

Semantic adaptation of multimedia documents

Sébastien Laborie, Jérôme Euzenat, Nabil Layaïda

▶ **To cite this version:**

Sébastien Laborie, Jérôme Euzenat, Nabil Layaïda. Semantic adaptation of multimedia documents. Multimedia Tools and Applications, 2011, 55 (3), pp.379-398. 10.1007/s11042-010-0552-9 . hal-00781024

HAL Id: hal-00781024
https://inria.hal.science/hal-00781024

Submitted on 25 Jan 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.

Multimedia tools and applications 55(3):379-398, 2011
To be published

Semantic Adaptation of Multimedia Documents

Sébastien Laborie · Jérôme Euzenat · Nabil Layaïda

Submission date: first version November 10th 2009 and revised version March 26th 2010

Abstract Multimedia documents have to be played on multiple device types. Hence, usage and platform diversity requires document adaptation according to execution contexts, not generally predictable at design time. In an earlier work, a semantic framework for multimedia document adaptation was proposed. In this framework, a multimedia document is interpreted as a set of potential executions corresponding to the author specification. To each target device corresponds a set of possible executions complying with the device constraints. In this context, adapting requires to select an execution that satisfies the target device constraints and which is as close as possible from the initial composition. This theoretical adaptation framework does not specifically consider the main multimedia document dimensions, i.e., temporal, spatial and hypermedia. In this paper, we propose a concrete application of this framework on standard multimedia documents. For that purpose, we first define an abstract structure that captures the spatio-temporal and hypermedia dimensions of multimedia documents, and we develop an adaptation algorithm which transforms in a minimal way such a structure according to device constraints. Then, we show how this can be used for adapting concrete multimedia documents in SMIL through converting the documents in the abstract structure, using the adaptation algorithm, and converting it back in SMIL. This can be used for other document formats without modifying the adaptation algorithm.

Keywords Multimedia document transformation, qualitative representation and reasoning, SMIL.

Sébastien Laborie
IRIT – Université Paul Sabatier, 118 Route de Narbonne 31062 Toulouse Cedex 9, France. Tel.:+33561556899. E-mail: Sebastien.Laborie@irit.fr.

Jérôme Euzenat · Nabil Layaïda
INRIA Grenoble Rhône-Alpes & LIG, 655 Avenue de l'Europe 38130 Montbonnot Saint-Martin, France. Tel.: +33476615200. E-mail: {Jerome.Euzenat, Nabil.Layaida}@inrialpes.fr.

1 Introduction

In the near future, mobile access to the Web will surpass personal computer access [40]. This includes access to multimedia content from mobile phones as well as other emerging environments, such as set-top boxes, PDAs, tablet PCs, interactive television, and even automobiles. However, mobile Web access today still lacks the necessary flexibility needed for the usability of multimedia content on such a variety of devices [39].

To tackle this problem there are two possible approaches: either define a device independent multimedia document format, e.g., [19], from which one can generate adapted content encoded in a specific format or define a language, e.g., [4], in which variations of a document can be expressed. The former approach would encode content with this new format but ignore legacy content already found on the web today. The latter requires to anticipate all possible contexts in which a document will have to be played. First, it requires at multimedia content design time the knowledge of all the target devices characteristics on which the content would be rendered in the future. Second, it requires designers to specify a content which would gracefully and seamlessly render correctly on all existing and future devices. In practice, designers produce a very limited set of such adapted content [31]: usually, one for desktop PCs and a second one for mobile terminals. The reason behind this is that they are more comfortable with specific designs tied to particular screen sizes and terminal characteristics. When they create content, they do generally engage in a cyclic process where they specify and check the result directly on a specific terminal. With such a process, it becomes hard to address correctly a richer set of terminals. Furthermore, multimedia documents formats are now defined as a progressively richer sets of sub-languages with increasing features. For example, the SMIL language has several versions [4]: basic, tiny, mobile, advanced mobile and full together with a scalability framework. This will likely make any adaptive design more complex if not impossible without automatic processing.

