

Microformats

A XML-document-immanent way for semantic annotations of web contents*

Matthieu-P. Schapranow

Hasso-Plattner-Institute for IT Systems Engineering
at the University of Potsdam,
Prof.-Dr.-Helmert-Str. 2-3,
D-14482 Potsdam, Germany,
matthieu.schapranow@hpi.uni-potsdam.de,
<http://www.hpi.uni-potsdam.de/>

Abstract. The nowadays World Wide Web is characterized by a lot of human-readable, but unstructured web content. The growing amount of published information makes it more and more harder to find desired information. With the idea of Semantic Web the need for qualified, structured semantic annotations of web contents arises. Although the Resource Description Framework for semantic annotations is available for years, its complexity inhibits its wide usage. This paper deals with a new user-driven approach for minimalist semantic annotations of web contents: the Microformats. The concepts of Microformats, their origins and their intention are outlined and compared with existing approaches for semantic annotations. A selection of Microformats implementers is mentioned and the benefits for their users are extracted and compared.

Key words: Microformats, Resource Description Framework, Semantic Web, XHTML

* A paper associated with the **Process & Knowledge Engineering Master Course** in Summer Term 2006 by Prof. Dr. Mathias Weske and M. Sc. Hagen Overdick, July 11th, 2006

1 The need for Semantic Annotations of Web Contents

In the given paper a user-driven development in semantic annotation of web contents is presented: the Microformats. After describing the Resource Description Framework as an already existing approach for semantic annotations the new concepts of Microformats, their origins and their intention are outlined and compared. A selection of Microformats implementers and the benefits for their users are extracted. Ultimately, the Microformats' impact for the development of Tim Berners-Lee's dream of the Semantic Web is addressed and expectations for future developments are drawn.

1.1 World Wide Web in a Nutshell

For years the World Wide Web (WWW) was only used in a one-way: scientists published their own work for other scientists, i.e. the originally intend of the WWW. During the 1990s more and more user discovered the WWW because of lower prices for necessary technical equipment and phone lines rates. Especially companies used the WWW to establish a more or less static market place. Information were produced and placed on their websites, but the wide area of users only acted as information consumers in the given scenario.

Within the last few years that development has been changed dramatically and a wide range of former information consumer become information provider. Especially the so-called blogging becomes famous, which offers everyone the possibility to publish any personal influenced information on the WWW. On the one hand, one side effect of this development is the establishment of an enormous large community of regular and non-regular information providers, the so-called bloggers. On the other hand, a lot of information consumers get involved in new topics and are enabled to give their comments on existing posts in the WWW.

Thus, an enormous information mesh-up is created by giving extensive feedback to a wide-range of published information. One way to create mesh connections is to use the so-called Trackback mechanism, which explicitly links a published article in response to another already published article on basis of hyperlinks – a real web emerges.

In March 2006, more than one billion people of the whole world population are marked as active users of the Internet as described in [1]. The persistent increase of Internet usage also results in a considerable increase of published information. In combination with this development the desire for finding the appropriate information becomes more and more essential.

A way of finding information in the WWW is to use search engines such as Google, Yahoo, MSN, etc. Those providers often use crawling robots, i.e. a tiny, multiple instantiated program that visit registered websites in rotation. The contents of the crawled websites are indexed in tremendous databases.

With the increase of published information and the rising number of websites the necessary resources for crawling websites grows. The traditional search are not able keep pace with the shortened lifecycle of published information. Information is published multiple times a day exactly when a related event occurs.

This latency often results in outdated information returned by search engines. To overcome this problem new ways of information propagation, such as RSS- and ATOM-feeds, have been developed and are established nowadays.

1.2 The Semantic Web

Currently, the WWW is used as an interactive omnipresent media for information transfer announces a next step in the evolution of this technical media. For the last few years a new idea of using the WWW has been born: the Semantic Web.

The idea of the Semantic Web was described by WWW-Consortium member Tim Berners-Lee et al. in 2001: "The Semantic Web is [...] an extension of the current one [WWW], in which information is given well-defined meaning, better enabling computers and people to work in cooperation." [2]

They underline the aspect of information and their "well-defined meaning" as a main target of the Semantic Web. This means, that information and their meaning are not clearly connected in their current representation in the WWW, i.e. computers do not *understand* the human-generated meaningful content. As a primary aim of the Semantic Web Berners-Lee et al. stress the interaction of human-beings and computers and vice-versa as well as the semi-automatic or even fully automatic interpretability of web contents by machines.

