RDFa versus Microformats: Exploring the Potential for Semantic Interoperability of Mash-up Personal Learning Environments

Vladimir Tomberg, Mart Laanpere

Tallinn University, Narva mnt. 25, 10120 Tallinn, Estonia
vtomberg@tlu.ee, mart.laanpere@tlu.ee

Abstract. This paper addresses the possibilities for increasing semantic interoperability of mash-up learning environments through the use of automatically processed metadata associated with both learning resources and learning process. We analyze and compare potential of two competing technologies for this purpose: microformats and RDFa.

1 Introduction

Mash-up Personal Learning Environments have become a fast developing trend in the world of technology-enhanced learning, partly because of their flexibility and lightweight integration features. Although it is quite easy to aggregate the RSS feeds from the blogs of learners, it is more difficult to get an overview of course and its learning activities. A course is not just a syllabus, it also involves various dynamic processes that can be described in many aspects. The course always has certain learning goals, a schedule that consists learning activities (assignments, discussions), registered participants like teachers and students, and different types of resources. It would be useful, if we would be able to extract such information also from mash-up personal learning environments (just like it can be done in traditional Learning Management Systems) and allow exchanging it between the course participants.
Today for semantic tagging of Web content in general and learning content as special case various technologies are used. But there are no tools and ways exist for semantic annotation of learning process that takes place in a distributed network of mash-up personal learning environments. The main aim of this paper is to compare possible ways for embedding the rich metadata about the course into mash-up personal learning environments.

The main format of presenting information in mash-up PLEs is an (X)HTML document; so any necessary metadata should be embedded into it. Because (X)HTML syntaxes were not designed for carrying the semantic data, different enhancement technologies were introduced in the past. We try to examine two of them that seem for us as much appropriate for the aims mentioned above — microformats and RDFa. The reuslts of such inquiry could be useful in first turn for designers of mash-up learning environments, because there are no common established issues that can be used for semantic annotation of learning process.

2 Scenario

Let us examine a typical usage scenario of using mash-up personal learning environments, involving exchange of semantic metadata associated with learning-related content and also the process of learning. As a starting point for initiating a new course, the teacher publishes an announcement using a Web application (blog, wiki, forum or personal Web site). This announcement contains (meta)data about the course syllabus, pre-requisite and target competencies, amount of credits, dates for start and end of the course, the criteria and form of a final assessment, contact information of teachers and other participants. Web servers by means of mash-ups will bring this information to potentially interested learners. It arrives to them through RSS channels, Web services and/or an e-mail subscription. After a learner enrolls him- or herself to a course, information about him/her is added to a course database.

The teacher attaches semantic significance to learning materials placed in a network. The learners use semantic search for finding additional learning resources. Besides, by means of metadata additional organizational possibilities appear. All participants of a course, learning materials, activities of the teacher and learners can be connected among themselves by means of semantic relations that give the possibility to trace progress of learners in real time or for reception of reports. In this case the classical learning system can be implemented as the thin client which role to search and analyze semantic relations. Problems of indexing and storage of data and representation of information take upon oneself distributed servers. The role of learning system at such scenario is reduced to aggregation and the analysis of the distributed data only.

Using only the metadata which is intended for the learning materials description will not be sufficient for current scenario. The metadata by means of which is possible to describe a course is required for this purpose. In wide understanding the course it is not a static product, it is a process which after starting occurs throughout time. Also the course is the community of people that work together. The curriculum can be implemented as dozens copies of course, but each of these copies, besides a set of the identical information will have also the own unique information associated with various resources, including the human resources. Usual course's metadata contains values that possibly will not change, for example the course name, amount of credits or form of the assessment. At the same time the course expiration date, amount of the learners registered on a course, names of teachers and a venue will change every time as new instance of course will initiate.

Thus, using a special semantic markup for courses, the teacher has possibility to constantly update the course information during time. The teacher assigns lectures, announces an assessment and evaluates learners. Learners thus constantly have the fresh information on everything that happens on a course. After finishing of a course the information about it as about a logical unit also can be demanded for the analysis and the reporting. We will consider two technologies available today that allow implementing semantic description of educational process occurring in a Web environment.

