Wikis, Semantics, and Collaboration
Symposium on Collaboration Analysis and Reasoning Systems, at the 2014 Conference on Collaboration Technologies and Systems

Mark Greaves
National Security Directorate
Pacific Northwest National Laboratory
Seattle, WA USA
mark.greaves@pnnl.gov

EXTENDED ABSTRACT

Designing software for collaborative sensemaking environments begins with a set of very challenging requirements. At a high level, the software needs to be flexible enough to support multiple lines of inquiry, contradictory hypotheses, and collaborative tasking by multiple analysts. It should also include support for managing evolving human-machine workflows and analytic products at various levels of strictness and formality, processing partial and ambiguous evidence arriving in streams, and developing explanatory scenarios based on both serendipitous and structured discovery. Eventually, it should support the analytic team as they evaluate multiple alternatives and converge on one or more consensus responses, while preserving the history and underlying reasoning. Finally, it should be delightful and simple to use, not require an inordinate degree of precision and exactness, and be quickly and inexpensively deployable in a variety of rapid-response analytic situations. It has not been possible thus far to create a single software architecture that adequately balances all these goals. However, we can shed useful light on this problem by looking at the experience of semantic wiki architectures: an emerging class of software that blends wikis, databases, social tagging systems, and Semantic Web representations.

A. Knowledge Representation on the Web

Our starting point for approaching the design of collaborative analytic support systems is the software and standards for formal knowledge publication on the Web. By formal knowledge, we mean information that is explicitly represented in Web documents using something other than natural language text – typically, as machine-readable Semantic Web expressions based on a formal syntax and a specific, intended semantics [1]. These systems are a counterpoint to traditional artificial intelligence (AI) based accounts of knowledge. Most symbolic AI systems are designed to support sophisticated logical inference over fully consistent (and often closed-world) knowledge, and the underlying research challenges are focused on characterizing formal properties such as the space of entailment relations, time-space complexity of inference, monotonicity, and expressiveness. In contrast, the Semantic Web places formal knowledge in a very different context, where partial and often contradictory data representations exist in a constantly changing, large-scale, highly distributed network of loosely-connected publishers and consumers, and are governed by a Web-derived set of social practices for publication, discovery, and trust. The underlying research questions for Semantic Web knowledge include difficulty of reuse, ontology mapping, consistency and paraconsistency, and versioning. Unlike classical AI knowledge bases, Semantic Web data are constantly evolving through a lifecycle including publication, use, integration with other knowledge, modification, and eventual deprecation.

While the Semantic Web includes microdata, Open Graph data, and several other markup formats, the core of the Semantic Web is rooted in a set web-friendly knowledge representation languages – principally dialects of the Resource Description Framework (RDF), the Web Ontology Language (OWL), and the SPARQL query language that are compatible with standard Web server and browser infrastructure [2]. These standards were designed to be logically open-world and consistent with a collaborative publication ethos. The Semantic Web has grown to encompasses more than 50 billion Semantic Web assertions (triples) shared across the world via large numbers of autonomous Web servers, processed by situation-specific combinations of local and remote logic engines, and consumed by a shifting collection of software and users.

B. Web Data and Semantic Wikis

Most data currently published on the Semantic Web are originally sourced from existing relational databases, either via front-end systems like D2R [3], or by offline loading of the relational data into an associated triplestore. However, because the Semantic Web exists within a framework of other Web practices, it supports additional techniques for data acquisition, structuring, and logical characterization—namely, those that derive from human collaborative behavior on the Web. In particular, work on the Semantic Web has spawned semantic wikis, a class of extended wiki that was originally developed as a way to interact with and publish Semantic Web data, and combines Semantic Web technology with wiki-based social mechanisms. The most mature example is the open-source...
Semantic MediaWiki (SMW), which is built as a set of extensions to the popular open-source MediaWiki software framework [4]. Semantic wikis combine the structural discipline and query power of a database with the crowdsourcing power of an ordinary text wiki. In particular, they take advantage of the consensus-promoting property that many encyclopedia-style text wikis exhibit: the revision stream of the wiki articles with a sufficiently large and diverse number of authors will tend towards a (temporally individuated) set of fixpoints that Wikipedia calls the neutral point of view (NPV). The NPV represents the effective consensus of all the wiki authors at a particular time. MediaWiki-based wikis like Wikipedia contain several software features that encourage convergence to the NPV in article text, including join editing and rollback, separate discussion spaces, traceability features, and notification.

