HAL
open science

Semantic multimedia document adaptation with functional annotations

Sébastien Laborie, Jérôme Euzenat, Nabil Layaïda

To cite this version:

Sébastien Laborie, Jérôme Euzenat, Nabil Layaïda. Semantic multimedia document adaptation with functional annotations. Proc. 4th international workshop on Semantic Media Adaptation and Personalization (SMAP2009), Dec 2009, San Sebastián, Spain. pp.44-49, 10.1109/SMAP.2009.19 . hal-00794919

HAL Id: hal-00794919

https://inria.hal.science/hal-00794919

Submitted on 26 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic Multimedia Document Adaptation with Functional Annotations

Sébastien Laborie
IRIT – Paul Sabatier University
Sebastien.Laborie@irit.fr

Jérôme Euzenat
INRIA Grenoble – Rhône-Alpes
Jerome.Euzenat@inrialpes.fr

Nabil Layaïda
INRIA Grenoble – Rhône-Alpes
Nabil.Layaida@inrialpes.fr

Abstract

The diversity of presentation contexts for multimedia documents requires the adaptation of document specifications. In an earlier work, we have proposed a semantic adaptation framework for multimedia documents. This framework captures the semantics of the document composition and transforms the relations between multimedia objects according to adaptation constraints. In this paper, we show that relying on document composition alone for adaptation restricts the set of relevant candidate solutions and may even divert the adaptation from the authors intent. Hence, we propose to introduce functional annotations to guide the adaptation process. Theses annotations allow to refine the role of multimedia objects in the document. We show that SMIL documents could embed functional annotations encoded in RDF. These multimedia documents are then adapted thanks to an interactive adaptation tool.

1 Introduction

Ideally, a multimedia document is authored once but can be played on different devices with different capabilities: phones, PDAs, desktop computers, set-top boxes, etc. These devices introduce different constraints on the presentation itself. For instance, CPU or display limitations (e.g., mobile phones) may prevent multimedia objects from being displayed correctly. Other constraints may also be introduced by user preferences, content protection issues or terminal capabilities [10]. The set of constraints related to a client is called a profile.

To satisfy profiles, multimedia documents must be adapted, i.e., transformed into documents compatible with target contexts before being played. We need to distinguish two types of adaptation: local adaptation (adaptation of multimedia objects individually) and global adaptation (adaptation of the document structure). This paper focuses on the latter.

In [5], a framework for adapting multimedia documents based on the qualitative semantics of the documents and constraints was proposed. This approach transforms the relations between multimedia objects and ensures two main properties: (1) all constraints specified in a target client profile are satisfied and (2) the computed adapted document is as close as possible to the initial document. We have shown, e.g., in [11, 12], that this framework can adapt the spatio-temporal and the hypermedia dimensions of standard multimedia documents such as SMIL documents [4].

However, this adaptation framework only deals with the initial document composition, i.e., explicit relations between multimedia objects specified by authors. In this paper, we show that relying only on this information for adapting documents is unsatisfactory because it leads to a limited form of adaptation. Hence, we propose to guide adaptation with functional annotations, i.e., annotations related to multimedia objects which express a function in the document, such as a title, a legend. Consequently, additional implicit relations between multimedia objects can be identified and these could enrich the adaptation process for producing new adapted solutions close to the initial document composition and the author intent.

Furthermore, we show that SMIL documents could embed functional annotations encoded in RDF [17]. From such multimedia documents, our interactive adaptation prototype provides relevant adaptations.

In Section 2, we present a general adaptation scheme. In Section 3, we define functional annotations and show how to integrate them into the document specification. In Section 4, we adapt a multimedia document example and compare adaptation results with and without functional annotations. We show that with these annotations, adaptation is more flexible. Finally, in Section 5, we adapt SMIL documents with functional annotations encoded in RDF.

