A Framework for Media Adaptation Using the Web and the Semantic Web

Sébastien Laborie and Antoine Zimmermann
INRIA Rhône-Alpes and LIG - France
{Sebastien.Laborie,Antoine.Zimmermann}@inrialpes.fr

Abstract

The World Wide Web can be accessed through a number of different devices, each having its own capabilities and limitations. Additionally, the content of the Web is increasing tremendously in size and variety. Yet, many devices do not embed support for all types of media and formats. Therefore, in order to provide as much information as possible to all kind of devices, media items have to be adapted. In this paper, we propose to adapt them by replacing incompatible media items by others found on the Web. The adapted media items must convey the same message as the original ones, while satisfying the target profile. We present a possible architecture to implement this and we show that search engines can already achieve this to a limited extent. Nonetheless, some results are unsatisfactory because media annotations lack semantics, are partial and are heterogeneous. Hence, we propose to use Semantic Web technologies, such as RDF descriptions, ontologies, ontology merging and matching, in order to select better alternatives, thus improving this adaptation framework.

1. Introduction

In only a few year period, the number of devices able to connect to the World Wide Web (e.g., desktop computers, laptops, PDAs, mobile phones) has increased steeply. Still, the wide variety of media types is mostly inaccessible to many devices due to their technical characteristics (e.g., screen size, memory, bandwidth), or user's personal profile (e.g., language, handicap). Indeed, media types (texts, images, sounds and videos), formats (e.g., avi, mpeg, jpeg, gif) or specific characteristics (e.g., quality, size) may not be executable because of these environment constraints. To avoid ambiguity, we call a *media item* an atomic multimedia object such as a single picture or video[1].

In order to ensure universal access to the Web, incompatible documents must be adapted, i.e., transformed into documents complying with the target context before being played. This paper focuses on media item adaptation.

We propose to deal with this problem by replacing incompatible media items by compatible ones selected among a set of possible alternatives, which are retrieved dynamically on the Web. This choice is motivated by concrete examples in Sect. 2. Accordingly, we define an adaptation framework in Sect. 3 composed of several modules which interact with the World Wide Web. As a first step, media items available on the Web are indexed according to their descriptions. Then, the adaptation process consists in retrieving the most similar description of an adapted media item satisfying the target profile. Although this framework has not been implemented yet, we simulate this approach by taking advantage of existing search engines and show that it is already capable of promising—yet limited—results.

Now, the World Wide Web is changing into the Semantic Web [2], where annotations are even more expressive. This ensures a far better retrieving process. Sect. 4 details how currently developed Semantic Web technologies not only can improve our adaptation mechanism, but also can overcome heterogeneity and incompleteness of semantic descriptions. Sect. 5 discusses possible limitations and open problematics. Finally, Sect. 6 presents other existing frameworks or systems that are related to this issue.

2. Motivating Examples

Consider a video about the 2004 tsunami in Indonesia which is found online. It can potentially be consulted by several net access devices, some of which are not able to execute the original media item. In order to adapt it, a simple method consists in browsing the Web and selecting an alternative media item that conforms to the target profile. Using existing search engines, the Web holds many versions of this very event in terms of media types, formats and characteristics as shown in Tab. 1. Whatever the target media type or characteristics, the response time is almost instantaneous. Obviously, this time does not take into account the selection of an appropriate solution, but a suitable result conveying the initial information is quickly found.

[1] Other terms are *media element*, *multimedia object*, *media object*.

Media type	Media characteristic	Results number	Response time
Videos[2]	All durations	171	0.025s
	$>$ 20 minutes	12	0.013s
	4-20 minutes	45	0.02s
	$<$ 4 minutes	104	0.027s
Images[3]	All sizes	77200	0.03s
	Large	2310	0.11s
	Medium	73700	0.09s
	Small	2720	0.1s
Texts[4]	Any	1540000	0.23s
	English	1230000	0.23s
	French	39700	0.21s
	Indonesian	76200	0.37s

Table 1. Results provided by Google search engines with the query “Tsunami December 2004”.

Another such situation occurs when a person sends a movie trailer to a mobile phone which is not able to play video. In this case, it would be acceptable to display the poster and play an audio track when it is possible, otherwise only display the synopsis. For most commercial movies, this information is easily found on the Web in several formats, especially on the movie studio's website.