SMW inherits the above capabilities from its MediaWiki ancestry, but it also includes a number of specialized software modules that extend the NPV consensus property to user-entered data types and values. In this way, SMW merges formal knowledge authorship (at roughly the RDF(S) level of expressivity) with the familiar edit, discussion, and feedback loops that are characteristic of text wikis. Additionally, SMW includes several modules that support capabilities found in modern data management environments, such as forms-based data entry, faceted search, query, data visualization, triplestore-based inference support, role-based security, and Linked Data mapping and integration tools [5].

C. Implications for Collaborative Analysis and Reasoning

What lessons can we draw for collaborative analysis and reasoning software? Ordinary MediaWiki-style wikis seem at first to be a promising class of approaches to this kind of analytic software: they are demonstrably quick to evolve their structure in response to new user threads, have a simple and proven UI, and would support a collaborative analytic convergence to a recommendation (the equivalent of the NPV). However, simple wikis have not been generally successful in analytic contexts. The reasons for this are not fully known, but include insufficient incentives to vigorously contribute as well as the following technical issues:

- **Difficulty of navigation:** Wikis generally support keyword search or explicit hyperlink traversal only, with search results limited to a set of linked pages under a freeform category system.
- **Unstructured:** No workflow system or support for structured tasks, and only a basic security system.
- **Stand alone:** It is cumbersome to integrate external data feeds or background reference data.
- **Article-oriented:** Significant technical skill needed to create entities that are not text-rooted, such as query-generated charts or dynamic maps.

Semantic wikis directly address many of these issues by blending a wiki with a database, and adopting as a data representation the flexible and open-world graph data models that underlie the Semantic Web.

Most importantly, though, semantic wikis give users the ability to collaborate and reach NPV-style consensus about structured entities. Beyond simple RDF(S) data and type assertions, SMW extends the class of wiki-supported basic types to include entities like workflow tags, data mappings, fragments of argument structure, taxonomies and controlled vocabularies, and data queries and visualizations. Finally, because these wiki entities can be translated into RDF(S) and OWL expressions, semantic wikis can employ off-the-shelf description logic reasoners to provide automatic support for structure and consequence reasoning in the wiki.

These properties give semantic wikis three major advantages over standard wikis and databases for collaborative analysis tasks. First, in contrast with the ‘schema-first’ designs of traditional relational databases, semantic wikis are “schema-last.” This means that, like the Semantic Web itself, the logical structure of data in SMW emerges from the actual patterns of properties and values that exist in the RDF(S) data in the wiki. This gives semantic wikis great flexibility with respect to embedded data, tasking, and user expectations. Second, because all data elements in the wiki are represented on pages, the explanatory metadata that goes with the data elements can be documented in ordinary text that is visually linked to the target data, promoting human readability, understandability, and comfort with the data. Third, the ability to embed SPARQL-based query mechanisms into wiki articles allows users to continuously assess wiki data scope and integrity, find conflicts and missing values, and locate areas where the emergent schema of the wiki data is not consistent or needs revision or refactoring.

D. Examples

Semantic wikis have started to be deployed in domains that share many of the requirements of analytic sensemaking environments. Two relevant examples are:

- **SCRUM wiki.** SCRUM is a software development team methodology that is based on agility and rapid response to changing customer requirements. It combines characteristic timeboxed workflows (called sprints) with lightly structured collaboration around software designs, functional tests, and feature definition. SCRUM wiki supports SCRUM developer teams with a management and collaboration platform that includes SCRUM metrics and entities, differential privileges according to SCRUM role, a flexible tasking system, and quick customization to the individual project requirements by the users themselves.

- **Wikidata.** Wikidata [6] is a new project of the Wikimedia Foundation, aimed at providing a new set of powerful methods for the Wikipedia community to separately collaborate on and manage the symbolic data currently locked in millions of Wikipedia articles and infoboxes. Wikidata is built on a custom semantic wiki which allows Wikipedia contributors to also contribute symbolic almanac-style data. Wikidata’s design allows a distributed set of Wikipedia authors to collaboratively curate semistructured data, manage contradictory assertions, and create dynamic visualizations directly in
Wikipedia. Wikidata has been a successful project – it currently contains around 30 million statements and has over 1 million human contributions per month [7].

Systems like SCRUM wiki and Wikidata show how the flexibility and collaborative convergence properties provided by familiar text-based wikis can be combined with structured data management based on Semantic Web knowledge representation techniques. Together, they illustrate a promising pattern for designing flexible user-driven software for collaborative analysis and reasoning in dynamic information environments.

REFERENCES