Based on the ideas of Berners-Lee et al. for the Semantic Web the given paper compares the ready for use Resource Description Framework with the new user-driven development of Microformats to offer machine-interpretability of web contents.

2 Semantic Annotations of Web Contents

In this section the Microformats concepts are outlined after a brief introduction of the basic aspects of the Resource Description Framework as one already existing technique for semantic annotations of web contents. Additionally, a possible classification of Microformats is presented and the hCard Microformat is described in detail. On the basis of the hCard example a comparison of the Microformats and the existing annotation approaches is acquired.

2.1 The Resource Description Framework Model

The WWW-Consortium has been working for years on a standard for semantic annotations of web contents called Resource Description Framework (RDF) and published it in 1998 as recommendation. This standard is based on structured documents using namespaces and markup-languages. Within this section the RDF model and an example is described to communicate the principles of this framework for semantic annotations.

The RDF model can be integrated in a technology stack as shown in Fig. 1 on page 5. With the help of the eXtensible Markup Language (XML) any hierarchical structured set of information can be encoded on a textual base. This established data format is vendor-independent, human-readable and interpretable

if the application domain is known. The space of applicable formal constructs in XML documents is restricted by a XML scheme which defines a co-domain for the underlying XML structure. On top of this XML fundament the RDF is added, i.e. a specification for a metadata model based on triples. With the help of the triples *subject*, *predicate* and *object* it is possible explain objects and their relationships as further described in Tab. 1 on the next page. Analog to the XML scheme a RDF scheme is used to define a subset of acceptable language constructs for a given domain. The Web Ontology Language (OWL) [3] adds ontologies for machine-processing of human web documents. For instances, OWL is capable to describe meta properties of classes and object, such as equality, cardinality, restrictions, intersections, unions, and 1:n or n:n relationships.

Figure 2 on the facing page illustrates exemplarily an excerpt of a RDF file. Lines one and two are used to specify RDF content and introduce the XML schema for RDF and the RDF specific schema Dublin Core (DC) 1.1. In line three the `rdf:Description` enclosing the DC attributes and the described subject in the `rdf:about` tag is defined. The subject is a set of presentation slides taken from [4] for the presentation about Microformats which are described using the URL `http://www.myhpi.de/~schapran/pke/`. In line four and five the DC attributes `dc:title` and `dc:publisher` are assigned. Within the RDF model context both attributes describe objects related to the introduced subject. For instances, the person identified by the name *Matthieu-P. Schapranow* is described as the publisher of the aforementioned presentation slides, i.e. the relationship "*is publisher of*" describes how the subject (presentation slides) and the object (the named published) belong to each other.

In this simple example it is visible that the RDF approach is highly formalized which needs a considerable specification overhead. Although the file is human-readable it contains complex meta information, thus it is especially created for machine-interpretability analyzing the file automatically. Obviously, for human-readable the semantic content has to be prepared separately in an additional document containing hyperlinks to the presented RDF file.

2.2 The Microformats intention

The core developer team consisting of Tantek Çelik, Dan Cederholm, Rohit Khare, Ryan King, and Kevin Marks develops Microformats to be used by a wide-range of information publishers for annotating information semantically keeping the overhead as small as possible.

Microformats are "...designed for humans first and machines second, Microformats are a set of simple, open data formats built upon existing and widely adopted standards."

This citation taken from [5] is comparable to the citation of Berners-Lee et al. mentioned on page 3. The intention of the Microformats concept underlines the user-centric aspect rather than making information interpretable by machines. However, the common agree to use well-established standards for development is stated in both citations. So, the Microformats approach may help to get one step further on the road to the Semantic Web.

