connect.php, accessed Dec 2008

party service can read or write what fragment of user data are necessary but very hard to translate to an intuitive user interface and user experience.

If this vision comes true, we will see a user centric, user friendly, privacy preserving and meaningful tool. Users can explicitly grant online marketers access rights to attention data or purchasing history data to empower them to target meaningful ads that may take into account what direct friends recommend.

For both, machines and applications of the emerging Social Web, the Web of Identities is a very important infrastructure for looking up user-related private, volatile personal and contextual data.

3 Interplay of the Webs

With the Trilogy of Webs as a backbone, the Internet as a tool will change because through interconnected knowledge and toolsets, machines are catapulted to a new ability level. New services will emerge based on the foundation of the Webs. From the Human Computer Interaction (HCI) side, the way we use existing services will change dramatically.

The distinction between data items stored and retrieved and the use of services will continue to blur until requests will freely traverse the webs, retrieving items of data, feeding chains of services that use personal information from the stored identities.

The following example scenarios give an impression of how the webs interlink.

3.1 Scenario: Social Recommendations

Imagine a user visiting a search engine system that is based on the Webs. He queries *Recommend books about Berlin for my mother for Christmas*. From the Web of Data, the system gathers general knowledge about the terms *mother*, *Berlin*, *Christmas*. The system proceeds by querying a service that indexed the Web of Data about all books covering Berlin or authors born or living in Berlin. Given permission from the user, his IDP is called to return his mother's identity Uniform Resource Identifier (URI) from his social graph. His IDP searches all of his personae for his mother and finds her in his private persona's social graph. The mother's IDP is called to access her interest information limited to the wider topic fields *books* and *Berlin*. The mother's personae's social graphs are searched for a link back to Peter. As the private persona is found, that persona's information is selected for access limitation. From the private persona, the mother's IDP returns a set of information the mother explicitly granted access to. The set contains general interests, some purchases, reviews, comments, ratings and some attention data. URIs of the mother's private friends are also returned. The system continues by querying the mother's closest friends' IDPs if one of them liked or recommends books about Berlin since friends' recommendations are the most valuable. The system identifies the term *recommend* as a service request term and searches the Web of Services for a recommendation service that can handle books, personal interests and recommendations as filtering and ranking

criteria. The initial set of books the system retrieved from the Web of Data and the information collected from the Web of Identities is now sent to a filtering and ranking service. As the term *Christmas* is recognized as contextual term for the task, the system now searches the Web of Services for e-commerce services offering books. The filtered and ranked list of books is sent there to retrieve price proposals with a delivery date before December 24th. Finally, the list of books is augmented by prices and dates and presented to the user. The system tracks feedback for the book recommendations and assigns it to its QoS ratings for the services invoked.

3.2 Scenario: Mass Customization

Think of a user that recently graduated from university. She knows that she needs an insurance package but has no idea what exactly it should consist of. She heard of this intelligent insurance packaging brokerage system which she now visits with her browser. She logs into the system with her ID. From the Web of Identities and with her permission, the system initiates a profile lookup at her IDP to gather information needed for the configuration of the components of the insurance package. It queries for information like private address, marriage status, age, and gender. Since it cannot find her current income, it prompts her directly. From the Web of Data, the system now queries for her neighborhood's crime statistics for risk estimates. The system now looks up all insurance services it can find in the Web of Services. It configures the services with the knowledge gathered, selects the best offers and combines them to a personalized insurance package. The package consists of products from different insurers around the world. She signs the contracts through the broker and logs out with the satisfaction that she now is optimally and neither under- nor over insured.

4 Conclusion

In this paper, we have outlined the concepts of the Web of Data and the Web of Services, and introduced the Web of Identities. We have demonstrated how, in parallel to the Web for humans, these interplaying Webs will provide a new level of machine understanding and interoperability which one could see as common sense for machines.