2 Multimedia Document Adaptation

In order to execute a multimedia document on different devices with different profiles, a fair amount of research has been conducted on multimedia document adaptation. We propose to group all techniques into three categories[1]:

- **Specification of alternatives:** During a multimedia document edition, an author may specify different versions based on some target profiles, e.g., thanks to the `switch` tag of SMIL [4]. However, when editing a document, it is cumbersome to specify all possible versions for every existing profiles.
- **Using transformation rules:** Inside an adaptation system, a set of transformations is defined a priori (e.g., speech to text, parallel to sequential) and stored into a repository. From a multimedia document and a target device profile, some transformations are selected by the system and applied on the document. [14] and [2] are two examples of such a system.
- **Multimedia document generation:** Other approaches for multimedia document adaptation are based on specialized specifications for generating documents, such as [8] and [3]. However, this requires the content to be initially described with these specifications instead of adapting existing documents.

Apart from the existing adaptation mechanisms, there are several languages or formats for specifying multimedia documents that have to be adapted, such as SMIL [4], Madeus [9] and XHTML [19]. Making the adaptation format-dependent requires an adaptation strategy for every single format. Instead, we propose an abstraction layer which allows to hide the format specific syntax and details, and capture the essence of a given document with respect to its main dimensions. We call this document abstraction a multimedia document specification and presents in Figure 1 a generic adaptation scheme for such representation.

Definition 1 (Multimedia document specification) *A multimedia document specification $s = \langle O, C \rangle$ is made of a set of multimedia objects O and a set of relations (or constraints) C between these objects.*

In the remainder, the set O will contain at least two multimedia objects and the relations of C will be considered as binary. As illustrated in Figure 1, from an initial multimedia document we identify all its involved objects (i.e., texts, images, audios, videos, hypermedia links) and all relations between them. These relations may come from the document composition (ⓐ) and functional annotations (ⓑ). The latter will be defined in Section 3.

[1] Other adaptation categories have been proposed, e.g., in [13].

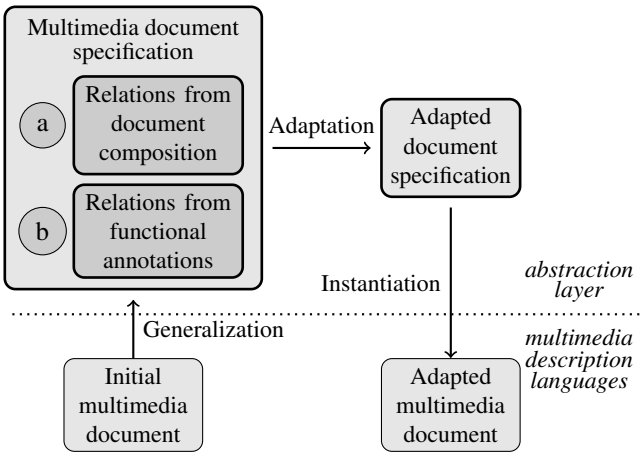


Figure 1. A generic multimedia document adaptation scheme.

Thereafter, the multimedia document specification is adapted according to target profiles. For that purpose, forbidden relations deduced from target profiles are transformed into valid ones. In the remainder, we will use the adaptation framework presented in [5] because it considers multimedia document specification and provides adapted documents which are as close as possible from the initial document.

Finally, an adapted document satisfying the adapted specification is instantiated. This one could be encoded as the initial multimedia document description language or another one.

The advantages of such adaptation scheme presented in Figure 1 are twofold. First, it permits to reuse the same adaptation strategy for different formats. Second, the abstract representation provides more flexibility for adaptation since the relations between multimedia objects can be described qualitatively.

In the following, we show that functional annotations enrich the adaptation process for producing new adapted solutions close to the initial document composition and the author intent.

3 Enhanced Multimedia Document Specifications with Functional Annotations

A multimedia document specification is composed of relations holding between objects (cf., Definition 1). Functional annotations may complete these relations, they are annotations related to objects which express a function in the document. We identified two categories of functional annotations: one concerns media individually and the other one concerns groups of objects.