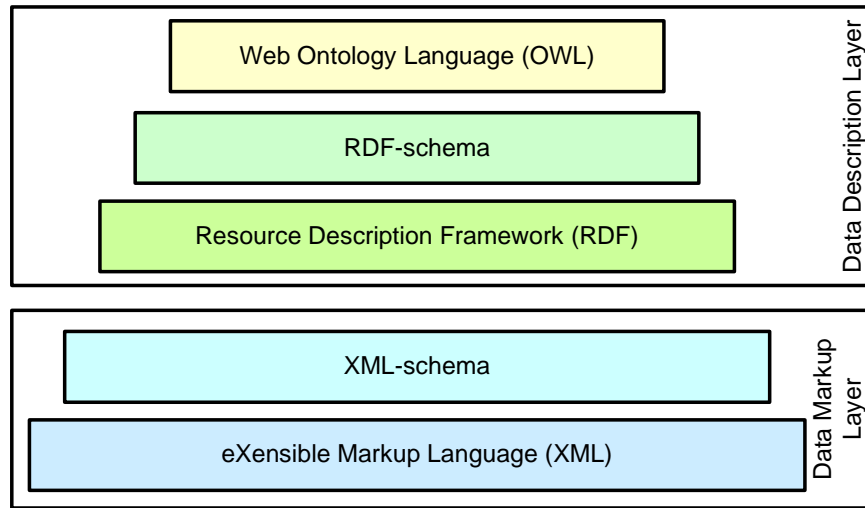


Fig. 1. Semantic stack – schematic view

```

1 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns"
2     xmlns:dc="http://purl.org/dc/elements/1.1/">
3 <rdf:Description rdf:about="http://www.myhpi.de/~schapran/pke/">
4 <dc:title>Microformats - Designed for humans first...</dc:title>
5 <dc:publisher>Matthieu-P. Schapranow</dc:publisher>
6 </rdf:Description>
7 </rdf:RDF>

```

Fig. 2. An example of Resource Description Framework file

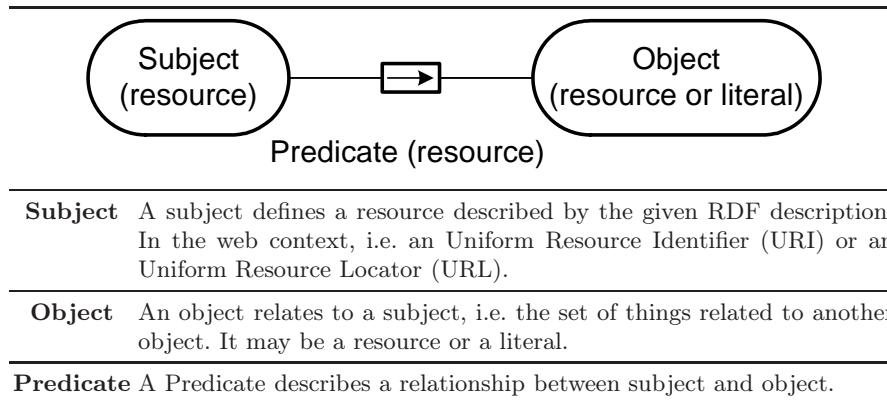


Table 1. Relationship of the triples in the RDF Model

2.3 Classification of Microformats

Table 2 on page 15 shows a list of currently available Microformats and a short description of their purpose. To categorize the set of existing Microformats two exemplarily criteria are presented below.

On the one hand, Microformats can be distinguished according to their divisibility, thus so-called *Elementary Microformats* and *Compound Microformats* as described in Fig. 3 are established. Elementary Microformats are atomic, i.e. they consist of one single tag or a not further refining functionality, e.g. link annotations such as the rel-tag in hyperlinks. This kind of Microformats forms at the moment a small subset of available Microformats. The larger subset contains compound Microformats which combine smaller functionalities to a more complex Microformat or data type.

On the other hand, it is possible to distinguish Microformats in *user data describing Microformats* and *data structuring Microformats* according to their solved problem. Different Microformats may be used to solve a similar aim, e.g. the Microformats hCard and hCalendar are used for description of personal user data. The Microformats eXtensible Open XHTML Outlines (XOXO) and eXtensible Open XHTML Tables (XOXT) are used for displaying data in a structured form, so they belong to the group of data structuring Microformats.

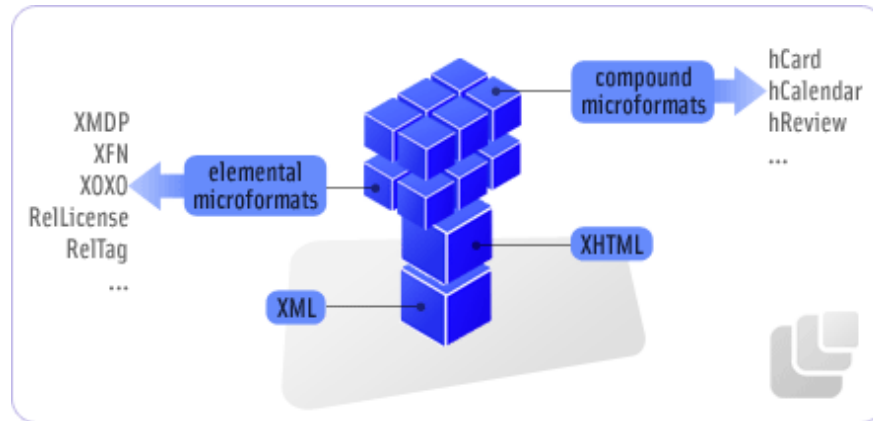


Fig. 3. Microformats – classification (taken from [5])

2.4 The hCard Format – A Microformat for Contact Information

In the following section the hCard Microformat for semantic annotation of contact information is compared to a RDF equivalent annotation. The name hCard is derived from the vCard data format, i.e. a standardized format for contact data exchange as described in [6].