We want to note that it is indispensable that all Webs ensure *security*, *privacy* and *trust*, internally as well as in their interaction. The notions as described here however allow including any mechanisms that support these, just as the World Wide Web provides the basis for secure transactions without prescribing technologies.

Our vision of three interlinked, yet clearly definable, areas within the Future Internet allows for focussed research and development in either each of the webs or in their interactions, some of which we hinted at in the scenarios. However, other interactions are of course possible. We believe that research areas and business cases can and will arise from the Web for Machines as described in this paper.

References

1. Adida, B., Birbeck, M., McCarron, S., Pemberton, S.: RDFa in XHTML: syntax and processing, W3C recommendations (2008)
2. Akkiraju, R., Farrell, J., J. Miller, Nagarajan, M., Schmidt, M., Sheth, A., Verma, K.: Web service semantics - WSDL-S, a joint UGA-IBM technical note, version 1.0. Tech. rep., IBM and the University of Georgia (April 2005)
3. Berners-Lee, T., Fischetti, M.: Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor. Harper (1999)
4. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American* 284(5), 28–37 (2001)
5. Bizer, C., Heath, T., Idehen, K., Berners-Lee, T.: Linked data on the web (LDOW2008). In: Proceedings of the 17th International Conference on World Wide Web (WWW). pp. 1265–1266. ACM (2008)
6. Dale, J., Burg, B., Willmott, S.: Innovative Concepts for Agent-Based Systems, Lecture Notes in Computer Science, vol. 2564/2003, chap. The Agentcities Initiative: Connecting Agents Across the World, pp. 453–457. Springer (2003)
7. Domingue, J., Fensel, D., Gonzalez-Cabero, R.: SOA4All, enabling the SOA revolution on a world wide scale. In: Proceeding of the 2nd IEEE International Conference on Semantic Computing (ICSC). pp. 530–537. IEEE Computer Society (2008)
8. Erl, T.: Service-Oriented Architecture: Concepts, Technology, and Design. The Prentice Hall Service-Oriented Computing Series from Thomas Erl, Prentice Hall, Indiana, USA (August 2005)
9. Fensel, D., Lausen, H., Polleres, A., Bruijn, D., Jos, Stöllberg, M., Roman, D., Domingue, J.: Enabling Semantic Web Services: The Web Service Modeling Ontology. Springer, Berlin, Heidelberg (2006)
10. Information Society Technologies Advisory Group: Working group on web-based service industry. ftp://ftp.cordis.europa.eu/pub/ist/docs/web-based-service-industry-ista%g_en.pdf (accessed Dec 08) (2008)
11. Karvounarakis, G., Magkanaraki, A., Alexaki, S., Christophides, V., Plexousakis, D., Scholl, M., Tolle, K.: RQL: A functional query language for RDF. In: Gray, P.M.D., Kerschberg, L., King, P.J.H., Poulouvasilis, A. (eds.) The Functional Approach to Data Management: Modelling, Analyzing and Integrating Heterogeneous Data. pp. 435–465. LNCS, Springer-Verlag (2004)
12. Khare, R.: Microformats: The next (small) thing on the semantic web? *IEEE Internet Computing* 10(1), 68–75 (2006)
13. Martin, D., et al.: OWL-S: Semantic markup for web services. <http://www.daml.org/services/owl-s/1.0/owl-s.pdf> (accessed Dec 08) (2004)
14. O'Reilly, T.: What is Web 2.0: Design patterns and business models for the next generation of software. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web%-20.html> (accessed December 08) (September 2005)
15. Prud'hommeaux, E., Seaborne, A.: SPARQL query language for RDF. www.w3.org/TR/rdf-sparql-query (accessed Dec 08) (January 2008)
16. Recordon, D., Reed, D.: OpenID 2.0: a platform for user-centric identity management. In: Proceedings of the second ACM workshop on Digital Identity Management (DIM), pp. 11–16. ACM, New York, NY, USA (2006)
17. Wooldridge, M.: An Introduction to Multiagent Systems. John Wiley & Sons (2002)