In an earlier work, we proposed a semantic framework for multimedia document adaptation in which a multimedia document is interpreted as a set of potential executions [11]. This framework defines adaptations as the search for a document compliant with the device constraints which is as close as possible to the executions specified by the author. It does it in a very general way, independent from the concrete multimedia document languages and independent from the main multimedia document dimensions, i.e., temporal, spatial and hypermedia.

In this paper, we show how this approach can be used for adapting the structure of multimedia documents in concrete multimedia description languages without having anticipated the type of constraints imposed by the devices. For that purpose, we define a structure that captures the spatio-temporal and hypermedia dimensions of multimedia documents. We have developed an adaptation algorithm which transforms in a minimal way such a structure according to device constraints.

Finally, we consider adaptation as a three-step process: (1) abstracting the original content format, followed by (2) adapting that abstract representation to the device profile, and (3) producing an adapted version in the same format as the input document (see Figure 1). This approach has many advantages: It allows for defining a device independent format in the form of some abstract relations

between multimedia document objects; as new format appears, it can be used by only defining rules for extracting the abstract format from the concrete one and vice-versa.

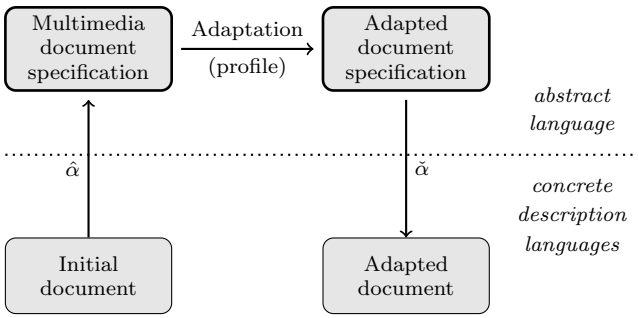


Fig. 1 Multimedia document adaptation strategy: a document in a concrete language is first ($\hat{\alpha}$) expressed as an abstract specification on which adaptation is performed; then, the result is transformed back in the initial format ($\check{\alpha}$).

The remainder of the paper is structured as follows. Section 2 presents a multimedia document example. In Section 3, we review existing adaptation techniques and how they are used to adapt multimedia documents. In particular, we focus on a semantic adaptation framework proposed in [11] and how it compares to the other approaches. In Section 4, we extend this framework in order to capture the spatio-temporal and hypermedia dimensions of multimedia documents. In Section 5, we show how to adapt documents semantically in a minimal way according to device profiles. For that purpose, an adaptation algorithm is also described. Finally, in Section 6 we present a prototype for SMIL documents.

2 A Multimedia Document Example

A multimedia document is an entity that combines pieces of information which come from various media types named multimedia objects (as known as media items) [2,33]. Typical examples are web documents including synchronized video or audio. Figure 2 represents a multimedia document example played on a laptop. This multimedia document is composed of four images of Athens city, namely *Acropolis*, *Agora*, *Temple* and *Museum*[1].

Multimedia documents can be modeled through three main dimensions:

- **Temporal:** Multimedia objects are synchronized temporally, e.g., some objects are played sequentially while others are played simultaneously.
- **Spatial:** Some multimedia objects are organized spatially and displayed in a graphic layout, e.g., two side by side images.
- **Hypermedia:** Links allow for interacting with the presentation and for navigating through its different parts.

In Figure 2, multimedia objects are synchronized temporally, along the presentation timeline, and organized spatially on the screen. They embed two hypermedia links, namely l_1 and l_2 (see the black thick solid lines in Figure 2). For instance, the images named *Acropolis* and *Agora* are presented from $t = 0s$ to $t = 10s$. These images have identical sizes, and are displayed side by side.

Moreover, both images embed one hypermedia link which points to the beginning of the presentation of the two other images, namely *Temple* and *Museum*. The hypermedia links may have specific durations corresponding to the period of time when they are clickable or active. Actually, if a user selects l_1 or l_2, time jumps to $t = 10s$.

This sample multimedia document may be played on other devices, such as PDAs or mobile phones. Figure 3 shows how it is rendered on a PDA. Notice that images are not displayed correctly because the PDA has a smaller display compared to a laptop. This illustrates the need to perform adaptation in order to view this document correctly.

