A Methodology for Building Microformats

Wilhelm Loibl, and
Dora Tüz

Institute for Tourism and Leisure Studies
Vienna University of Economics and Business, Austria
{wloibl, dora.tuez}@wu.ac.at

Abstract

As tourism is a very information intense industry, many prospective customers resort to well-known generic search engines like Google to find information about touristic offers. Often the results are not completely satisfactory. One possibility to enhance search engine results is through the use of microformats, XHTML-code which is inserted directly into the webpage and which gives the information contained within well defined meaning. To date several such microformats exist, all of which are for general purposes and not specifically for the tourism domain. In order to encourage the creation of tourism-related microformats the authors propose a methodology for microformats engineering based on existing ontology engineering methodologies. An example in the domain of events exemplifies the usage.

Keywords: microformats, tourism, engineering methodology, semantic web

1 Introduction

Travel and tourism is a very information-intense industry (Werthner & Klein, 1999) consisting mostly of small and medium-sized businesses (Dell'Erba, Fodor, Höpken, & Werthner, 2005). Along with the growing importance of the World Wide Web as a source of information and the decreasing cost of hard- and software many of these businesses set up their own website to provide information and communicate directly with prospective customers (Gratzer, Werthner, & Winiwartner, 2004). With such a huge wealth of information from many different sources available, online travellers and tourism professionals alike struggle to gather information of high quality (Scharl, Dickinger, & Weichselbraun, 2008). In order to find the information necessary for their trip planning many tourists resort to generic search engines like Google. These search engines act as "gateways" to travel-related information on the web and are important traffic generators for tourism-related websites (Xiang & Gretzel, 2010). The usefulness of current search engine technology is limited though. On the one hand this is because of the inability of many users to form precise queries. On the other hand most search engines still rely on matching the terms of the user's query against the (weighted) terms of the individual documents without considering the content itself (Berry & Browne, 1999).

A more thorough discussion about the shortcomings of search engines can be found in Henzinger (2007). What is needed in order to improve search results are so called Semantic Searches (Maedche & Staab, 2002). Web pages are enriched with additional metadata describing the content of the page (Davis, Studer, & Warren, 2006). The data taken from these documents can now be combined with information gathered from other documents to answer more complex queries. Constructing a web where machines can read and interpret information is the basic idea behind the "Semantic Web" (Antoniou & van Harmelen, 2008). Over the course of the last decade several

M. Fuchs et al. (eds.), *Information and Communication Technologies in Tourism 2012*

initiatives have created such systems of metadata for the tourism industry. These range from full-fledged ontologies, describing the whole domain of discourse in greater detail, to embeddable semantics which are inserted directly into the HTML-code of a webpage. Although the technology already exists, uptake by the industry has been rather slow (McDowell et al. 2003, Haustein & Pleumann, 2002). The reason lies in the missing motivation because much technical sophistication and substantial effort on the part of the website designer is necessary to structure the data. In order to justify structuring the data it must be rewarding and as easy as possible to do so. Microformats satisfy these conditions.

In order to encourage the creation of microformats for the tourism industry, the authors propose a new microformats engineering methodology building on existing methodologies for ontology engineering. This greatly facilitates the engineering process using well known tools and reduces the costs. In the remainder of section 1 the basic terminology used throughout the paper will be explained followed by a discussion about ontologies and microformats in the domain of tourism. Section 2 contains a discussion about the state of the art in ontology as well as microformats engineering and thereby laying the foundation for the research methodology (section 3) and the new microformats engineering methodology presented in section 4. The example given in section 5 helps to understand how the methodology is used in practice. The managerial implications given in section 6 outline how tourism businesses may benefit from this work while section 7 points out which steps must still be taken for the industry to fully profit from these technologies.

1.1 Basic Terminology

Some terms like domain or ontology have already been mentioned without further explanation of their meaning. The following section gives a short overview of important Semantic Web terminology used in this paper.