3 Microformats

Microformats are defined as "a way of adding simple markup to human-readable data items such as events, contact details or locations, on Web pages, so that the

information in them can be extracted by software and indexed, searched for, saved, cross-referenced or combined' [1].

Microformats allow to use for metadata storing purpose the standard (X)HTML attributes 'class', 'rel' and in a case with microformat VoteLinks as well 'rev'. The Most important advantage of microformats consists in that that they use existing HTML syntax, thus not changing it at all. The HTML text with correctly used microformats always will pass a validation. Most of the Web browsers simply ignore the presence of microformats in the HTML code. Recently some developers of browsers already had announced further microformats compatibility, though at the current moment visualization of microformats is possible only by means of additional add-ons or coding of special page behavior, e.g. by means of JavaScript.

Microformats are categorised into two groups: elementary and compound microformats. Elementary microformats like *geo* or *rel-license*, are intended for the describing of small, individual things and concepts. They can also be used as building blocks for compound microformats. Compound microformats have complex nested structures. Examples of compound microformats are hCalendar, hCard that precisely reflect data of standardized iCalendar and vCard.

The basic advantages of microformats are their simplicity of implementation, a consistency with existing standards and broad support of developers. Possibility to place in the text additional information which without unnecessary expenses can be transferred to an address book, a calendar or an electronic map attracts users. These possibilities undoubtedly add accessibility to data and make the web friendlier to the user. Support of microformats through special extensions is implemented today almost in all popular web browsers. In comparison with other solutions microformats seriously facilitate a life to developers. They should not implement specific behavior for Web clients again if someone already implemented it. Microformats are not standardized by any standardization agency, but they are well specified and widely known. If one developer writes the program for processing of microformats it will process such microformats from any sources.

The vocabulary of microformats is constantly updated. Unlike well-known standards microformats have no versions, they are being developed continuously. People responsible for uniformity of microformats try to maintain the vocabulary of microformats in as much as possible compact condition. In most cases they advise instead of introduction of a new microformat to try to use already existing microformat. At the same time in many cases microformats' vocabulary is not sufficient for the description of all necessary data. In such case content creators appear in front of a deadlock.

Biggest drawback of microformats is that they have no ontologies, formal descriptions or schemes. The vocabulary can be checked up manually under specifications, but there is no way of automatic check. The existing values of 'rel' and 'class' attributes can to be not coinciding with the vocabulary of microformats, but thus HTML code remains completely valid [2]. In HTML code any value for these attributes can be used, but not in microformats' case.

4 RDFa

RDFa it is the syntax standardized by W3C. RDFa is an acronym from RDF in Attributes; it uses the mechanism of implementation of metadata similar to microformats in XHTML attributes. Instead of three attributes of microformats RDFa

uses ten; sometimes at first sight it is difficult to distinguish code of RDFa from microformats one. For example:

```
<a rel="cc:license" href="http://creativecommons.org/ licenses/by-nc-nd/3.0/">  Creative Commons License
</a>
```

Difference of RDFa in the given example is syntax of 'rel' attribute value. A 'cc:' prefix specifies namespace, in this case Creative Commons. The prefix allows to RDFa to describe such concepts which by means of microformats it is hard to describe, especially because of the limited vocabulary of microformats.

The fundamental advantage of RDFa that with its help it is possible to use different namespaces in one document, for example 'dc:' as Dublin Core and 'cc:' as Creative Commons simultaneously. RDFa allows combining, remixing, and extending existing vocabularies easily [3].

Using of RDFa assumes for developers also some considerable restrictions. It is possible to apply RDFa only to XHTML version 2, not to HTML. XHTML version 1.1 also can be used with RDFa in practice but with considerable limitations [4]. Prospects of use RDFa together with HTML (not XHTML) do not looks as promising at present. On the one hand attempt to introduce RDFa code in HTML will lead to situation, where Web browsers can process such page without any problems. However the validation of such HTML+RDFa page will be impossible, as in HTML syntax is absent necessary for RDFa attributes. In addition RDFa code can be casually damaged at attempt to pass HTML through Tidy or other code cleaning tool. Such tools will simply break embedded RDFa semantics [5]. In consideration of mentioned above using RDFa in HTML should have no trust from developers.