In the following section, we present an overview of current adaptation approaches and compare them with a semantic adaptation framework which was first proposed in [11].

[1] During the presentation, a narrative audio content commenting the images is played in parallel, however to avoid visual overloading we do not consider it in this paper.

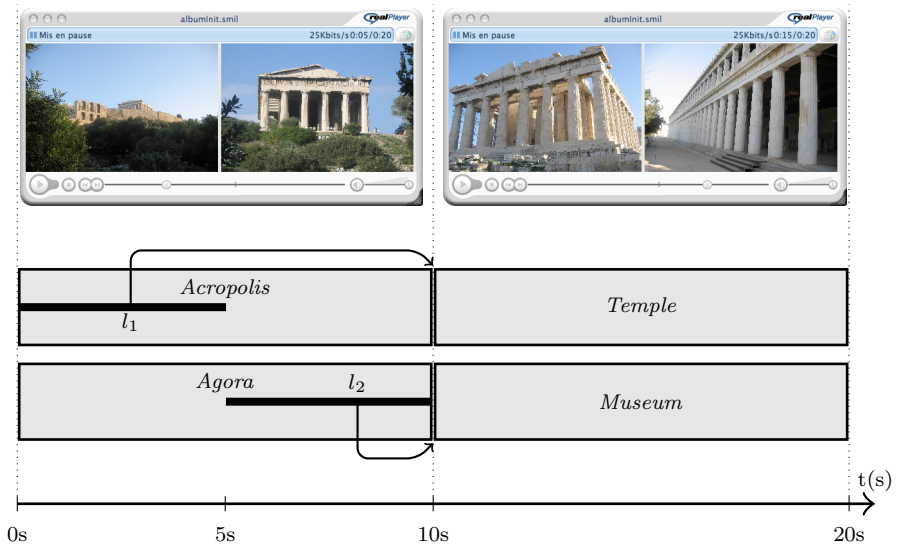


Fig. 2 A multimedia document played with Real Player on a laptop.

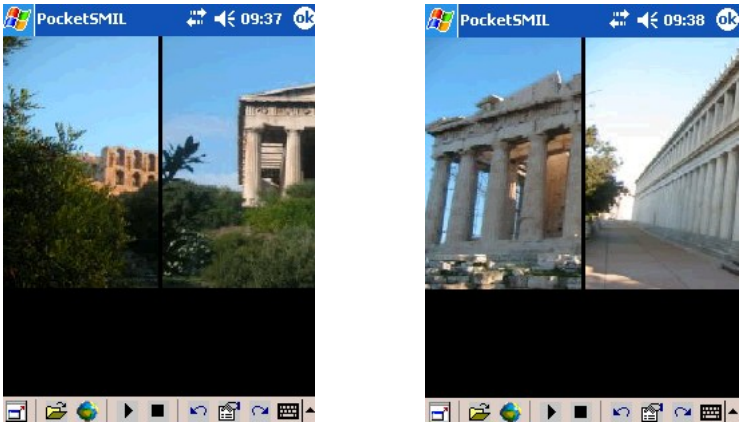


Fig. 3 The multimedia document presented in Figure 2 played on a PDA.

3 Related Work

To play a multimedia document correctly on different devices, the document needs to be adapted, i.e., transformed in order to comply with target device constraints. In Section 3.1, we review existing adaptation techniques and highlight some of their limitations. We propose, in Section 3.2, a more general semantic framework for multimedia document adaptation.

3.1 Current Multimedia Document Adaptation Techniques

A profile is a description of all the capabilities and limitations of a given terminal [20]. Usually, profiles contain the description of both the software and hardware features of a terminal. Software capabilities list, for example, the supported codecs, whether or not the operating system is capable of handling parallel videos or audio streams, etc. The hardware capabilities are given in terms of the physical display size, the input modalities, such as keyboard and pointers. Profiles may also capture user preferences. For example, the user may prefer a particular text magnification for readability purposes. In a more abstract way, a profile can be viewed as a set of constraints on content introduced by profile descriptions. More specifically, for a given multimedia document, the profile imposes constraints on the whole document which in turn can be translated as constraints on each media object or on every pair of media objects contained in it.