- Syntax vs. Semantics. The syntax is grammatical rules whereas semantics denote the meaning of an expression. In a web context semantics can be added to a webpage by including metadata (i.e. data about data) to an online document describing the content or the document itself (e.g. the author or creation date).
- Domain. A domain is a collection of all entities about a specific subject (Hjorland & Albrechtsen, 1995). The domain of tourism therefore consists of all technological elements and information entities related to travel (Xiang, Wöber, & Fesenmaier, 2008).
- Ontology. Ontologies provide a formal description of a certain domain (a finite list of terms and the relationships between these). Terms are classes of objects (concepts) which are important in the domain of discourse. The relationships are defined using ontology languages (Antoniou & van Harmelen, 2008). The process of manually constructing domain ontologies may become rather complicated according to the size and complexity of the domain.
- Embeddable Semantics is an umbrella term for all of those formats which use (X)HTML tags to integrate metadata into the code of a webpage itself. Two important such formats are RDFa and microformats.
- RDFa only defines the syntax and is interoperable with RDF. All vocabularies (the terms allowed in the document) are defined independently and are freely

intermixable (Adida, Birbeck, McCarron, & Pemberton, 2008[Sept. 3, 2011]). To extend RDFa in such a way, the RDF schema mechanism can be used (Adida & Birbeck, 2008[Sept. 3, 2011]).

- Microformats define the syntax as well as the vocabulary. They are generally designed for human consumption first, machine readability and interpretability comes only second. As they can only be used for describing explicit information more sophisticated knowledge representation like inference is not possible (Khare & Celik, 2006).
- A schema either defines the structure of a document (XML) or a set of terms (a vocabulary) which can be used in a data model (RDF(S)) (Antoniou and van Harmelen, 2008). In the latter case a schema adds additional meaning to data, providing guidelines for its interpretation like certain relationships or constraints (Allemang & Hendler, 2008).

1.2 Ontologies and Microformats in the Tourism Domain

Despite the difficulties inherent in constructing domain ontologies, several projects for the domain of tourism exist. An overview of these ontologies is given by Feilmayr and Pröll (2009). Although RDFa offers more expressivity (Adida, 2008), microformats are currently deployed in greater numbers (Lewis, 2010). Unfortunately, overall usage numbers are still quite limited. As Loibl (2011) shows for the Austrian market only about 1.2 % of all websites of tourism-related businesses contain semantic markup. Most of this markup is comprised of microformats giving general information about the address e.g. in the form of a vcard microformat. An example of such a format, taken from the website of an Austrian golf resort at http://www.golfresort.at is given below.

```
<address id="c-footer" class="vcard">
  <span class="org fn">
    <span class="organization-name">Golfresort
    Haugschlag</span>
  </span>,
  <span class="adr">
    <span class="postal-code">A-3874</span>
    <span class="locality none">Haugschlag</span>
    <span class="street-address">Haugschlag 160</span>
    <span class="country-name none">Austria</span>
  </span>, T:
  <span class="tel">
    <span class="type none">home</span>
    <span class="value">+43 (0)2865/8441-0</span>
  </span>
</address>
```

As illustrated by this example of a virtual business card above, microformats are added to web pages by inserting class-attributes with specified values (Lewis, 2010). For the human user this HTML code is rendered as a normal web page with address data. Any program capable of reading and interpreting the additional metadata the information contained in the page can be extracted and used for further purposes like

exporting the data into the address book of an e-mail program. In the tourism domain microformats could be used to automatise communication between companies. For example restaurants have to manually update data kept on restaurant search engines. Using microformats the changes must only be made on the company website and all partner businesses can automatically extract the new data from it. This is especially helpful for data that changes more often like weekly menu cards, room availability with hotels or air fares.

According to Lewis (2010) only 8 stable microformats exist with another 15 being in a draft status. None of these explicitly support the description of information specifically for tourists. In order to encourage the creation of microformats for the use in tourism businesses this paper provides an easy to follow step-by-step engineering approach.

2 State of the Art

As creating microformats is to some extent quite similar to the process of ontology engineering, the following section provides an overview of prevalent methodologies in this area.