Even if the media item is executable on the target, the profile may specify other constraints such as language, content protection, or personal preferences that necessitate adaptation. Our approach is still usable and effective in this context. Eventually, adaptation is also useful when, though the media item is executable, the end-user is not satisfied and wants an alternative (e.g., finding different points of view, camera angles).

Other cases may arise when sharing images or videos of famous persons (e.g., Albert Einstein, Brad Pitt), masterpieces (e.g., Mona Lisa, The Last Supper), monuments (e.g., Eiffel Tower), sport events, etc.

In the following section, we describe an architecture that takes advantage of the profusion and diversity of the content of the Web in order to adapt media information.

3. A Framework for Media Adaptation

Our goal is to retrieve from the Web a media item that matches two requirements: (1) it must be executable on the target device (i.e., it must conform to the target profile) and (2) it has to convey the same information as the original media item, or at least, provide a message as close as possible to the original one. Obviously, we expect the system to do it automatically. For that purpose, we first present a possible architecture to implement it (§3.1). It is generic enough to be adapted to various types of description. Thereafter, we show that existing technologies can already achieve this to a limited extent (§3.2).

3.1. The general strategy

The adaptation approach consists in replacing incompatible media by compatible ones selected among a set of possible alternatives, which are retrieved dynamically on the Web. Fig. 1 gives an overview of the framework.

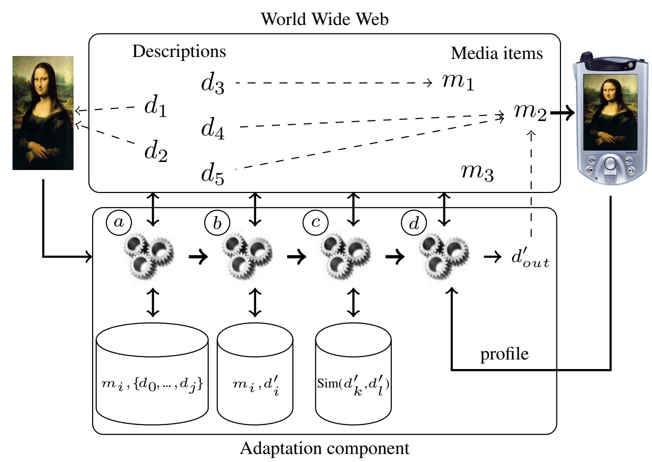

Figure 1. Media adaptation scheme.

We assume that media items are identified by their URI, including the initial one. In Fig. 1, media items are represented by m_i, each description d_i refers to exactly one media item's URI (dashed arrows).

The descriptions may be web pages, formal annotations, or automatically generated metadata available on the Web. The presented architecture uses four modules (ⓐ, ⓑ, ⓒ and ⓓ in Fig. 1), that can be implemented as Web services. The composition of these modules or services is a software component that takes a source media URI and a target profile as input, and produces as output a replacement media item that conforms to the profile. Each module and the overall component are detailed in the following.

Description association ⓐ: This module is used to retrieve all the descriptions related to a particular media item. The descriptions are indexed and cached in a database, constantly updated by automatic Web mining techniques. It

[2] http://video.google.com
[3] http://images.google.com
[4] http://www.google.com

may happen that a media item is related to several descriptions. More precisely, the index table associates to each media item's URI the set of descriptions that refers to it on the Web. The stored descriptions can be in any format, as long as their structure can be understood by the next module.

Description aggregation b: This module takes a set of descriptions as input and produces a description that aggregates all the information found in the input descriptions. This operation is needed because media annotations are usually poor individually, while merging them would enrich the media description. Data integration techniques are used to do this, that can be as simple as concatenating textual descriptions or making an index of keywords, to very elaborate database integration methods [18]. The aggregated descriptions are cached and updated as soon as a new description is found by the previous module, or when an existing description is updated.

Description similarity c: When media descriptions have been aggregated in such a way that only one unique description corresponds to a given media item, these have to be compared in order to find the most similar ones according to the input document. More precisely, a pairwise comparison using data or metadata similarities is computed for each pair of descriptions, and the results are cached too, and updated when an aggregated description changes. An example of a similarity measure applied on Web pages can be found in [3].