The former expresses a specific role of a media involved in a multimedia presentation. For instance, one media might be a title, a menu, a logo or an advertisement. Some multimedia description standards, e.g., the multimedia description schemes of MPEG-7 [18], allow to specify such roles.

The latter expresses semantic relationships between a group of objects and describes how several objects relate in a narrative or story. For instance, these semantic relationships may describe relations that draw upon typical pragmatic relations such as an object being the agent, patient, cause, goal of other objects. The rhetorical structure theory [16] offers some possible semantic relationships between objects. These have been used, e.g., in [20], for generating multimedia presentations.

During a multimedia document edition, objects are usually placed according to these functional annotations. For example, a title may stay on top of a slide, a legend may stay around the object it describes. From particular functional annotations, [15] and [7] identify mappings to qualitative spatio-temporal relations. Consequently, from a multimedia document specification the set of relations could contain both explicit spatio-temporal relations made by an author and implicit spatio-temporal relations deduced from functional annotations. In the following, we show how to embed functional annotation information into a multimedia document specification.

Suppose the multimedia presentation example illustrated in Figure 2. This one is composed of four objects: two texts T_1 and T_2, one video V and an image I.

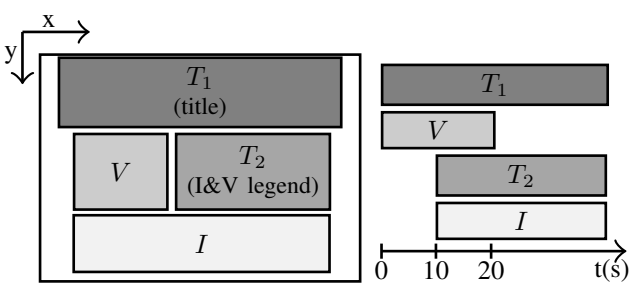
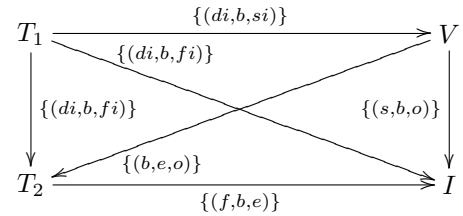


Figure 2. A multimedia presentation example.

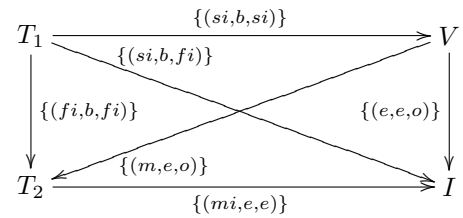
The multimedia document specification corresponding to this example may be represented with a relation graph where each node refers to an object of the presentation and each edge is labeled by some spatio-temporal relations. Figure 3 presents three relation graphs related to the multimedia document example of Figure 2.

In this example, the Allen algebra [1] (cf., Table 1) is used to describe spatio-temporal relations. More precisely, each spatio-temporal relation forms a triple (r_x, r_y, r_t) such that r_x, r_y and r_t are Allen's relations which correspond respectively to projections over the x, y and t axis.

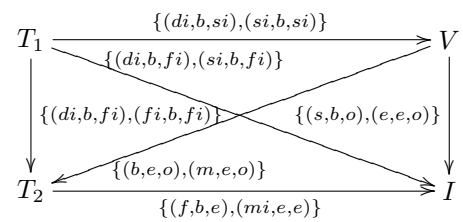
Figure 3(a) presents the explicit spatio-temporal relations related to the document composition of Figure 2. From functional annotations, additional implicit spatio-temporal relations are identified in Figure 3(b). For example, T_1 is a title and should stay on top of the other objects and played simultaneously (e.g., $T_1\{(si, b, fi)\}I$), T_2 is the legend of I and V, and should stay around them (e.g., $T_2\{(mi, e, e)\}I$). It may happen that no functional annotations are provided, e.g., between I and V.