2.1 State of the Art in Ontology Engineering

The ontology engineering methodology used depends on what type of ontology is to be created. Representation ontologies do not make claims about the world but instead provide a representational framework. The primitives provided form the basis for the description used in domain and generic ontologies. Domain ontologies describe concepts which are specific for a certain domain, e.g. accommodations while generic or upper level ontologies define concepts that are generic across many different domains. Application ontologies model all knowledge required for a particular application including method- and task-specific extensions. Such ontologies are not reusable themselves (van Heijst, Schreiber, & Wielinga, 1997).

Although different methodologies for building different types of ontologies exist, a common sequence of steps can be distinguished (Pinto & Martins, 2004). During the specification phase the purpose of the ontology is identified together with its intended areas of application. In the conceptual phase a conceptual model is formulated describing the concepts in the domain of discourse together with the relationships among them. Often groups of highly interlinked concepts can be partitioned into sub-ontologies. This conceptual model is then formalized using axioms which restrict possible interpretations of the meaning of the concepts contained in the ontology (formalization stage). The formal model is then implemented using a representation language and the implementation is committed to updates and corrections until the end of its life cycle. In addition to these sequential steps, other activities have to be performed throughout the life cycle. One of the most important activities for ontology engineering is the acquisition of knowledge. This can be accomplished through expert interviews, brainstorming, text analysis or any number of other techniques.

Evaluating the quality of prototypes according to the criteria defined in the specification phase helps guiding the evolutionary process. An extensive documentation regarding all decisions taken during the development process

facilitates maintenance as well as possible reuses of the ontology (Pinto & Martins, 2004). Generic ontologies are mostly built from scratch or by merging other generic ontologies. Several methodologies like TOVE (Grüninger & Fox, 1995) and METHONTOLOGY (Fernandez, Gomez-Perez, & Juristo, 1997), designed for building ontologies from scratch, gained a lot of attention in the relevant literature. According to Suarez-Figueroa (2010) even newer developments like DILIGENT (Pinto, Staab, & Tempich, 2004) and On-To-Knowledge (Staab, Schnurr, Studer, & Sure, 2001) do not provide detailed guidelines for reusing existing ontologies. The focus of these methodologies lies on collaborative engineering. Domain ontologies can be built from scratch as well or by reusing modules from other ontologies.

The methodology proposed by Pinto and Martins (2001) recommends that the process of composition should already start in the conceptualization phase and consists of seven consecutive steps. First the general integration possibility must be assessed. This depends on the framework that is being used for building the ontology. The individual building blocks (sub-ontologies) making up the future ontology together with the knowledge represented in each module and the assumptions related to the modules, have to be identified. Next possible candidate ontologies, representing the main concepts, have to be identified. If necessary, these ontologies must be reengineered before they can be evaluated. After the assessment, the most adequate source ontologies can be chosen and integrated into the resulting ontology. This integration process may involve adapting the information contained in a module, specialization (resulting in a more specific domain ontology) or inserting more general information (augmentation). The last step involves the evaluation of the resulting ontology according to specific evaluation criteria (Pinto & Martins, 2001).

2.2 State of the Art in the Creation of Microformats

For the design of microformats only guidelines and design principles exist. Allsopp (2007) even discourages the premature creation of a new microformat for a specific task. Instead all other possibilities for solving the problem at hand must be taken into consideration and only if these prove to be insufficient a new microformat may be created. Reuse is the first of three design principles defined by Khare (2006). An example would be the hCalender microformat which is a 1:1 implementation of the iCalendar standard defined in RFC2445 (Celik & Suda, 2011[Sept. 3, 2011]). Another important design principle is to reduce complexity. A microformat should always focus on solving a specific problem by employing the simplest solution.