Description selection d: Among an ordered set of alternative media descriptions, this process selects the first description d'_{out} that conforms to the target device environment constraints and personal profile.

The adaptation component: The inputs of the overall component are the initial media URI and the target profile. The URI is used to retrieve the corresponding aggregated description d'_i from module b. Then, it computes the list of aggregated descriptions ordered by similarity with d'_i. Finally, this ordered list and the target profile are sent to the last module which returns a description that refers to the adapted media item (m_2 in Fig. 1) that will be sent to the target device.

3.2. A case study

To demonstrate that this approach is concretely feasible, this section describes how existing technologies already implement most of the framework presented above. Additionally, we simulate the execution of the framework using a Web-based search engine on a concrete example.

Someone wants to send a picture of Mona Lisa to someone else having a mobile phone. The initial image is in PNG format in resolution 560×864.[5] The target mobile phone cannot display PNG images and is limited to a maximum resolution of 400×600. A web page refers to the picture URI[6], and the `img` HTML tag contains an `alt` attribute which serves as a textual description for the picture.

In order to find alternatives and adapt this media item, one can use a search engine using the content of the `alt` attribute as a query. In our case, this attribute value is "Mona Lisa". A screenshot of the query result is shown in Fig. 2.

(a) Mona Lisa medium images.

(b) Mona Lisa small images.

Figure 2. Mona Lisa provided by Google.

In this context, the module a can be equated to the indexing process of search engines. The only difference is that a typical search engine gets media items according to a description given as a textual query, while a provides the descriptions according to a media item's URI (i.e., the opposite). However, reversing the index table is a trivial operation. We simulate this step by considering that the description associated to a picture URI is found in its corresponding `alt` attribute. Although the module b may not have equivalent in search engines, this only affect the accuracy of the description and consequently the result quality. Indeed, search engines return an ordered set of results based

[5] The picture URI is `http://www.inrialpes.fr/exmo/people/zimmer/smap2007/monalisa.png`.

[6] The web page is `http://www.inrialpes.fr/exmo/people/zimmer/smap2007/monalisa.html`.

on the similarity between the query and the media descriptions. This can be likened to module ⓒ. In this example, no aggregation is made, hence we directly compare the description in the `alt` attribute with descriptions of other images on the Web. This is simulated by submitting the query "Mona Lisa" to Google images search engine. Finally, results are filtered thanks to advanced search options or preferences as done in ⓓ. Our simulation takes advantage of Google's search options, such as "medium images".

As we may notice, in Fig. 2(a) the first answer is conveying the same message as the original one, i.e., both are the same paintings of Mona Lisa. However, for the same description and a different profile (small images), Fig. 2(b) provides unsatisfactory results because even if there exists a small Mona Lisa image, the first results are completely different from the initial Mona Lisa image.

Even though the World Wide Web is vast and diverse, currently descriptions about content available on the web are heterogeneous and partial. Moreover, in the description above, we focused on textual descriptions, while more and more annotations are using semantics. Considering these issues and to improve our adaptation mechanism, we propose in the next section to use Semantic Web technologies, such as RDF descriptions [12], ontologies and ontology matching and merging.

4. Media Adaptation Using the Semantic Web

When media items are annotated with semantic descriptions, the selection of the alternative media item can be far more accurate. In this section, we describe how our framework could be implemented with currently developed Semantic Web technologies. First of all, we assume that media descriptions are now represented as RDF descriptions [12] and use terminologies specified in OWL ontologies [13] as shown in Fig. 3.[7] In this example, d_1 and d_2 describes the initial media item, while d_3, d_4 and d_5 are descriptions of potential alternatives. Notice that the properties and concepts are defined in ontologies, hence it allows inferring knowledge that is not explicitly written in the description. For instance, the axiom $Painting \equiv \exists hasPainter$ in ontology O_1 means that if something has a painter, then it is a painting and vice versa. Now, d_2 stipulates that the initial media item was painted by "Da Vinci". Therefore, the image is a painting. Furthermore, $Cartoon \perp MasterPiece$ in O_2 signifies that $Cartoon$ and $MasterPiece$ are disjoint concepts. Thus, d_5 does not describe a cartoon.