Figure 4 on the next page shows an example of a vCard which are separate files containing attribute value pair separated by colons. It contains the attributes Name (N), Formatted Name (FN), Website URL (URL), Organization (ORG) and vCard version information. This format is human-readable, but it does not satisfy the common way of reading contact information. Furthermore, redundancy name information are stored for the attributes *Name* and *Formatted Name* differing only in the given word order. Additionally, vCard files are not embeddable in valid Extensible Hypertext Markup Language (XHTML) documents, i.e. they are referenced by hyperlinks on a webpage. The surrounding BEGIN and END attribute suggest the existence of a hierarchical structure which owns an isomorphism mapping all information lossless to a valid XML format.

With the help of Microformats it is possible to combine the semantic information of the vCard with its human-readable representation in a valid XHTML document. Figure 5 on the following page illustrates an hCard Microformat as a XHTML excerpt which represents exactly the same content as the aforementioned vCard in Fig. 4. To realize the semantic annotation the `class` attribute of the XHTML standard is used: `class="vcard"` in line one. The Microformat hCard uses this attribute for semantic annotation purposes without influencing the attributes initial task. The `class` attribute was intentionally introduced to separate the graphical representation from the data within (X)HTML documents by using cascading style sheet (CSS) metadata.

To understand the aforementioned isomorphism the following algorithm simplifies the transformation from vCard to hCard.

1. Enclose each vCard attribute value in `div` or `span` (X)HTML regions,
2. Add to the inserted (X)HTML regions' `class` attribute the appropriate vCard attribute name 1:1, e.g. FN.

For instances, in line two of Fig. 5 on the next page the initial vCard attributes Formatted Name (FN) and Website (URL) are combined by creating a hyperlink. The relationship is encoded using the (X)HTML `class` attribute `class="url fn"`. People browsing the resulting (X)HTML document containing the hCard Microformat are able to identify the hyperlink as the personal Website of the named person. Furthermore, it is easy to find another transformation from the given (X)HTML excerpt to vCard or even to RDF or any other compatible format and vice versa.

2.5 RDFa vs. Microformats

As described in section 2.1 on page 3 the high-formal approach for semantic annotation manifests in RDF. But it stores annotations in an additional XML document which is connected to the XHTML document containing the initial information via hyperlinks. To overcome this gap the RDFa format is switched to developing progress for future XHTML 2.0 standard. This should help to eliminate the additional maintenance to keep the RDF document up to date when the initial XHTML document changes. The key concept of RDFa is to

```

1 BEGIN:VCARD
2 VERSION:3.0
3 N:Schapranow, Matthieu-P.
4 FN:Matthieu-P. Schapranow
5 URL:http://myhpi.de/~schapran/pke/
6 ORG:Hasso-Plattner-Insitute for IT Systems Engineering (HPI)
7 END:VCARD

```

Fig. 4. The content of a vCard

```

1 <div class="vcard">
2 <a class="url fn" href="http://www.myhpi.de/~schapran/pke/">
3   Matthieu-P. Schapranow
4 </a>
5 (<span class="org">
6   Hasso-Plattner-Insitute for IT-Systems Engineering (HPI)
7 </span>)
8 </div>

```

Fig. 5. The content of a mapped hCard

embed native RDF structures into valid XHTML documents without any data redundancy for meta data descriptions. Thus, the RDFa approach is similar to the Microformats approach and both converge to the same aim.

Figure 6 on the next page shows an excerpt of an hCard Microformat and its equivalent using the RDFa format. Both examples contain an email address and its corresponding owner as a formatted name. The Microformat example uses the `class` attribute carrying a set of values containing the hCard values `email` and `fn` to realize the aim. A set of values is always unordered, i.e. for machines `class="fn email"` is identical to `class="email fn"`.