Support of RDFa syntax likely is not planned to introduce in following version HTML 5 also [4], though advocates of a RDFa do not lose till now hope that WHATWG will pay the attention to problems associated with semantics [6].

Other barrier for fast RDFa introduction is using in it a new syntax for Uniform Resource Identifier (URI). Instead of traditional URI a new CURIE syntax is used in RDFa. CURIE simplifies work with links on URI and shortens a code. CURIE is W3C Candidate Recommendation [7]; it is intended for use in SPARQL, RDFa and XHTML 2. However, CURIE technology is still in its early phase of development.

5 Microdata — a new player

At the moment of writing this paper (July 2009) important changes in situation around standards have happened. The W3C announced that ‘XHTML 2 Working Group expected to stop work end of 2009, W3C to increase resources on HTML 5’ [http://www.w3.org/News/2009#item119]. Obviously reason for this decision was changes in relationship between W3C and WHATWG as browser’s business representative. W3C has changed its attitude to HTML 5 from full rejection to unstinting support.

Dozens discussions in the W3C Mailing lists and blogs related to standards have lead to much unexpected result. Ian Hickson, a leader of HTML 5 working group finally has agreed to arguments of his opponents and has proposed a Microdata — completely new semantic syntax for HTML 5 [8]. By its nature Microdata is a certain kind of hybrid made from microformats, RDFa and HTML 5 syntaxes. It is based at three new HTML 5 attributes: "item", "itemprop" and "subject"; at their base RDF triples can be generated.

At the moment of writing of this paper a discussion about strengths and weaknesses of Microdata is continuing. Advocates of Microdata, for their part, are proposing analysis of use cases that describe advantages of Microdata compared to RDFa [9]. Advocates of RDFa in turn publish their own RDFa at the HTML 5 specification's proposal [10] and express optimism about future perspectives of RDFa.

Situation with RDFa is remaining to be very unclear. As W3C says, work on XHTML 2.0 and CURIE will stop, although both of these unimplemented standards very close related to RDFa. In the same time W3C says also: 'Whether and how to include RDFa into HTML 5 is an open question on which we expect further discussion from the community' [11]. Uncertainty of RDFa's perspectives and appearance of new player Microdata only complicated tasks for developers and will slow down widespread use of semantic tools on the Web platform.

Because Microdata for today is not established and well defined yet, we do not consider it in this paper.

6 Microformats vs. RDFa on the semantic level

After outlining tech differences of Microformats and RDFa we try to define a differences between two technologies in semantic aspects. Both of them have own historical background, both are driven by forces from very different sources, and especially important, these technologies have very different perspectives.

From the point of view of *data description* the principal difference between microformats and RDFa consists that microformats can use only one vocabulary. This vocabulary is being approved by community and cannot be modified by the separate developers. Actually it is possible to concede such independent modification in practice, but it should be considered as violation of microformats' principles, so we not examine such case. RDFa vocabularies on the contrary are not limited by a specific data set [12]; they are developed separately in different places. In connection with the aforesaid it becomes obvious that implementation of semantic data in big scales is easier with using of microformats. The probability that software will easy manage, define and process any syntax of microformats is much higher than probability of recognition of RDFa, just because this vocabulary is small, well known, and is just one.

At present there are some implementations of add-ons for Firefox and Internet Explorer browsers exist which are able to recognize both formats. But only vocabulary of microformats is well known for everyone who interested to work with them.

Both technologies can be considered as from side of supporting of end users, as from side of machine processing for indexing and mashups needs. The developers already have many possibilities to implement practical functions for end users using microformats today. For example, if Operator or Tails Export add-ons for Firefox have a detection of hCard microformat's presence, they suggest to the user an export of web-content to a file with *.vcf* extension or a method for adding the information to personal address book. Detection of hCalendar microformat calls a suggestion from browser to export the content to a file with *.ics* extension or place data from web page directly to the calendar software of the user. Tails Export add-on if meet a web-page with Rel-licence microformat can show to the user the specific information about this license, or in a case if there is microformat Geo presented, the software suggests an

allocating of the address on the Google Maps. All above listed examples are finished and working implementations of possibilities of microformats use for end users.