(a) Explicit relations from the document composition of Figure 2.



(b) Implicit relations from the functional annotations of Figure 2.



(c) The multimedia document specification of Figure 2 with explicit and implicit relations.

Figure 3. Multimedia document specifications represented with relation graphs.

relation (r): x r y	x / y	inverse: y r^{-1} x
before (b)		(bi) after
meets (m)		(mi) met-by
during (d)		(di) contains
overlaps (o)		(oi) overlapped-by
starts (s)		(si) started-by
finishes (f)		(fi) finished-by
equals (e)		(e)

Table 1. The thirteen Allen's relation.

To maintain complete information in Figure 3(b), we compute the transitive closure of spatio-temporal relations.

For example, $V\{(m,e,o)\}T_2$ and $T_2\{(mi,e,e)\}I$ implies $V\{(e,e,o)\}I$. For visibility reasons, we only illustrate in Figure 3(b) one spatio-temporal relation in each set which differs from those of Figure 3(a).

Figure 3(c) corresponds to the global multimedia document specification of Figure 2. It combines the explicit and implicit information provided respectively by Figure 3(a) and Figure 3(b). In the next section, the relation graph illustrated in Figure 3(c) will be adapted in order to show the benefits of using functional annotations.

4 Adaptation of Multimedia Document Specifications

We want to execute the multimedia presentation example illustrated in Figure 2 on a mobile phone. However, its profile indicates that it is impossible to play more than three multimedia objects concurrently. The multimedia presentation presented in Figure 2 does not comply with this target profile because from $t = 10s$ to $t = 20s$ four objects are executed simultaneously. We thus have to adapt the document, i.e., transform forbidden relations into valid ones and compute an adapted document as close as possible to the initial document and author intent.

In this context, adapting consists of finding a set of relation graphs satisfying the profile at a minimal distance from the initial document specification. We consider that the proximity between two relation graphs depends on the proximity between relations borne by the same edge in both graphs. This proximity relies on the conceptual neighborhood between these relations and is measured by the shortest path distance in the corresponding conceptual neighborhood graph (Figure 4, [6]).

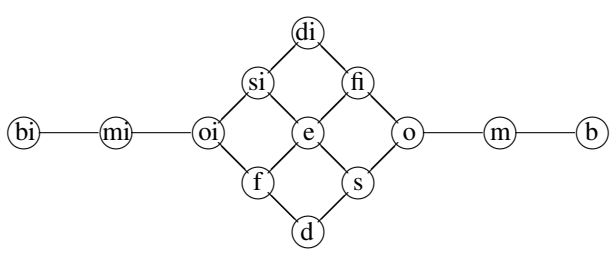
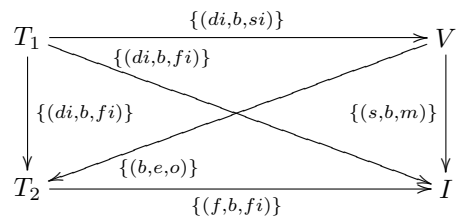


Figure 4. A conceptual neighborhood graph.

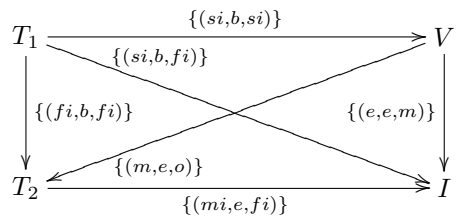
Thus, a distance d between relation graphs is obtained by summing up all the conceptual distances δ between relationships used in both graphs (cf., Definition 2).

Definition 2 *Let G and G' be two relation graphs with a set of nodes N and a labeling function λ (resp., λ') which returns a set of relations between two nodes.* $d(G,G') = \Sigma_{n_1,n_2 \in N} \mathrm{Min}_{r \in \lambda(\langle n_1,n_2 \rangle), r' \in \lambda'(\langle n_1,n_2 \rangle)} \delta(r_x, r'_x) + \delta(r_y, r'_y) + \delta(r_t, r'_t)$.