For example, a possible profile $\mathcal{P}$ may introduce the following constraints:

ⓐ Multimedia objects cannot be executed simultaneously.
ⓑ Multimedia objects cannot be presented side by side.
ⓒ A hypermedia link must remain active during its entire presentation.

The sample document of Figure 2 does not comply with the profile $\mathcal{P}$. Actually, the document violates all constraints specified in $\mathcal{P}$. These violations mean that the document cannot be executed correctly on the target device containing the profile $\mathcal{P}$. In order to handle different devices with different profiles, a fair amount of research has been conducted on Universal Multimedia Access (UMA) [38]. Besides, many recent approaches are also focusing on multimodal fission [15]. Multimodal fission consists of information output arrangement and organization for different interaction devices.

Two types of adaptation are generally considered: adaptation of multimedia objects individually and adaptation of the document structure or composition. This paper focuses on the latter. For that purpose, we group document adaptation approaches into three categories[2]: specification of alternatives (§3.1.1), using transformation rules (§3.1.2) and multimedia document generation (§3.1.3).

3.1.1 Specification of Alternatives

During the document edition, an author may specify different document versions, based on its content or target profiles. Several languages allow for specifying different versions of a multimedia document, such as standard multimedia languages, like SMIL [4], or multimedia models, like [16,34]. For that purpose, they specify alternative channels that are played only if they comply with the target device profile.

As mentioned earlier, it is often tedious to specify all versions for every possible profile. Furthermore, when a document is created, it is difficult to foresee all constraints, in particular those introduced by futures devices.

[2] Our proposed classification extends the one presented in [26], especially the static and dynamic adaptation cases.

3.1.2 Using Transformation Rules

To avoid the specification of alternatives, given a target device profile, some systems use transformation rules to select a set of predefined transformations that produce automatically a playable document.

For instance, NAC [27] (Negotiation and Adaptation Core) is a software infrastructure for multimedia services adaptation and negotiation in heterogeneous environments. It can adapt any HTML and SMIL documents using several media transcoders combined with XSLT transformations for the document structure.

MPEG-21 [7] is an ISO standard which allows for specifying and play adaptable multimedia documents. Many frameworks use MPEG-21 in order to adapt multimedia contents, such as the one described in [18] which proposes to compose MPEG-21 semantics-based adaptation services. However, these frameworks are limited to single media content adaptation and do not consider the adaptation of the document composition. The work of [3] proposes spatio-temporal transformations applied to MPEG-21 descriptions, especially semantic dependencies between content portions, in order to produce an adapted document. [17] proposes another set of transformation rules that can be applied on HTML web pages.

AHA! [8] adapts the hypermedia dimension only of multimedia documents. This system is mainly based on hypermedia transformation rules described in [9] and considers three different levels of adaptation: content, navigation and presentation.

The approaches based on transformation rules are usually language dependent. Moreover, it is hard to control the quality of the adaptation at the level of the entire document since the resulting document is obtained by applying the rules iteratively.

3.1.3 Multimedia Document Generation

Other approaches for multimedia document adaptation are based on automatic generation for particular classes of documents. For instance, Cuypers [14] is a system that automatically generates Web-based presentations from multimedia databases. It uses a set of abstractions, both on the document and on the presentation level, that are geared towards interactive, time-based and media centric presentations.

MM4U [35] proposes a generic and application-independent framework for supporting multimedia content personalization applications. This framework provides generic components for typical tasks of the general process chain for creating personalized multimedia presentations, i.e., selection, composition, transformation and presentation.

The STAMP model [6] (Synchronized Templates for Adaptable Multimedia Presentations) addresses the dynamic generation of multimedia presentations in the domain of multimedia Web-based information systems. STAMP allows the presentation of multimedia data obtained from XML compatible data sources by means of queries. Furthermore, it proposes templates which are specific document classes to describe the spatial, temporal, navigational structure of multimedia presentations. The instantiation of a template is achieved with respect to the set of spatial and temporal constraints associated with the delivery context.