In this context, module ⓐ corresponds to a Semantic Web search engine like Swoogle [5]. It retrieves all RDF descriptions relative to a media item. We notice that, as already mentioned in Sect. 3.1, there might be several descriptions of a media item that are rather weak all alone. So the next step, i.e., module ⓑ, consists in merging them, taking into account their semantics.

The RDF language offers a built-in merge operation, which merely consists in a set-theoretic union of RDF triples. In many cases, this notion of merging is not satisfactory. For instance, descriptions d_4 and d_5 refers to the same media item, and yet use a different terminology. Still, one can intuitively assume that $hasTitle$ and $title$ have the same meaning. Besides, other relations may exist between two ontologies, such as subsumption (e.g., a $Painting$ in O_1 is a kind of $Drawing$ in O_2). This is due to the fact that two ontologies, however different, may describe the same domain of knowledge. In order to reason simultaneously with descriptions from different ontologies, it is necessary to discover and define semantic relationships that exist between multiple ontologies. This activity is called *ontology matching* [6]. Matching ontologies results is an ontology alignment, i.e., a set of correspondences represented as dashed arrows in Fig. 3.[8] In our case, they are used to adequately merge individual descriptions. A basic yet efficient way to achieve this is to unify terms that have been proved equivalent by a distributed reasoner, and then apply a simple RDF-merge. In the general case, there are several possible ways to merge data, and a proper merging tool still has to be defined.

Computing similarities is also quite difficult when comparing heterogeneous semantic descriptions. However, examples of similarity measures for comparing ontology-based metadata can be found in [11]. Also, semantic similarities are defined for comparing terms or ontologies, for instance with the algorithm OLA [7]. These existing technologies can correctly implement module ⓒ. More precisely, the similarity between two individuals is usually computed by evaluating and combining the similarities of their attributes and properties. In the example of Fig. 3, the attribute $hasTitle$ of d_2 and d_4 can be directly compared and are quite different. This will decrease the similarity. Conversely, d_2 and d_4 both imply that the two pictures are paintings, which should increase the similarity. Indeed, according to ontology O_1, something that has a painter is a painting, and all portraits are paintings. Moreover, a correspondence asserts that $hasTitle$ is equivalent to $title$, so the two attributes can be compared. Therefore, d_2 and d_5 will be quite similar because of their titles. As a result, the merging of d_4 and d_5 will have a high overall similarity with the merged description $d_1 \oplus d_2$, whereas the similarity of d_3 with $d_1 \oplus d_2$ will be assessed as quite low because they do not have much in common. Consequently, the bottom right picture will most likely be preferred to the other one.

[7] The examples of RDF descriptions and OWL ontologies shown in Fig. 3 can be found at `http://www.inrialpes.fr/exmo/people/zimmer/smap2007/monalisa.html`.

[8] The alignment of Fig. 3 is defined in `http://www.inrialpes.fr/exmo/people/zimmer/smap2007/alignment.rdf`.

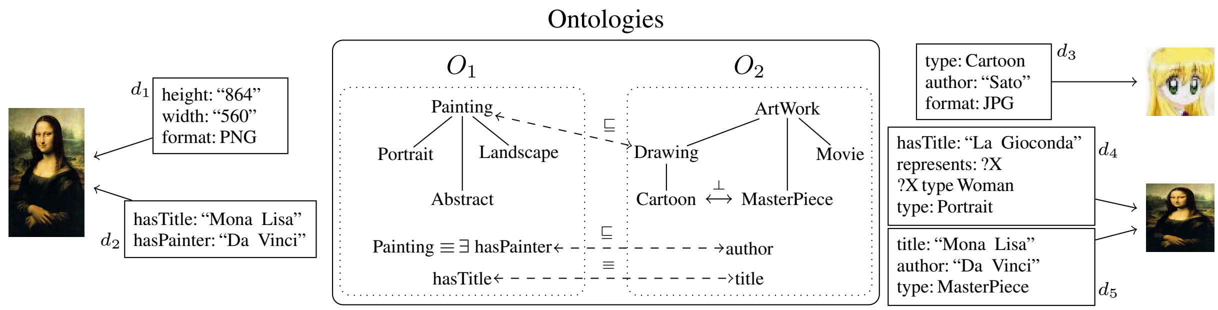

Figure 3. Media descriptions using Semantic Web technologies.