Figure 5 presents two adapted relation graphs satisfying the mobile phone profile. The former (Figure 5(a)) considers only the explicit relations presented in Figure 3(a) which were extracted from Figure 2. The latter (Figure 5(b)) considers the explicit and implicit relations presented in Figure 3(c). Their distance from their corresponding initial specification are both $d = 2$ because the temporal relation r_t between V and I changes from o to m ($\delta(o,m) = 1$ in Figure 4) and the one between T_2 and I changes from e to fi ($\delta(e,fi) = 1$ in Figure 4).



(a) Adaptation of the relation graph illustrated in Figure 3(a) without considering functional annotations.



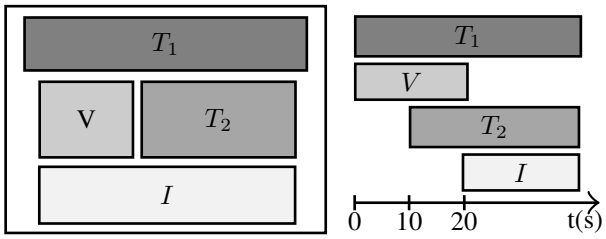
(b) Adaptation of the relation graph illustrated in Figure 3(c) which considers explicit and implicit relations.

Figure 5. Adapted relation graphs which satisfy the mobile phone profile.

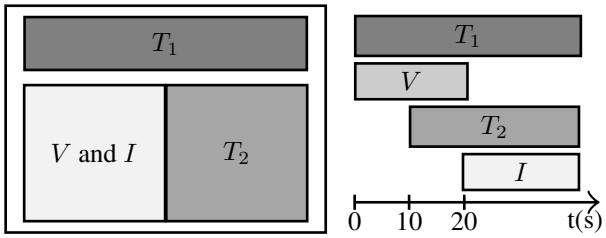
Figure 6 presents two adapted executions of the adapted relation graphs presented in Figure 5. Figure 6(a) presents an execution without considering functional annotations. This execution satisfies the mobile phone profile and is close to the initial document.

However, the solution proposed is not suitable because at $t = 20s$ (i.e., when V disappears and I appears) spatial information is not well presented. This is due to the explicit relations illustrated in Figure 3(a) which are too restrictive because it strictly interprets the document composition constraints and does not reflects functional annotations.

Figure 6(b) presents an execution when taking into account functional annotations, especially their implicit spatio-temporal relations. This execution satisfies the profile and is as close as possible to the initial document composition and the author intent. Furthermore, thanks to the flexibility added by these annotations the previous display problem did not appear in this adapted solution.



(a) An execution corresponding to Figure 5(a).



(b) An execution corresponding to Figure 5(b).

Figure 6. Executions corresponding to the relation graphs presented in Figure 5.

5 Implementation

This adaptation framework has been implemented in an interactive adaptation prototype. Its architecture is based on Figure 1 and uses the *Adaptation* algorithm (cf., Algorithm 1) to compute adapted solutions.

The user of our system may edit a SMIL document containing functional annotations encoded in RDF [17]. The SMIL Metainformation module [4] is used to specify such annotations in RDF. Figure 7 presents such a multimedia document, this one corresponds to the presentation illustrated in Figure 2.

From such document, explicit and implicit relations relative to the initial document specification, and possible relations satisfying the target device profile are encoded into two matrix $I_{i,j}$ and $P_{i,j}$. Before the call to $Adaptation(I_{i,j}, P_{i,j})$, we sort each label of the matrix $P_{i,j}$ according to the distance δ (cf., Section 4) from each label of the matrix $I_{i,j}$. Thereafter, Algorithm 1 computes consistent relation graphs with help from the *pathConsistency* function defined in [1]. The closest relation graphs from the initial document specification are selected in $\mathcal{S}$. When Algorithm 1 stops, $\mathcal{S}$ contains adapted relation graph solutions.

