Content-sensitive User Interfaces for Annotated Web Pages

Florian Schmedding
University of Freiburg, Germany
schmeddi@informatik.uni-freiburg.de

This is a resubmission of work published at the AST Workshop 2009 [Schmedding, 2009].

1 Introduction

A broad range of different semantic technologies have been developed and standardized over the last years, but browsing on web sites rarely takes advantage of them although semantic descriptions can be supplied with RDF. We see one reason for it in the missing linkage between this data and the page content being presented to the user. Consequently, the underlying data cannot be addressed via an interaction with the visual representation. Despite having already found what one was looking for in the document, selecting the according data in a separate location is necessary for further automatic processing.

Recently, a new formalism—RDFa [Adida and Birbeck, 2008]—that seeks to close this gap has been standardized by the W3C. It defines some additional attributes for XHTML and a mapping from the attribute values and the document content into RDF triples. An important feature in our context is the usage of selected text parts for literal properties and hence the explicit linkage between the human- and the machine-readable representations. In contrast to the mapping, the handling of this linkage is left open in the reference specification.

2 Contribution

In this work, we will introduce a novel way of handling this linkage within the document object model and show its utility for several new use cases for RDFa documents. We propose an integration of the RDF statements into the document object model of the XHTML pages as a generic way to manage RDFa annotations. Covering resemblances between the use cases it makes their handling easier.

In particular, we focus on applications and user interfaces that relate the annotations on web pages to knowledge from other sources. This means they operate in two directions: extracted data from the page can be forwarded and response data can be integrated right into the displayed page in the browser, close to the respective annotated elements. Of course, this approach is not limited to web browsers but applies to HTML documents in general. For instance, it could be used also for HTML e-mails.

3 Extension of the Document Object Model

With our integration of RDFa and DOM we are targeting the visual representation of web documents. As the interaction with web pages regards the DOM representation, the extension of this model with functions to manage RDF statements seems a natural derivative. Our approach makes it possible to retrieve DOM elements by SPARQL queries; DOM elements in turn provide information about the contained statements and resources. Therefore we add properties to store subject, predicates, objects, datatype, and type information of RDFa to each XML element; for the statements we include a triple store. Similar to existing DOM-methods, we provide methods to retrieve XML elements that contain a given resource or statement. More specific conditions can be expressed by SPARQL queries. Instead of variable bindings, the respective method returns the XML elements where the resulting resources are defined. The necessary linkage information is contained directly in the data structures of the used triple store.

4 Applications

In our use cases, we use RDFa not only in one but in two directions: we let the web browser extract data and receive further data from other sources to provide advanced feedback for users. Knowing the linkage between text and meaning is especially necessary in the second case.

For example, appointments on a web page can be compared to entries in the user’s personal calendar. To notify the user about overlaps, the passages describing conflicting dates are highlighted. Thus one only perceives feedback for content one is actually reading. Additional actions can be provided via context-menus on the respective elements. Similarly, input suggestions can be supplied, e.g. when entering a surname into an annotated form of an online telephone directory: the user’s address book can be searched for matching forenames, expecting better results.

References