The last design principle is recyclability. Microformats should be modular and easily embeddable. Based on these principles, Allsopp (2007) describes a series of tasks as a guideline for creating a microformat. The starting point is always a precise problem and not some hypothetical issue. This is important because it means that some solution in the form of a web page already exists. Wherever data is provided for human consumption first and machines second, there is an ideal field of application for microformats. If the problem at hand is generalisable for a greater number of users this constitutes an ideal candidate for a microformat. In order to determine generalisability, the engineer may turn to the community for advice. The

microformats community provides a mailing list for such discussions[1]. The next step involves researching current behaviour. A new format should contain only a minimal set of properties that is really needed and leave all excess data. To get an overview of the data the microformat has to structure, a large sample of real-world web pages of the respective domain ought to be investigated. The microformats community again helps by providing a wiki where the research process can be documented[2]. When a consensus has been reached as to the data that should be structured it is beneficial to search for existing standards already describing this data. Not only does reusing existing standards save a lot of time and effort it also improves the quality of the solution and interoperability with existing applications. At this stage the resulting format can be specified as a draft schema using an XHTML Meta Data Profile (Celik, 2003[Sept. 3, 2011]) and is open for discussion. An important indicator for maturity is usage. When a microformat is widely used it can be seen as mature. Such a pattern is characteristic for an evolving prototyping life cycle.

3 Research Methodology

A methodology in the context of this work is an integrated series of stages or phases that is used to structure a class of thought intensive work. A stage consists of a series of activities. Each activity is composed of tasks, which are the smallest units of work. These tasks are allocated to individual project members who are held accountable for their execution (Suarez Figueroa, 2010). The research methodology used for this work can be divided into two steps. In a first step, methodologies for software and ontology engineering as well as available design principles for microformats were contrasted with each other. This enabled the authors to recognize the focal point in the microformats engineering process. In addition, a preliminary methodology based on ontology engineering methodologies was created. In a next step this preliminary microformat engineering methodology was utilized to create a microformat for the tourism domain. The lessons learned from this practical application were then used to refine the initial engineering methodology.

4 Proposition of a Methodology for Creating Microformats

Ontology engineering methodologies stress the importance of knowledge acquisition (Pinto & Martins, 2004). This is especially true for microformats. As has already been noted, a microformat should be based on existing standards. Furthermore, it has to contain only the data necessary to solve a specific problem. Because of this importance of knowledge acquisition it has a very prominent position in the methodology and has to be explained more thoroughly. Design activities are often differentiated into conceptualization and formalization phases (Pinto & Martins, 2004). For building microformats such a distinction is not absolutely necessary as microformats are not as strictly formalized as ontologies are. Fig. 1 gives an overview of the methodology proposed in this work. It is based on ontology engineering methodologies and puts special emphasis on knowledge discovery. The remainder of section 4 describes the phases and individual steps of the methodology.

[1] See: http://microformats.org/mailman/listinfo/microformats-discuss
[2] See: http://microformats.org/wiki

In the specification phase the first step is problem definition. A problem that could be solved using microformats usually stems from some shortcoming of the website. Capturing motivating scenarios as defined by Grüninger & Fox (1995) help to uncover examples where a problem was not adequately addressed by current technology. These scenarios already hold the core of a possible solution. The next step defines the field of application. Useful tools are UML use cases showing certain scenarios how users interact with a system (Fowler, 2003) or informal competency questions (Fernandez-Lopez & Gomez-Perez, 2002). Any suitable microformat must be usable for answering these questions. In a web context a good starting point for formulating these competency questions lies in the search behaviour of the users of the website. Another use case would be the automatic extraction of a weekly menu card by gastronomy search engines. Here the main question would certainly be what type of information these companies present on their search platforms. The answer may be found on the administration web page of the respective search engines or by analysing the search web-interface.

Knowledge acquisition is definitely the most important. First, the information provided on the own website must be analyzed by extracting all information that is to be structured. For a restaurant this may be the online menu card. This information is collected into classes and provided with a label. In the remainder of this work these labels are referred to as terms. If someone wanted to create a microformat to structure an online menu card, the first source to investigate would most certainly be his or her own website. The information there likely would inform the user about different types of dishes like soup or desserts. Under the “soup”-heading the user would find different objects of type “soup” like French onion soup together with additional information about that soup like price or ingredients. Consequently “French onion soup” would be an object of type “dish” (class label) whereas the heading grouping all objects of type “soup” together would be generalized into the term “dishtype”.

