Relationship Web: Realizing the Memex Vision with the Help of Semantic Web

Amit Sheth
Kno.e.sis Center, Wright State University
Dayton, Ohio, 45435 USA
amit.sheth@wright.edu

ABSTRACT

Relationship Web takes us from "which document" could have information I need to "what's in the resources" that gives me the insight and knowledge I need for decision making.

Dr. Vannevar Bush outlined his vision for Memex in a 1945 *Atlantic Monthly* article [1]. Describing how the human brain navigates an information space in what he called trailblazing, Dr. Bush said, "It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain." Now that we can label content to associate semantics (meaning) to data and build information processing in which relationships rather than keywords and entities play the central role, the possibility of realizing the Memex vision seems tantalizingly close.

Although through much of the recent past attention has been on search, finding a document is seldom the end goal of a human activity. Aligned with the Memex vision, human need for information is related to a desire and need for information processing that goes well beyond delivering a list of documents that matches the keywords or even the implied intent. Human information seeking is likely to be driven by more demanding activities such as interaction and entertainment, finding associations and answers, performing analysis, gaining insights, or making decisions. The Memex vision provides an interesting paradigm for supporting these objectives.

Changing the computing paradigm to one that focuses on relationships is the key to realizing the Memex vision. We term our realization of Memex *Relationship Web*. In past work we observed the changing focus from documents to entities to relationships. We also investigated a broad variety of issues related to modeling, validating, discovering, and exploiting the many types of relationships between entities in content [2]. The first result of these efforts was the concept of Metadata Reference Links (MREFs), which proposed associating semantic metadata with hypertext links [3]. MREF faced several limitations, but recent significant advances resulting from research, standards, and technology development associated with Semantic Web provide building blocks for realizing the Relationship Web. We outline below some recent relationship-centric research to which we have had the opportunity to contribute, at the same time acknowledging extensive work in each area by many researchers and practitioners.

MS'07, September 28, 2007, Augsburg, Bavaria, Germany.
ACM 978-1-59593-782-7/07/0009.

Ontology Development: We now have languages, techniques, methodologies, technologies, and tools for creating and managing ontologies. In domains such as biomedicine, many open-source, extensible, and reusable ontologies have been created and are in use, while in other domains we know how to build ontologies as needed. Equally important, we understanding that agreement (ontological commitment) is critical to interoperability and integration. We can represent agreements that range from "little semantics" to "deep semantics" and that span domain vocabularies, nomenclatures, folkonomies, taxonomies, classifications, and ontologies, each of which provides different capabilities to improve search, integration, and analysis of information.

Semantic Metadata Extraction and Annotation: It is now possible to extract focused (persons, addresses, events, etc.) or domain-specific metadata from unstructured, semistructured, and structured data, including enterprise and scientific data, open Web and deep Web data, in a scalable way [4]. In addition to entities, it is also possible to extract *Simple Relationships* and *Complex Relationships* [5]. Imagine representing much of the information in large scientific repositories such as PubMed in RDF that models relationships as first class objects. This is now in the process of being realized!

Semantic enrichment: Enterprise and scientific databases, along with other deep Web content, store a lot of the world's information. What if these databases were represented in RDF and, in the process, additional implicit or indirectly represented semantic information were explicitly represented as named relationships? This is indeed being done, making the original data even more valuable, as exemplified by mapping NCBI's Entrez Gene into an RDF dataset [6].

Modeling and representation of relationships: We now understand various types of relationships and have found way to model, represent, and compute them. An interesting dimension along which we can study relationships, including implicit and formal relationships, was introduced in [7] and recently extended with a type we call Explicit Linguistic Relationship. Representation of complex relationships such as those needed to model complex natural phenomena is also attempted [8].

Query language with support for relationships: Extending query languages with the ability to compute and rank semantic associations and complex relationships [9] is important to support powerful relationship computing. Semantic Web languages such as SPARQL can be extended with powerful operators, as seen by the addition of a path operator in SPARQ2L [10]. Emergent Relationship Web capabilities have already started to power new techniques such as semantic browsing, hypothesis-based document retrieval, and knowledge discovery over all forms of data (text, scientific data, structured data repositories and deep Web data, Web 1.0, Web 2.0 content, and so on).

How does the Relationship Web relate to the Semantic Web? We see the Semantic Web as an enabler of the Relationship Web. What metadata, annotation, and labeling are to the Semantic Web, relationships of all forms (implicit, explicit, and formal) are to the Relationship Web. The primary goal of the Semantic Web has been described (by Tim Berners-Lee and many others) as integration of data or labeling of Web resources for more precise exploitation by both machines and humans. At the next level, the Relationship Web organizes Web resources for analysis that goes beyond integration to trailblazing, leading to deeper insights and better decision making.