Human readers expect the hyperlink to contain the email address as reference and its enclosing value as the owners formatted name (FN). This assumption based on semantic understanding of the hyperlink concept in XHTML cannot be assumed for machines. Therefore, an extension of the existing RDF concept, the RDFa approach, has been developed to embed adapted RDF data directly into XHTML documents as described in [7]. In the RDFa example, the attribute-value-pair `rel="email"` describes the relationship between the subject specified by the resource – the abstract person – referred to by the `href` attribute – the email address. Mapped to the RDF model, this means: the abstract person is the subject, the "has got email address" is one relationship and the concrete email address `matthieu.schapranow@hpi.uni-potsdam.de` is the corresponding object. The `property="fn"` pair indicates which property is being defined by the current XHTML element. In the given example the property of the hyperlink element is the formatted name (FN) attribute of the vCard attributes.

The aforementioned example outlines the importance of independent semantic annotations, i.e. interpretability without knowledge of the data domain. Additionally, the large overlap of both examples shows that Microformats are a reasonable way of annotating web contents while keeping overhead small.

```

1 <a class="email fn"
2   href="mailto:matthieu.schapranow@hpi.uni-potsdam.de">
3   Matthieu-P. Schapranow</a>

1 <a rel="email" property="fn"
2   href="mailto:matthieu.schapranow@hpi.uni-potsdam.de">
3   Matthieu-P. Schapranow</a>

```

Fig. 6. Excerpt of an hCard (top) and its RDFa equivalent (bottom)

3 The Real World Usage of Microformats

Within the following section the real world usage of the new phenomenon Microformats is analyzed and corresponding examples are described. While surveying the real world usage it is necessary to keep in mind that the development of Microformats was initiated from the web blog area. Thus, a lot of the following examples are adoptions within this area of information publishing, but more and more other scenarios are also involved in this adoption process.

3.1 Semantic Browser

With the Semantic Web idea Tim Berners-Lee also introduced the concept of Semantic Web-Browsers. This adapted browser should bridge the gap between the personal computer system with its installed software and the abstract information within the WWW.

Today, Firefox is a famous, easy-extendable web-browser using pluggable extensions on basis of JavaScript functionality. The usage of JavaScript as a simple programming language and the plug-in principle offer the possibility to write small programs which processes web pages supplementary on client-side, e.g. to extract Microformat content.

For Firefox extensions the Tails and Tails Export extension are currently available. Tails Export created by Robert de Bruin is able to detect and display contents of the Microformats hCard, hCalendar, hReview, hAtom, xFolk, and rel-license in Version 0.3.1. Furthermore, it is possible to transform the detected Microformats via Extensible Stylesheet Language Transformation (XSLT) into other formats. It is capable to export hCard content to the standardized Virtual Card File (VCF [6]) format and hCalendar to the standardized Internet Calendar

and Scheduling (ICS [8]) format. Both formats are interpretable by a wide-range of already existing desktop application software such as Outlook or iCal.

Obviously, users of these extensions profit by seamless export functionality and reduced time needed to acquire contact or event information. When adding an interesting contact or event it is no longer necessary to type all data by hand.

3.2 The Live Clipboard

Ray Ozzie, Chief Technology Officer and Chief Software Architect at Microsoft, introduced on O'Reilly's Emerging Technology Conference 2006 the Live Clipboard. It is a Dynamic HTML control which allows to copy and paste prepared content between personal computer and websites and vice versa seamlessly. The integration idea helps to overcome the gap between web content and content on a local user desktop.

The Live Clipboard consists of a User Interface component, and three components based on Java-Script: the Live Clipboard object model, serialization of Live Clipboard XML data, and callback methods for event driven application notification as described in [9]. By copying a prepared component from a website, the `copyCallback` method gets invoked, the wrapped component gets serialized using the Live Clipboard XML data format and is copied to the user's clipboard application on the local computer. To paste the copied component the same process is involved in complement direction. At the end of the process the `pasteCallback` is invoked to reveal the copied component.

To overcome the variety of application data formats the Live Clipboard XML data structure may contain multiple source data formats. For instances it may contain the data in the vCard format as well as in the derived hCard format. Thus, if the component is copied to another website the `pasteCallback` method of the destination container will use the Microformat representation. If the same clipboard content is pasted to Microsoft Outlook it will use the vCard representation to derive local contact information from the given component. Live Clipboard user profit from seamless integration of published web content in many different desktop applications, thus the border between desktop applications and WWW applications further and further.