Of course, our implementation could adapt multimedia documents on the client side or on the server side without any GUI. Our interactive prototype is useful for guiding an author during a multimedia document edition under constraints.

Algorithm 1: Adaptation

Input: A matrix $I_{i,j}$ corresponding to the initial multimedia document specification and a matrix $P_{i,j}$ corresponding to possible relations.
Data: $\mathcal{S}$ is a set of adapted relation graphs; Min is a current minimum computed distance (initialized with a high value).

```
pathConsistency(P);
if P does not contain an empty relation then
    Choose an unprocessed label P_{i,j} and split P_{i,j}
    into r_l = r_1, ..., r_k;
    if no label can be split then
        tmp ← d(I, P);
        if tmp < Min then
            Min ← tmp; S ← {P};
        if tmp = Min then
            S ← S ∪ {P};
    else
        for all label r_l (1 ≤ l ≤ k) do
            P_{i,j} ← r_l;
            if d(I, P) ≤ Min then
                Adaptation(I,P);
```

```
<?xml version="1.0" ?>
<!DOCTYPE smil PUBLIC "-//W3C//DTD SMIL 2.0//EN"
      "http://www.w3.org/2001/SMIL20/SMIL20.dtd"
     [<!ENTITY funannot "http://www.example.com/funannot#">]>
<smil xmlns="http://www.w3.org/2001/SMIL20/Language"
  xmlns:funannot="http://www.example.com/funannot#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#">
 <head>
  <metadata id="meta-rdf">
   <rdf:RDF>
    <rdf:Description rdf:about="#T1" rdf:type="&funannot;Title"/>
    <rdf:Description rdf:about="#T2">
     <funannot:isLegendOf rdf:resource="#V"/>
     <funannot:isLegendOf rdf:resource="#I"/>
    </rdf:Description>
   </rdf:RDF>
  </metadata>
  <layout>
   <root-layout height="250" width="200"/>
   <region id="rT1" height="40" width="200" left="0" top="0"/>
   <region id="rT2" height="100" width="115" left="70" top="55"/>
   <region id="rV" height="100" width="55" left="10" top="55"/>
   <region id="rI" height="75" width="175" left="10" top="170"/>
  </layout>
 </head>
 <body>
  <par>
   <text id="T1" region="rT1" src="Introduction.rt"/>
   <text id="T2" region="rT2" src="Information.rt" begin="10s"/>
   <video id="V" region="rV" src="ResearchTeam.avi" dur="10s"/>
   <img id="I" region="rI" src="Lab.jpg" begin="10s"/>
  </par>
 </body>
</smil>
```

Figure 7. The SMIL code of Figure 2 with functional annotations encoded in RDF.

Actually, one may specify some constraints (i.e., a profile) inside our prototype and if an author creates a forbidden relation between two objects the system provides other possible configurations close to the initial invalid one which satisfy the given profile. Hence, a user can be assisted, for instance, during the edition of a mobile multimedia document.

6 Conclusion

We have defined an adaptation framework which uses functional annotations in order to guide adaptation and make it more flexible. This adaptation approach is generic, i.e., it may adapt several document dimensions and may be applied to any multimedia description language. In order to validate the framework, we have implemented a prototype which adapts SMIL documents containing functional annotations encoded in RDF.

In the future, we plan to identify automatically, from the document composition, functional annotations (e.g., using emergent semantics as proposed in [21]) and select particular privileged adapted solutions, thus refining our adaptation approach.

References

[1] J. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.

[2] M. K. Asadi and J.-C. Dufourd. Knowledge-based and semantic adaptation of multimedia content. In P. Hobson, E. Izquierdo, Y. Kompatsiaris, and N. E. O'Connor, editors, *Knowledge-Based Media Analysis for Self-Adaptive and Agile Multimedia Technology*, pages 285–293, 2004.