Finally, the filter module ⓓ can select the adapted media item not only according to its technical specification, but also to its semantic description. For instance, if the target has a content protection, only media items that are asserted to be non violent would be retained. Likewise, if the target's language is set to French, only media items in French will be selected. This again may require ontology matching techniques if the profile is not encoded with the same vocabulary as the media description.

To sum up, all the presented modules can be adapted to the Semantic Web using technologies that are either implemented or on the verge of being so. Nevertheless, the framework we have presented has some limitations, and raises several issues that we discuss in the next section.

5. Discussion

As we point out in Sect. 2, it is useful to find media item alternatives on the Web in some relevant circumstances. Nonetheless, it may happen that no appropriate alternative can be retrieved. For instance, it is disputable that a relevant replacement for personal data (e.g., family photographs), small events or uncommon objects would be found at all on the Web.

In such a case, a media transformation technique certainly leads to better results. However, more and more personal media items are now stored on publicly available web sites like Flickr[9] or Picasa[10] for pictures, Youtube[11], Dailymotion[12] or Metacafe[13] for videos, not to mention personal weblogs.

Another critical issue for which our method would probably not offer the best solution is the problem of multimedia document adaptation. In multimedia, the different media items composing a document are generally connected temporally, spatially, rhetorically or semantically. For that matter, other adaptation frameworks (like [9]) are better adapted. By contrast, replacing media items individually would most likely destroy the articulation of the author's discourse. This issue is therefore left to some further investigation.

Furthermore, it is still unclear how the Semantic Web version of our approach would perform concretely, because most of the works mentioned in this paper are still preliminary when it comes to implementation. Some of them do not even have a prototype, while others, though fully developed, are still lacking large scale test cases. Anyhow, it is expected that semantic-based retrieval techniques will operate at least as well as fully syntactic systems.

Ultimately, our last but not least self-criticism is about copyright issues. Indeed, while the author or sender of the original media item is responsible of choosing an adequate media, our fully automatic approach might carelessly select an authored item that should not be transmitted to any target individuals. Again, if copyright annotations are attached to candidate media items (e.g., Creative Commons), the filter module would easily discriminate copyrighted elements.

6. Related Work

A fair amount of research has been conducted on media item transformation and summarization: InfoPyramid [15] manages the different variations of media items with different modalities and fidelities; NAC [10] seeks to transform incompatible media items efficiently, thanks to predefined transcoding components; [1] use MPEG-21 resource adaptation tools; and [16] use web service compositions for media transformation.

Unfortunately, these systems change a specific format into another specific format. As a result, an implementation must be conducted for each format. Moreover, the computation costs of media transformation are considerable for large data such as long videos and would overload low capacity devices.

[9] `http://www.flickr.com`
[10] `http://picasa.google.com`
[11] `http://www.youtube.com`
[12] `http://www.dailymotion.com`
[13] `http://www.metacafe.com`

In order to avoid excessive response time, some multimedia description languages offer authors with the capability of specifying explicit alternatives (e.g., [4, 14]). However, doing so is rather cumbersome and must be conducted for each conceivable execution profile. Additionally, it cannot take advantage of a dynamically evolving network like the Web, e.g., if a referenced media item moves, it will not be accessible anymore, and the alternative will not work.

Similarly to our approach, the work described in [17] uses the World Wide Web to select translation of text from Chinese to English and vice versa. More precisely, they use web-based search engines to automatically find probable translations by looking for the most frequent co-occurrences of terms. Nevertheless, it only translates words or small phrases, while one could need a summarized text or a new formulation of the text, which could be captured by a semantic annotation.

7. Conclusion

In this paper, we propose to use the World Wide Web diversity and profusion in order to adapt media items. We use Semantic Web technologies in order to deal with semantic gaps, heterogeneity and partial descriptions, thus improving this adaptation framework.

We showed that a naive yet efficient and sometimes effective implementation can already be carried out with current Web-based machinery. We also envisioned how current Semantic Web technologies nearly accomplish with accuracy the task that we motivated here.