3.3 Named Ranges in Microsoft Excel 2003

Another application scenario for Microformats is Microsoft Excel sheets with a lot of numerical information enclosed, but missing semantic annotations. A lot of this Excel sheets are produced in combination with statistical analysis or insurance calculations. This aggregated data is often hard to store and to search for. To annotate this numeric information Microsoft Excel introduced the concept of named ranges which are identifier for a cell, a group of cells, formulas, etc. with a local scope. This offers the possibility to annotate ranges of numeric values with additional meta data, e.g. semantic data as described in [10].

Starting with Microsoft Excel 2003 the export functionality was extended to export also valid XML documents from any Excel sheet. This is the base for

annotation of numeric values. With the help of a proper Extensible Stylesheet Language Transformation (XSLT) it is possible to transform any XML exported Excel sheet to either RDF or Microformats or further formats. In [10] the usage of the Gleaning Resource Descriptions from Dialects of Languages (GRDDL) is described to transform exported XML Excel sheets to RDF.

This offers the possibility for searching numeric data in a completely new way. On the one hand, office user may profit from this small annotation overhead by finding appropriate data much easier and faster. On the other hand, annotated named ranges can be easily reused within Excel sheets by inserting their name at any appropriate other cell. Microformats offer the additional possibility to directly publish data included in Excel sheets as a valid XHTML on a website with semantic annotations.

3.4 Aggregation Services

Semantic annotations are often mentioned in combination with the aim to create more suitable search engine results. In fact, the aggregation of precise information is much easier for machines and users may profit of it. The following real world usage examples show early adopters of the Microformats technology and the benefit for their users.

Yahoo! Local. Yahoo Local aggregates contact information such as telephone numbers and addresses of many "...things you need in your neighborhood...", thus its aim is to provide online regional yellow pages.

On 2006, Jun 21st the development team announced that Yahoo! local now fully supports the Microformats hCalender, hCard, and hReview. One reason for introducing Microformats is to open "...up new data and new possibilities for the developer community to build upon, to make tools that will be genuinely useful to all our users." as described in [11]. At the same time a script for the Firefox extension Greasemonkey occurred at [12]. This script extracts event information from any website containing hCalendar Microformats and directly puts the extracted data into the personal Yahoo! Calendar if available.

technorati.com. Technorati is a ping-based crawler for blogs which also extracts several kinds of Microformats. Additionally, a search engine [13] for contacts, events, reviews and other Microformats is provided. Technorati pushes the vision of Microformats drastically by creating a lot of new key functionality, because the Microformats idea has its origins here. Some main actors in the Microformats movement also worked for Technorati long before Microformats.org was launched.

4 Conclusions

With the knowledge of real world implementers using Microformats and the outlined benefits for their users, it is now possible to compare advantages and

disadvantages while using Microformats or RDF respectively. In the following section advantages for the Microformats usage are outlined and expectations for future developments in the area of semantic annotations are drawn.

4.1 Comparison of RDF and the Microformat approach

The given example in Fig. 5 on page 8 points out one immanent advantage of Microformats: they can be directly included in any valid XHTML document. Therefore, no additional hyperlinks to external files have to be created, such as links to additional RDF or vCard files.

Furthermore, it is not necessary to manage knowledge respectively information twice. In a RDF file the relationships are outlined by including all initial information and adding additional semantic information to it to make them readable by machines. However, the initial information has also to be prepared for human readability, because RDF files coexist next to the real XHTML document containing the human readable information representation. Microformats merge the semantic Meta data directly with the real information in one common XHTML document, thus only a small set of additional data using existing well-known techniques has to be added. This reduces the processing overhead for semantic annotation dramatically and keeps associated semantic annotations as up to date as the initial information is. Each time the initial information gets changed the semantic annotations are also updated.

Additionally, by using Microformats it is possible to keep the overhead for semantic annotations small. People do only use standards if they are quite simple to use and if tool support exists – both is given in case of Microformats.

The origins of Microformats are mainly influenced by the bloggers movement in the Web within the last few years. Therefore, currently a lot of tool support for Microformats exists for blogging engines which provide plugins with semantic annotation capabilities for blog entries. The Microformats hReview, hCard and hCalendar are provided as plugins for blogging engines such as *Wordpress* or *Movable Type*. This helps readers to extract information from blog entries easily for further processing.