Despite the fact that these frameworks produce correctly adapted multimedia documents, the generation process requires the documents to be initially described with these templates. These approaches do share the same drawbacks with the device independent format and are limited to some particular classes of documents.

To overcome these drawbacks, we present in the following section a semantic framework for multimedia document adaptation.

3.2 A Semantic Framework for Multimedia Document Adaptation

In [11], a semantic framework for multimedia document adaptation was proposed. This approach is semantic because it is based on interpretations of multimedia documents. In this framework, a multimedia document is interpreted as the set of its potential executions. For instance, in Figure 4, $\mathcal{I}_s$ corresponds to all possible timelines of the multimedia document presented in Figure 2 (these timelines are themselves abstractions of sets of executions). A model of a multimedia document specification is an interpretation satisfying all the constraints of the document. The set of models $\mathcal{M}_s$ of a specification s is thus a subset of the set of interpretations $\mathcal{I}_s$. Actually, e_5 is a model (potential execution) which corresponds to the timeline depicted in Figure 2.

The profile of a target device can be used to identify inside $\mathcal{I}_s$ possible executions, i.e., a set $\mathcal{M}_p$ of executions which comply with all constraints specified in the profile. Therefore, "adapting" is reduced to finding the set of potential executions that are possible, i.e., $\mathcal{M}_s \cap \mathcal{M}_p$. When none is possible, i.e., $\mathcal{M}_s \cap \mathcal{M}_p = \emptyset$, the goal of adaptation is to find executions as close as possible to potential executions that satisfy the profile.

Hence, we distinguish three cases of adaptation illustrated in Figure 4:

- **Compliant adaptation** when $\mathcal{M}_s \subseteq \mathcal{M}_p$;
- **Refining adaptation** when $\mathcal{M}_s \cap \mathcal{M}_p \neq \emptyset$;
- **Transgressive adaptation** when $\mathcal{M}_s \cap \mathcal{M}_p = \emptyset$.

In the following, we bridge the gap between this semantic adaptation theory and its concrete application to standard multimedia documents, such as SMIL documents [4]. In some earlier works [21,23,24], we have proposed some instantiations and extensions of the proposed semantic framework. In this paper, we take the opportunity to combine these contributions. More precisely, we consider the adaptation that deals simultaneously with the spatio-temporal and hypermedia dimensions of standard documents, such as SMIL documents.

4 An Abstract Description of Multimedia Documents

There are several languages or formats for specifying multimedia documents, such as SMIL [4] or Madeus [25]. Instantiating the semantic adaptation framework for a specific multimedia document format requires a dedicated adaptation strategy. Instead, we propose an abstraction layer which allows for hiding the specific syntax and details, and capture the essence of a given document with respect to its main dimensions. We call this document abstraction an abstract multimedia document specification.

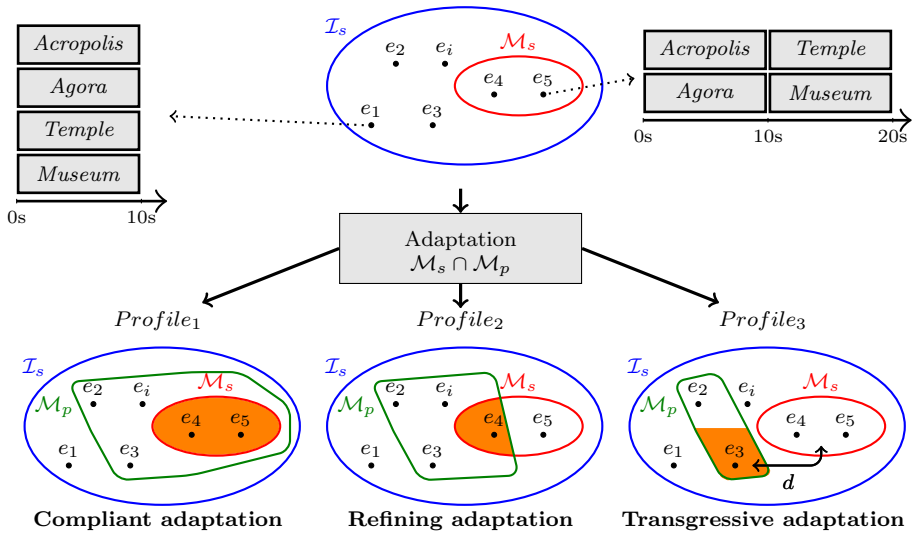


Fig. 4 A semantic framework for multimedia document adaptation.