We believe the time is right for a shift from the keyword-in-document-out paradigm to something that will provide more insight into content. Leveraging the power of relationships, we have proposed the idea of a Relationship Web [11]. We have suggested how this will transform search, browsing, and retrieval of Web resources, taking the user one step closer to an automated interpretation of Web content. We anticipate that preliminary forms of a Relationship Web will take shape in specific domains such as the biomedical. We plan to create one such Relationship Web for literature from PubMed pertaining to urological diseases. As the scalability and quality of semantic metadata extraction grows, larger community-wide Relationship Webs over heterogeneous Web content will emerge. This reminds us of a dialog by Jack in Terry Gilliam's 1985 film, *Brazil*: "Everything's connected, all along the line. Cause and effect. That's the beauty of it. Our job is to trace the connections and reveal them."

Categories and Subject Descriptors
H.2 Database Management, H.3 Information Storage and Retrieval, H.5 Information Interfaces and Presentation

General Terms
Algorithms, Design, Experimentation, Human Factors.

Keywords
Relationship Web, Semantic Web, Trailblazing, Semantic Browsing

SPEAKER BIO
Amit P. Sheth is LexisNexis Ohio Eminent Scholar and director of the Knowledge Enabled Information and Services Center (http://knoesis.wright.edu) at Wright State University. Earlier he founded and directed the LSDIS lab at the University of Georgia and worked at Bellcore, Unisys, and Honeywell. His research interests include the Semantic Web, Services Science, and information integration and analysis. His research has resulted in several commercial products; he founded two companies in the area of workflow management and Semantic Web. Sheth is a fellow of the IEEE. He is also Editor-in-Chief (EIC) of the International Journal on Semantic Web and Information Systems (http://ijswis.org), co-EIC of Distributed and Parallel Databases Journal, and editor of IEEE Internet Computing and other journals.

ACKNOWLEDGEMENTS
Cartic Ramakrishnan is my primary collaborator on this work. This work is funded by NSF-ITR- IDM Award #0714441, titled "SemDIS: Discovering Complex Relationships in the Semantic Web."

3. REFERENCES

[1] Bush, V. As We May Think. *Atlantic Monthly, 176, 1* (1945), 101-108.

[2] Sheth, A., Arpinar, I.B., and Kashyap, V. Relationships at the Heart of Semantic Web: Modeling, Discovering and Exploiting Complex Semantic Relationships. In *Enhancing the Power of the Internet Studies in Fuzziness and Soft Computing*, Nikravesh, et al., Eds. Springer-Verlag, 2003.

[3] Shah, K., and Sheth, A. *Logical Information Modeling of Web-Accessible Heterogeneous Digital Assets*, in *Proc. Forum on Research and Technology Advances in Digital Libraries*. IEEE Computer Soc. Press, Los Alamitos, CA, 1998, 266-275.

[4] Hammond, B., Sheth, A., and Kochut, K. Semantic Enhancement Engine: A Modular Document Enhancement Platform for Semantic Applications over Heterogeneous Content. In *Real World Semantic Web Applications*, V. Kashyap and L. Shklar, Eds. Ios Press, 2002, 29-49.

[5] Ramakrishnan, C., Kochut, K.J., and Sheth, A.P. A Framework for Schema-Driven Relationship Discovery from Unstructured Text, in *ISWC 2006*. 2006. p. 583-596.

[6] Sahoo S., Bodenreider, O., Zeng, K., and Sheth, A. Adapting resources to the Semantic Web: Experience with Entrez Gene (pdf). W3C Semantic Web Health Care & Life Sciences Workshop at ISWC 2006.

[7] Sheth, A.P., Ramakrishnan, C., and Thomas, C. Semantics for the Semantic Web: the Implicit, the Formal and the Powerful. *Int'l Journal on Semantic Web and Information Systems, 1,1* (2007), 1-18.

[8] Sheth, A., Thacker, S., and Patel, S. Complex relationships and knowledge discovery support in the InfoQuilt system. *VLDB Journal: Int'l Journal on Very Large Data Bases,12, 1* (2003), 2-27.

[9] Anyanwu, K., Maduko, A., and Sheth, A. SemRank: Ranking Complex Relationship Search Results on the Semantic Web. In *Proceedings of the 14th International World Wide Web Conference (WWW2005)*(Chiba Japan, May 2005). 117-127.

[10] Anyanwu, K., Maduko, A., and Sheth, A. SPARQ2L: Towards Support for Subgraph Extraction Queries in RDF Databases, in *Proceedings of the 16th International World Wide Web Conference (WWW2007)* (Banff, Canada, 2007) 117-127.

[11] Sheth, A., and Ramakrishnan, C. Realizing the Relationships Web: Blazing Semantic Trails Between Resources, *IEEE Internet Computing* (2007) (in press).