4.2 Future Development

The WWW consortium introduced the RDF format for semantic annotations in 1998, but within the last few years the usage of RDF did not increase very fast: "The results of this survey suggest that RDF has not caught on with a large user community." [14]

The Microformats movement was initiated by a group of developers around Tantek Çelik getting to public area at Supernova conference 2005, June 20th. The initial revolution already started in early 2004 with the working draft for the hReview Microformat which was adopted by the large area of bloggers within few months. One main reason for the success of the hReview Microformat is the simplicity of use to include it in already published as well in new published reviews, e.g. on web blogs. It also introduces a basis for a common understanding

of review data as described in [15]. This results in an informal standardization of reviews written on thousands of web blog websites.

Another example for accepted Microformats is the rel-tag Microformat. Its adoption rate is shown in Fig. 7 on the next page which reflects a dramatic exponential growth. But why is it possible to get such a high adoption, although Microformats are not official web standards such as RDF or XML are. The development process of Microformats is completely open and supported by a continual growing user community. Drafts are discussed in a dedicated wiki on the Microformats website [5] where each contributor is able to collaborate without any centralized committee or decision group.

Pessimistic people may say that Microformats are only the next hype in the Web 2.0 which will disappear within the next months when the next hype arises. The probability for this extreme seems to be very small, because global players such as Yahoo already implemented Microformats in their current web contents. The costs to migrate existing contents to provide Microformats are relatively small; the benefits for users are extremely high especially if new application scenarios arise in nearer future. The example of Yahoo boosts the self-confidence of Microformats and will keep the idea alive. This helps to reduce the gap between vendor-specific desktop applications and WWW applications.

Currently, the Microformat spreading is viral, i.e. user-driven and widely in adoption. However, it will not represent the ultimate version of semantic annotations for web contents, because Microformats are not able to cover the whole wide variety of possible application cases. They cover only a very small but really often used application case often used on blogging websites. However, the usage of semantic annotations may grow faster than it did for the last years by using RDF. Because the Microformats approach does not use a complete formal way, it is less complex and will establish the need for semantic annotations of web content widely. However, as it becomes much easier to use semantic annotations this may create a new risk for personal data.

Embedded RDF approaches such as RDFa will result in a similar ease of use as it is currently possible by using Microformats. RDFa is currently not ready for use in valid XHTML documents, because its basis XHTML 2.0 is not released in a final version, yet. The time will show whether RDFa or Microformats will be accepted by the majority of users or whether both approaches can coexist.

Microformats are the early adopters of semantic annotation of web contents by provide a simple mechanism that can be used by everyone publishing information in the WWW. This movement will strengthen the understanding for the need of annotations which builds the basis for upcoming developments. Semantic annotations in any XML valid format may be transformed into other formats using a proper XSLT as described exemplarily in [16]. Thus, the existence of annotations in any XML format – Microformats, RDF, RDFa or any other – is better than their absence. The future of web content annotation will not be decided in the end of this paper, but the decision for the use of semantic annotations has been passed.

Finally, it has to be stated, that also new risks arise by using semantic annotations. It becomes easier for machines to aggregate personal data, thus it is also much easier to create lists of addresses which may be misused. For instances email addresses may be aggregated by spammers for advertisement misuse in a much faster way than it is possible today. Therefore, with the widely establishment of semantic annotations it becomes more and more important for everyone to protected personal data within the WWW.

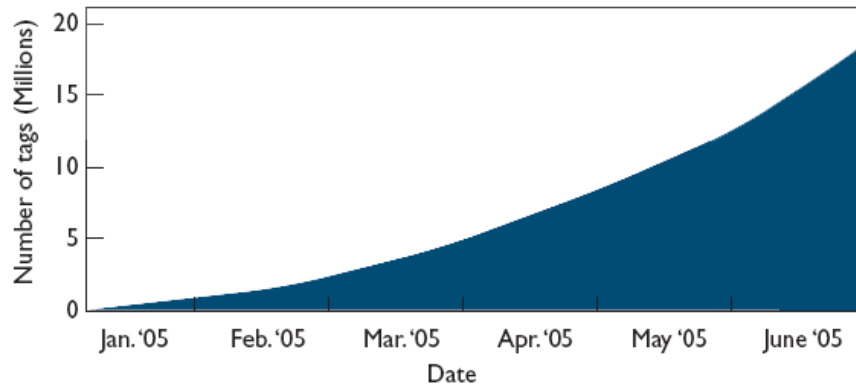


Fig. 7. Microformat adoption rate. Data from Technorati and analysis by researchers at Carnegie Mellon University show that blog posters adopted the `rel-tag` Microformat extremely rapidly (taken from [17])