The resulting microformat would contain a class “dishtype” containing one or several objects of type “dish”. On a web page these more general objects like “soup” would certainly be marked up using certain tags like header-tags or using tables. A resulting DOM-subtree may be used for structuring the emerging microformat. The Document Object Model offers a system-independent convention for representing and interacting with objects found in HTML-documents. Therefore a HTML-document may be represented as a tree (or more than one tree) where a subtree is a smaller part of that structure. As microformats are primary for human consumption and only secondary for machines retaining the structure discovered on domain specific websites is mandatory.

At this stage though it is still not determined that a new microformat has to be created at all. Possible alternatives may take the form of existing microformats, ontologies, XML schemas or some other way to structure online data like e. g. semantic HTML. Only if no suitable alternatives (as compared to the competency questions) emerge, the creation of a microformat is warranted. Based on the terms extracted from the own website, other relevant sources of information are examined.

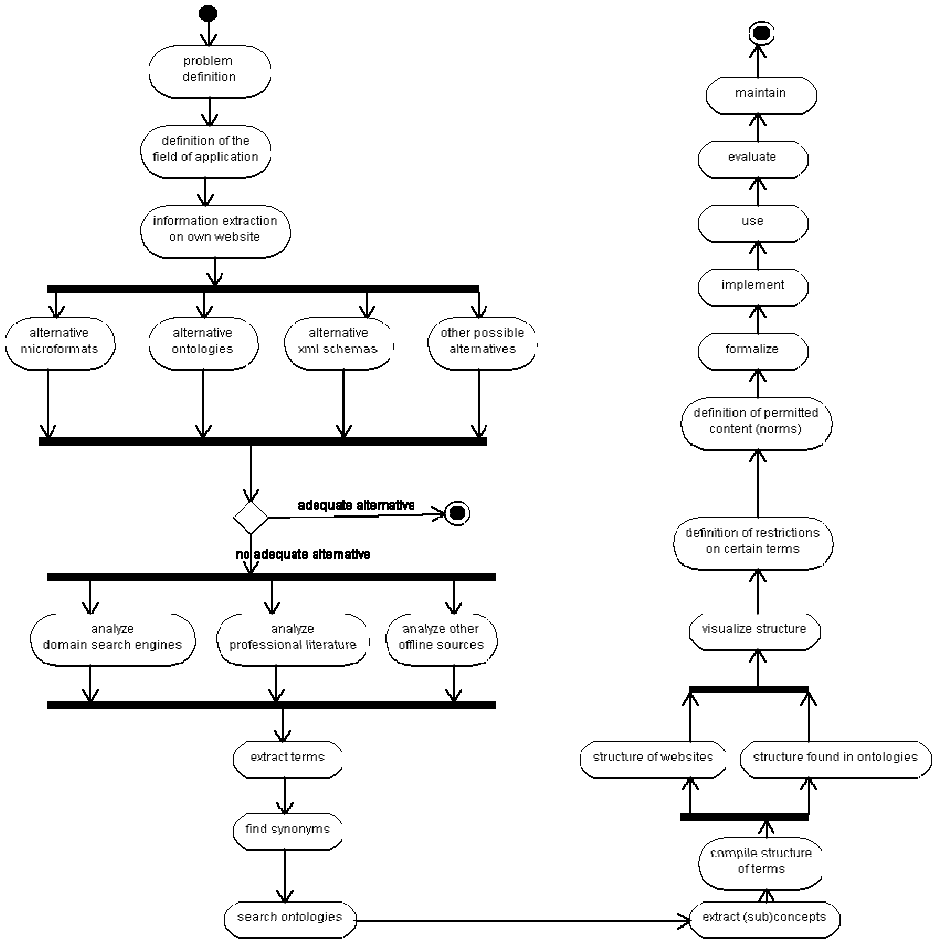

Fig. 3 Microformats Engineering Methodology

These sources are found offline, like catalogues or offline menu cards. Another source of inspiration is the search interfaces of domain search engines and professional literature. As with all external sources the legal terms must be considered. They regulate how if and how these sources may be used. All terms as well as their synonyms are then used for searching ontologies. A synonym for “beverage” would be “drink”. At this stage a semantic search engine is used to see if the concepts of “beverage” or “drink” are already described somewhere else. These descriptions may be utilized for defining the structure of the microformat (conceptualization phase).