As a future development, this framework could be implemented as a Web service, while environment constraints (profiles) may be defined using CC/PP [8], ontologies based on OWL and description written in RDF. Another interesting prospect that we envisage is the following. Since this framework is used to discover alternative media items according to the proximity of their description to the original media description, it is also possible to broadcast several alternatives that are sufficiently close to an initial description, *even when there is no initial media item*. This can be paralleled with the act of automatically generating a document according to a description. Lastly, this framework could be integrated into a more general multimedia adaptation system mixing semantic media adaptation with structural and compositional adaptation.

References

[1] M. K. Asadi and J.-C. Dufourd. Knowledge-based and semantic adaptation of multimedia content. In *Knowledge-Based Media Analysis for Self-Adaptive and Agile Multimedia Technology*, pages 285–293, 2004.

[2] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.

[3] A. Z. Broder, S. C. Glassman, M. S. Manasse, and G. Zweig. Syntactic clustering of the web. *Computer Networks and ISDN Systems*, 29(8-13):1157–1166, 1997.

[4] D. Bulterman, G. Grassel, J. Jansen, A. Koivisto, N. Layaïda, T. Michel, S. Mullender, and D. Zucker. *Synchronized Multimedia Integration Language (SMIL 2.1)*. W3C, 2005. http://www.w3.org/TR/SMIL2/.

[5] L. Ding, R. Pan, T. Finin, A. Joshi, Y. Peng, and P. Kolari. Finding and ranking knowledge on the semantic web. In *Proc. of the 4th International Semantic Web Conference*, LNCS 3729, pages 156–170. Springer, Nov. 2005.

[6] J. Euzenat and P. Shvaiko. *Ontology matching*. Springer-Verlag, Heidelberg (DE), 2007.

[7] J. Euzenat and P. Valtchev. Similarity-based ontology alignment in owl-lite. In R. López de Mantaras and L. Saitta, editors, *Proc. 16th European Conference on Artificial Intelligence*, pages 333–337. IOS press, 2004.

[8] G. Klyne, F. Reynolds, C. Woodrow, H. Ohto, J. Hjelm, M. H. Butler, and L. Tran. *Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies 1.0*. W3C, 2001. http://www.w3.org/TR/CCPP-struct-vocab/.

[9] S. Laborie, J. Euzenat, and N. Layaïda. Multimedia document summarization based on a semantic adaptation framework. In *Proc. of the International Workshop on Semantically Aware Document Processing and Indexing*, pages 87–94. ACM Press, 2007.

[10] T. Lemlouma and N. Layaïda. The negotiation of multimedia content services in heterogeneous environments. In *Proc. of the 8th International Conference on Multimedia Modeling*, pages 187–206, Amsterdam, 2001.

[11] A. Maedche and V. Zacharias. Clustering ontology-based metadata in the semantic web. In *Proc. of the 6th European Conference on Principles of Data Mining and Knowledge Discovery*, pages 348–360, London, 2002. Springer-Verlag.

[12] F. Manola and E. Miller. *RDF Primer*. W3C, 2004. http://www.w3.org/RDF/.

[13] D. L. McGuinness and F. van Harmelen. *OWL Web Ontology Language Overview*. W3C, 2004. http://www.w3.org/TR/owl-features/.

[14] F. Rousseau, J. A. García-Marcías, J. V. de Lima, and A. Duda. User adaptable multimedia presentations for the world wide web. In *Proc. of the 8th international conference on World Wide Web*, pages 1273–1290. Elsevier, 1999.

[15] J. R. Smith, R. Mohan, and C.-S. Li. Scalable multimedia delivery for pervasive computing. In *Proc. of the 7th ACM International Conference on Multimedia (Part 1)*, pages 131–140, New York, 1999. ACM Press.

[16] P. Soetens, M. D. Geyter, and S. Decneut. Multi-step media adaptation with semantic web services. In *Proc. of the 3rd International Semantic Web conference*, Hiroshima, 2004.

[17] Y. Zhang and P. Vines. Using the web for automated translation extraction in cross-language information retrieval. In *Proc. of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 162–169. ACM Press, 2004.

[18] P. Ziegler and K. R. Dittrich. Three decades of data integration - all problems solved? In *IFIP Congress Topical Sessions*, pages 3–12. Kluwer, 2004.