With RDFa the situation differs radically. Some add-ons for Firefox declaring ability to work with RDFa, but today they are able only to find out RDFa code on the web-page. A maximum that a browser with RDFa code can make, it is a retrieving RDF triplets from source code and sending the user on a special page with the RDF triplets' interpreter. It is obvious, that such services can bring some advantage for developers, but have no sense for end users, and actually there is nothing especially useful to the end user in RDFa.

The implementation of machine processing on a global scale is much easier too with the vocabulary of microformats than RDFa's because vocabulary of microformats is uniform and developed centrally. The software from different vendors should be able to process microformats of one type equally. The set of microformats' types is limited by community therefore it is possible to state high enough level of interoperability between the different software intended for processing of microformats.

At the same time uniform vocabulary of microformats has also drawbacks. Developers continually face a deficiency of the vocabulary, with absence of new microformats which are necessary to cover continually arising new requirements. Other possible problem consists in a non-typical way of microformats' development, because unlike traditional standards, today it does not assume any logical completion. Specifications of existing microformats, their syntax and the processing rules are not described in schemas, they are only can be searched manually through a web [13]. Such approach obviously does not promote fast updating of the software after appearance of new microformats. The majority of developers simply learns nothing about new developments if does not monitor a web site microformats.org on one's own.

7 Educational application

Above mentioned semantic characteristics of technologies under comparison set quite different limits to their educational use. We try to examine these limits closer now.

The vocabulary of microformats is laconic and almost perfect. We can solve many tasks by means of the microformats' vocabulary if it is necessary to describe activities concerning a course. Developers of microformats suggest using actively a nesting of microformats, and by means of that it is possible a semantic marking of rich data structures.

Let's consider some examples of possible using of standard vocabulary of microformats for the course description. By referring to use case described above, lets examine the situation, where the teacher announces a course. We can use for the description of this course a microformat hCard. This microformat allows naming a course using a class "organization-name". Also with the help of hCard we can specify geographic location of the course. With the help of hCalendar microformat we can announce lessons, specify their duration and a venue. The learner can then automatically import a schedule of lesson to their personal calendars.

Similarly we will supply with hCard microformats all data about the teacher and learners who take part in a course. By means of the address information learners can find each other and communicate among themselves and with the teacher online. We can also define a difference between them by means of a 'role' class.

Shared knowledge building assumes the organization of workgroups from learners. Here we at the first time face shortage of microformats' vocabulary. If a necessity to specify the relation between participants of a course (teacher / learner) or participants of working group (senior of group, rank-and-file member of group) it would be logical to use XFN microformat which is created to show relations between people. However the best possible solution (which is thus far not ideal) that can be found in XFN are values of 'rel' attribute "co-resident" and "colleague". It is certainly not the best method to show the exact relationships in groups. Values of attribute "parent" and "child" for such situation not so good suits also, as in a context of course have ambiguous sounding.

This example well illustrates problems of implementation of specific ontologies by means of microformats. On a web-site microformats.org there are special pages on which the community discusses new microformats for education. There even the prototype of the future microformat hCourse is mentioned. Unfortunately, planned functionality of it would more suit for the description of the curriculum than for description of the course process. In existing attempts to implement the description of the course, the course often is considered as the set of certain curriculum data, learning materials and other accompanying information. For example such data can be rooms where the course will be spent, teachers' names, final examination's date and so forth. The problem consists that the information on a course is considered first of all as the marketing tool by means of which is possible to inform a buyer about university products. However such information will be almost useless after the course will begin. We wish to notice that course is not only static set of the information. It is also selection of activities into that many people and resources are involved.

Table 1. Summative comparison of Microformats and RDFa. Better features marked with gray background.