The fundamental structure of the microformat is taken from the websites mimicking the visible structure of the web pages. If the visible structuring of the data is constant over all investigated websites the microformat should be grounded on that structure (e. g. most menu cards found on restaurant websites are similarly structured). Another important aspect is the reduction of the terms. Only terms that appeared in most

sources and can be seen as absolutely necessary to answer the competency questions should be incorporated into the microformat. A “dish”-element may provide information that a diabetic may eat it. But if no application uses that information it may be left out. The final step in the conceptualization phase is to visualize the resulting structure using either trees or graphs.

In the formalization phase certain restrictions on the terms may be defined. It must be noted that microformats are considerably less formalized than ontologies. Therefore this step is not mandatory. Much more important is the definition of the permitted content. When describing the price for the French onion soup it may be necessary to give details about the currency the price is in. For describing currency the norm ISO 4217 exists (http://www.currency-iso.org/iso_index/iso_tables/iso_tables_a1.htm). The results of the last two steps are summarized using an XHTML Meta Data Profile Description (Celik, 2003[Sept. 3, 2011]). The authors regard the description in the profile as a prototype that may be utilized for the uses defined during specification phase. All following phases aim at enhancing the prototype until suitable stability has been reached.

5 Example in the Tourism Domain

In order to learn more about the applicability of the preliminary methodology, a microformat for describing information about touristic events was created. The focus was on structuring information about events as well as any necessary additional information like tickets. The requirements were defined based on use cases. First a search for existing microformats and ontologies yielded the hCalendar microformat which is suitable for describing events as well as the ebSemantics event ontology (ebSemantics, n. d.[March 3, 2011]). These were found to be inadequate because they did not support the mark up of ticketing information. Consequently the existing microformat was to be extended. Next, a sample of online event calendars found on community websites as well as a sample of specialized event websites was examined manually. The majority of Austrian community websites provide an event calendar for local events aimed at locals and visitors alike.

A list of 2328 links to websites was compiled and a sample of 30 pages was chosen randomly. For each of these online event calendars the information presented on them was extracted manually and categorized into classes and labelled. The same was done for online event calendars. The results from the first result page of the search engine Google, using the keywords “event” and “österreich”, were analyzed. These terms together with their synonyms (gathered using WordNet[1]) were used as input into the semantic search engine Swoogle[2]. The results were analyzed manually to find any sub-ontologies that could be used for further describing the data discovered on the websites. In addition, a literature review was done to find ontologies not yet analyzed. The common structure discovered on the websites was enriched with suitable sub-ontologies. Afterwards, the core structure, necessary to accomplish the requirements defined at the beginning, was extracted and visualized using graphs.

[1] See: http://wordnetweb.princeton.edu/perl/webwn
[2] See: http://swoogle.umbc.edu/

At the end two XMD profiles, one for pricing the other for ticket-information, were defined. These two microformats are modular, the ticketing-microformats uses the pricing-microformat to mark up information about prices.

6 Managerial Implications

Most tourism businesses nowadays depend to a large extent on internet communication services for contacting their customers. Unfortunately the majority of companies is of small or medium size and often lacks the financial resources or the knowledge to build sophisticated websites. Microformats may help these firms solve some problems like poor visibility on the web. With this work the authors pursue several goals. First and foremost they want to spread the idea of using microformats. With rising usage numbers even more tourism businesses or web agencies may be persuaded to use them creating a snowball effect. Of course this vision can only become true if microformats exist which are useful for tourism businesses. To this end, a methodology for creating such formats is proposed in the work at hand which is precise enough to be used out of the box but sufficiently easy to be used by businessmen with less experience in microformats engineering. For the most part, the methodology was prepared for academics though and should be seen as an invitation to provide the industry with ready to use microformats.