In section 1 on page 2 of the given paper the reasons for the rapid growth in available web contents in the last few years and the necessity for new content annotations is outlined. The Resource Description Framework (RDF) is described exemplarily as an existing approach for semantic annotations of web contents introduced by the WWW-Consortium. Afterwards, in section 2 on page 3 the emergence of the new approach Microformats is described and a classification of Microformats is given in section 2.3 on page 6. A Microformat example is developed in section 2.4 on page 6 by mapping an existing vCard file to the corresponding Microformat named hCard. An adapted RDF approach called RDFa which is in draft status and should be integrated in future XHTML 2.0 standard is compared to the existing Microformat approach in section 2.5 on page 7. Additionally, the current real world usage of Microformats is analyzed in section 3 on page 9 and chosen providers of Microformats are described and the benefit for their users is evaluated. Ultimately, possible future developments in the area of semantic annotations of web content using Microformats or other equal approaches is outlined in section 4.2 on page 12.

A Available Microformats

Microformat	Description
hCalendar	Calendaring and events format based on iCal-Format [8]
hCard	Contact information format based on vCard-Format [6]
rel="license"	Indicates content licenses
rel="nofollow"	Indicates content which should not be ranked, e.g. by search-engines
rel="tag"	Indicates that the referred resource serves as a tag or keyword/subject for the referring page
VoteLinks	Set of attributes for hyperlinks (vote-for, vote-against, vote-abstain) which are mutually exclusive representing agreement, disagreement, or indifference corresponding to a voting choice
XFN	XHTML Friends Network (XFN) indicates relationship hyperlinks between information authors and other peoples using the rel-attribute of hyperlinks
XMDP	XHTML Meta Data Profiles (XMDP) is a format for defining meta data profiles using elements such as definition lists (DL), definition description (DD), definition terms (DT)
XOXO	Extensible Open XHTML Outlines (XOXO) is an XHTML outline format, e.g. suitable for lists
XOXT	Extensible Open XHTML Tables (XOXT) is an XHTML outline format, e.g. suitable for tables

Table 2. Excerpt of available Microformats (taken from [5])

All Microformats are embeddable into (X)HTML, Atom, RSS, and other XML-compliant document types. Additionally, several Microformats are currently in draft state, e.g. `adr`, `geo`, `hAtom`, `hResume`, `hReview`, `rel-directory`, `rel-enclosure`, `rel-home`, `rel-payment`, `Robots-exclusion`, `xFolk`. For a complete list of currently available Microformats see the wiki on the Microformats Website [5].

References

1. Miniwatts Marketing Group: Internet World Stats. <http://www.internetworldstats.com/stats.htm> (2006)
2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001)

3. McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language. <http://www.w3.org/TR/owl-features/> (2004)
4. Schapranow, M.P.: Microformats – Designed for humans first. <http://myhpi.de/~schapran/pke/> (2006)
5. Çelik, T., et al.: Microformats Website. <http://www.microformats.org> (2006)
6. Dawson, F., Howes, T.: vCard MIME Directory Profile. RFC 2426 (1998)
7. Prodromou, E.: RDFa vs. Microformats. http://evan.prodromou.name/RDFa_vs_microformats (2006)
8. Dawson, F., Stenerson, D.: Internet Calendaring and Scheduling Core Object Specification (iCalendar). RFC 2445 (1998)
9. Ozzie, R.: Live Clipboard Technical Introduction. <http://spaces.msn.com/editorial/rayozzie/demo/liveclip/liveclipsample/techPreview.html> (2006)
10. Nottingham, M.: Adding Semantics to Excel with Microformats and GRDDL. http://www.mnot.net/blog/2005/08/13/excel_microformats (2005)
11. Yahoo! Local MAPS Blog: Yahoo! Local MAPS Blog. <http://ylocalblog.com/blog/2006/06/21/we-now-support-microformats/> (2006)
12. Luk, G.: Greasemonkey Script for Yahoo! Local. <http://upcoming.org/tools/microformats/ycalhcal.user.js> (2006)
13. Çelik, T., et al.: Technorati kitchen. <http://kitchen.technorati.com/> (2006)
14. Eberhart, A.: Survey of RDF data on the Web. Technical report, International University in Germany (2002)
15. Khare, R.: hReview in Review – Purpose — Predecessors — Specification — Implications. <http://cnlabs.commerce.net/~rohit/hReview-in-Review/> (2005)
16. Herrington, J.D.: Separate data and formatting with microformats. <http://www-128.ibm.com/developerworks/library/x-microformats/> (2006)
17. Khare, R.: Microformats – The Next (Small) Thing on the Semantic Web? (2006)