[3] S. Boll. MM4U - a framework for creating personalized multimedia content. In *Proceeding of the International Conference on Distributed Multimedia Systems*, 2003.

[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). Recommendation, W3C, 2005. `http://www.w3.org/TR/SMIL/`.

[5] J. Euzenat, N. Layaïda, and V. Dias. A semantic framework for multimedia document adaptation. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 31–36, 2003.

[6] C. Freksa. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54(1–2):199–227, 1992.

[7] L. Gaillard, J. Nanard, B. Bachimont, and L. Chaming's. Intentions based authoring process from audiovisual resources. In *Proceedings of the International Workshop on Semantically Aware Document Processing and Indexing*, pages 21–30, 2007.

[8] J. Geurts, J. van Ossenbruggen, and L. Hardman. Application-specific constraints for multimedia presentation generation. In *Proceedings of the International Conference on Multimedia Modeling (MMM)*, pages 247–266, 2001.

[9] M. Jourdan, N. Layaïda, C. Roisin, L. Sabry-Ismaïl, and L. Tardif. Madeus, an authoring environment for interactive multimedia documents. In *Proceedings of the 6th ACM Multimedia conference*, pages 267–272, 1998.

[10] 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. Recommendation, W3C, 2001. `http://www.w3.org/TR/CCPP-struct-vocab/`.

[11] S. Laborie. Spatio-temporal proximities for multimedia document adaptation. In *Proceedings of The Twelfth International Conference on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA)*, pages 128–137, Varna (Bulgaria), September 2006. Springer, LNAI 4183.

[12] S. Laborie and J. Euzenat. An incremental framework for adapting the hypermedia structure of multimedia documents. In M. Wallace, M. Angelides, and P. Mylonas, editors, *Advances in Semantic Media Adaptation and Personalization*, volume 93 of *Studies in Computational Intelligence Series*, chapter 8, pages 157–176. Springer, 2008.

[13] Z. Lei and N. D. Georganas. Context-based media adaptation in pervasive computing. In *Proceedings of Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 913–918, 2001.

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

[15] S. Little, J. Geurts, and J. Hunter. Dynamic generation of intelligent multimedia presentations through semantic inferencing. In *Proceedings of the 6th European Conference on Research and Advanced Technology for Digital Libraries*, pages 158–175, 2002.

[16] W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: Towards a functional theory of text organisation. *Text*, 8(3):243–281, 1988.

[17] F. Manola and E. Miller. RDF primer. Recommendation, W3C, 2004. `http://www.w3.org/TR/rdf-primer/`.

[18] MPEG MDS Group. *Text of ISO/IEC 15938-5 FDIS Information Technology - Multimedia Content Description Interface - Part 5 Multimedia Description Schemes*, 2001. ISO/IEC JTC1/SC29/WG11 MPEG01/N4242.

[19] S. Pemberton, D. Austin, J. Axelsson, T. Çelik, D. Dominiak, H. Elenbaas, B. Epperson, M. Ishikawa, S. Matsui, S. McCarron, A. Navarro, S. Peruvemba, R. Relyea, S. Schnitzenbaumer, and P. Stark. XHTML™ 1.0 the extensible hypertext markup language (second edition). Recommendation, W3C, 2000. `http://www.w3.org/TR/xhtml1/`.

[20] L. Rutledge, B. Bailey, J. van Ossenbruggen, L. Hardman, and J. Geurts. Generating presentation constraints from rhetorical structure. In *Proceedings of the 11th ACM on Hypertext and hypermedia*, pages 19–28. ACM Press, 2000.

[21] A. Scherp, S. Boll, and H. Cremer. Emergent semantics in personalized multimedia content. In *Proceedings of the 4th special workshop on Multimedia Semantics*, pages 53–61, 2006.