7 Conclusions and Further Research

This work is supposed to provide a methodology that is easy to use and precise enough to create microformats for immediate use. The first issue, easy usage, is tackled by providing a clear sequence of steps that have to be followed when creating a new microformat. As the methodology is still in a very early stage of development, two main areas of discussion remain. Further research may discover that some important steps are still missing in the methodology. The second issue is automation. The work for compiling the example presented in section 5 was mostly done manually. Better tool support eases usability for less experienced microformats. This also applies to the second issue, precision. For each step shown in Fig. 1 an assortment of adequate methods has to be presented. These methods must be described in great detail so the user can decide which method to use. The two issues raised above show quite clearly that the current methodology is only a first step. Further research has to improve this methodology by employing it for the creation of microformats. All experience gained through application can then be fed back into a next revision. Other approaches like the stage-gate approach may further enhance the methodology (Cooper, Edgett, & Kleinschmidt, 2002)

Another problem is the slow uptake of the technology. In order to uncover barriers which impede the usage of microformats the authors will analyse tourism related websites in Austria according to their use of semantic metadata. Based on the insights gained through this analysis users and non-users of semantic formats will be questioned as to their motivations.

References

Adida. B. (2008). hGRDDL: Bridging microformats and RDFa. Web Semantics. *Science, Services and Agents on the World Wide Web* 6(1): 54-60.

Adida, B. & Birbeck, M. (2008). *RDFa Primer.* Retrieved from http://www.w3.org/TR/xhtml-rdfa-primer/

Adida, B., Birbeck, M., McCarron, S. & Pemberton, S. (2008). *RDFA in XHTML: Syntax and Processing*. Retrieved from http://www.w3.org/TR/2008/REC-rdfa-syntax-20081014

Allemang, D. & Hendler, J. (2008). *Semantic Web for the Working Ontologist modelling in RDF, RDFS, and OWL*. Amsterdam: Morgan Kaufmann.

Allsopp, J. (2007). *Microformats. Empowering Your Markup for Web 2.0*. Berkeley. Friends of ED.

Antoniou, G. & van Harmelen, F. (2008). *A Semantic Web Primer, second edition*. Cambridge: The MIT Press.

Berry, M. & Browne, M. (1999). *Understanding Search Engines. Mathematical Modeling and Text Retrieval*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

Celik, T. (2003). *XMDP: Introduction and Format Description.* Retrieved from http://gmpg.org/xmdp/description

Celik, T. & Suda, B. (2011). *hCalendar 1.0.* Retrieved from http://microformats.org/wiki/hcalendar

Cooper, R.J., Edgett, S.J. & Kleinschmidt, E.J. (2002). Optimizing the Stage-Gate Process: What Best-practice Companies Do-I. *Research Technology Management* 45(5): 21-27.

Davis, J., Studer, R. & Warren, P. (2006). *Semantic Web Technologies: Trends and Reseach in Ontology-based Systems*. Chichester: Wiley.

Dell'Erba, M., Fodor, O., Höpken, W. & Werthner, H. (2005). Exploiting Semantic Web Technologies for Harmonizing E-Markets. *Information Technology & Tourism* 7: 201-219.

ebSemantics (n. d.). Retrieved from: http://www.ebsemantics.net/doc

Feilmayr, C. & Pröll, B. (2009). Ontologiebasierte Informationsextraktion im eTourismus. In M., Lassnig and S., Reich (Eds.), *eTourismus, Praxis der Wirtschaftsinformatik, Heft 270*. Heidelberg: Dpunkt.verlag.

Fernandez-Lopez, M. & Gomez-Perez, A. (2002). Overview and Analysis of Methodologies for Building Ontologies. *The Knowledge Engineering Review* 17(2): 129-156.

Fernandez-Lopez, M., Gomez-Perez, A. & Juristo, N. (1997). METHONTOLOGY: from ontological art towards ontological engineering. In *Proceedings of AAAI97 spring symposium series, workshop on ontological engineering*. Stanford, CA.