	Microformats	**RDFa**
Technological properties		
Can be applied to	HTML, XHTML	In current state for XHTML 2 and limited for XHTML 1
Have useful implementations for end user today	Yes, different add-ons for browsers allow that	Not very useful for end user directly
Can be used in mash-ups	Yes	Yes
Practice using for semantic indexing examples	Google and Yahoo now indexing such microformats as hCard and hReview	Google, US Government Website and Slideshare use RDFa
Standardized	No	Yes
Drawbacks	It is very difficult to use one limited dictionary for all purposes. Because of absence of standards support from developers is more enthusiastic than industrial	Because of finishing of XHTML2 development the perspectives of RDFa are very unclear at the current moment
Semantic properties		
Vocabulary	Only one	Unlimited amount is possible
Vocabulary can be extended	Yes, but only through community	Yes, freely
Interoperability level	High	Possible only if common vocabularies are used
Possibilities to add semantic value to data	Average	High
Application for educational needs		
Possible semantic data use	Course's membership via hCard Learner and Teacher information via hCard Members relations via XFN Course events via hCalendar	Can use any semantic vocabulary that suitable for course description.

At RDFa the situation with vocabularies is much better. Possibility simultaneously to use at once such popular vocabularies as Creative Commons, Dublin Core, XCRi and other removes both problems related to description of a course resources and to description of activities. In case of absence necessary metadata vocabulary it is

possible to create any schemes by oneself. However it is necessary to mean, that interoperability of system will decrease proportionally to appearance of new vocabularies, because any software cannot be ready to right process all possible vocabularies in the world.
A general outline of microformats and RDFa comparing results is shown in table 1.

5 Conclusion

So which technology is more suitable if we need to exchange easily the rich metadata about the course and its learning activities between several mash-up personal learning environments? The answer is depending on our choice between the universality of language and simplicity of adaptation of the selected annotation technology. If developers of microformats will start paying more attention to educational possibilities of microformats (and they probably will), it is possible in the near future we will see new microformats that are powerful enough and simple for adaptation. If we decide to implement the RDFa approach, there is still a lot of work to do. We will harvest on much smaller field, than in a case with microformats. The output however will be much more flexible and semantically rich than with microformats.

6 Acknowledgements

This study was funded by ESF grant 7663 and MER targeted research 0130159s08.

References

1. Introduction to Microformats. Microformats Wiki, Retrieved April, 2009, from http://microformats.org/wiki/introduction
2. McLellan, D.: Can Microformats be Validated? — All in the head. Retrieved April, 2009, http://allinthehead.com/retro/302/can-microformats-be-validated
3. Adida, B., & Hausenblas, M. RDFa Use Cases: Scenarios for Embedding RDF in HTML, W3C, Retrieved April, 2009, http://www.w3.org/TR/xhtml-rdfa-scenarios
4. Birbeck, M.: XForms and Internet Applications: Using RDFa in XHTML 1. Retrieved March, 2009, http://internet-apps.blogspot.com/2007/02/using-rdfa-in-xhtml-1.html
5. Graf, A.: RDFa vs. Microformats. Digital Enterprise Research Institute, Innsbruck (2007)
6. @rel attribute discussion - RDFaWiki. Retrieved April, 2009, from http://rdfa.info/wiki/@rel_attribute_discussion.
7. W3C. CURIE Syntax 1.0. (M. Birbeck, & S. McCarron, Eds.) Retrieved March, 2009, W3C: http://www.w3.org/TR/curie
8. HTML 5, Draft Standard. Retrieved July, 2009, http://www.whatwg.org/specs/web-apps/current-work/multipage/microdata.html#microdata
9. Hickson I.: Annotating structured data that HTML has no semantics for. lists.whatwg.org Mailing Lists. Retrieved July, 2009, http://lists.whatwg.org/htdig.cgi/whatwg-whatwg.org/2009-May/019681.html
10. Sporny, M. First Editors Draft of HTML5+RDFa Published. Blog @ Digital Bazaar. Retrieved July, 2009, http://blog.digitalbazaar.com/2009/07/13/html5rdfa/2/
11. Frequently Asked Questions (FAQ) about the future of XHTML. Retrieved July 14, 2009, from http://www.w3.org/2009/06/xhtml-faq.html
12. Simpson, J.: Microformats vs. RDF: How Microformats relate to the Semantic Web. Retrieved April, 2009, http://www.semanticfocus.com/blog/entry/title/microformats-vs-rdf-how-microformats-relate-to-the-semantic-web
13. Ogbuji, U.: Microformats in Context. Retrieved April, 2009, http://www.xml.com/pub/a/2006/04/26/microformats-grddl-rdfa-nvdl.html