Fowler, M. (2004). *UML Distilled. Third Edition. A Brief Guide to the Standard Object Modeling Language*. Amsterdam: Addison Wesley.

Gratzer, M., Werthner, H. & Winiwarter, W. (2004). Electronic business in tourism. *International Journal of Electronic Business* 2(5): 450-459.

Grüninger, M. & Fox, M. (1995). Methodology for the design and evaluation of ontologies. In *Proceedings of IJCAI95's workshop on basic ontological issues in knowledge sharing*. Montreal, Canada

Haustein, S. & Pleumann, J. (2002). Is Participation in the Semantic Web Too Difficult? In I. Horrocks & J. Hendler (Eds.), *ISWC 2002, LNCS 2342*.

Henzinger, M. (2007). Search Technologies for the Internet. *Science* 317(5837): 468-471.

Hjorland, B. & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain-analysis. *Journal of the American Society for Information Science* 46(6): 400-425.

Khare, R. & Celik, T. (2006). Microformats: a pragmatic path to the semantic web. In *Proceedings of the 15th international conference on World Wide Web*. Edinburgh, Scotland: ACM Press.

Khare, R. (2006). Microformats: The Next (Small) Thing on the Semantic Web? *IEEE Internet Computing* 10 (1): 68-75.

Lewis E. (2010). *Microformats made simple*. Berkeley: New Riders.

Loibl, W. (2011). An Overview of the Use of Semantic Markup on Tourism Websites in Austria. In *TTRA Europe 2011 Conference Proceedings. Creativity and Innovation in Tourism*. Archamps, France.

Maedche, A. & Staab S. (2002). Applying Semantic Web Technologies for Tourism Information Systems. *Proceedings from ENTER 2002*. Innsbruck, Springer.

McDowell, L., Etzioni, O., Gribble, S., Halevy, A., Levy, H., Pentney, W. & Vlasseva, S. (2003). *Evolving the Semantic Web with Mangrove*. Technical Report UW-CSE-03-02-01, February 2003.

Pinto, H. S. & Martins, J. P. (2001). A methodology for ontology integration. In *Proceedings of the 1st international conference on knowledge capture*. New York: ACM Press.

Pinto, H. S. & Martins, J. P. (2004). Ontologies: How can They be Built? *Knowledge and Information Systems* 6. London: Springer.

Pinto, H. S., Staab, S. & Tempich, C. (2004). DILIGENT: Towards a fine-grained methodology for Distributed, Loosely-controlled and evolving Engineering of oNTologies. In R. L. Mantaras & L. Saitta. *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)*. Valencia, Spain.

Scharl, A., Dickinger, A. & Weichselbraun, A. (2008) Analyzing News Media Coverage to acquire and structure Tourism Knowledge. *Information Technology & Tourism* 10: 3-17.

Simperl, E. P. B. & Mochol, M. (2006). Cost Estimation for Ontology Development. In W. Abramowicz (Ed.). Proceedings of BIS 2006. Poznan, Poland.

Staab, S., Schnurr, H. P., Studer, R. & Sure, Y. (2001). Knowledge Processes and Ontologies. *IEEE Intelligent Systems* 16(1): 26-34.

Suarez Figueroa, M. C. (2010). NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse. (Unpublished doctoral dissertation). Universidad Politecnica de Madrid.

Van Heijst, G., Schreiber, A. T. & Wielinga, B., J. (1997). Using explicit ontologies in KBS development. *Int. J. Human-Computer Studies* 45: 183-292.

Werthner, H. & Klein, S. (1999). *Information Technology and Tourism - A Challenging Relationship*. Wien, New York: Springer.

Xiang, Z. & Gretzel, U. (2010) Role of social media in online travel information search. *Tourism Management* 31: 179-188.

Xiang Z., Wöber K. & Fesenmaier D. (2008). Representation of the Online Tourism Domain in Search Engines. *Journal of Travel Research* 47(2): 137-150.