Definition 1 (Abstract multimedia document specification) An abstract multimedia document specification $s = \langle O, C \rangle$ is made of a set of objects O and a set of relations (or constraints) C between these objects. In the remainder, the relations will be considered as binary.

For instance, the presentation illustrated in Figure 2 refers to the execution e_5 in Figure 4. This execution satisfies the following (temporal) abstract multimedia document specification:

$s_1 = \langle$ {Acropolis, Agora, Temple, Museum},
{$\langle Acropolis\{=\}Agora\rangle$, // Acropolis and Agora are played at the same time
$\langle Temple\{=\}Museum\rangle$, // Temple and Museum are played at the same time
$\langle Acropolis\{m\}Temple\rangle$, // Acropolis is played just before Temple
$\langle Agora\{m\}Museum\rangle$}$\rangle$ // Agora is played just before Museum

Many other executions could reflect the abstract multimedia document specification s_1, e.g., with different multimedia object durations. The advantages of adapting semantic multimedia representations are twofold. First, it allows for reusing the same strategy for different formats. Second, the abstract representation provides more flexibility for the adaptation since the relations between objects can be described qualitatively.

In order to capture the spatio-temporal and hypermedia dimensions of the document illustrated in Figure 2, we propose the following definition for the spatio-temporal and hypermedia specification of a multimedia document.

Definition 2 (Spatio-temporal and hypermedia specification) Let O_m be a set of multimedia objects and O_l a set of hypermedia links. A spatio-temporal

and hypermedia specification $s = \langle O, C \rangle$ is composed of a set of objects O with $O = O_m \cup O_l$ and a set of spatio-temporal relations C between the elements of O.

A spatio-temporal relation $r = \langle r_t, r_s \rangle$ of C is composed of a temporal relation r_t and a spatial relation r_s. In this paper, r_t refers to Allen relations [1] and r_s refers to RCC8 relations [32] (see Figure 5 and 6). An abstract multimedia document specification may be represented as a relation graph.

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)

Fig. 5 The Allen interval algebra (13 temporal relations).

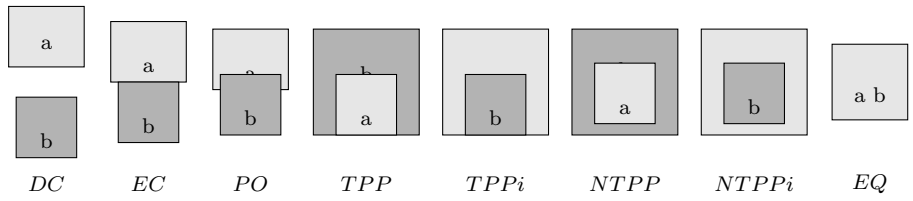


Fig. 6 The RCC8 representation (8 spatial relations).

Definition 3 (Relation graph) Let O be a set of objects and C a set of relations between the elements of O. An abstract multimedia document specification $s = \langle O, C \rangle$ can be represented as a directed labeled graph $g_s = \langle O, E, \lambda \rangle$ called a relation graph. The elements of O and E correspond respectively to the nodes and the edges of g_s, and $\lambda : E \to 2^{\mathcal{R}}$ is a total function from the edges of g_s to the set of relations of a representation $\mathcal{R}$ such that for each $x\ r\ y \in C$ with $x, y \in O$ and $r \in \mathcal{R}$, $r \in \lambda(\langle x, y \rangle)$.

In the remainder, we will see that the finite set $\mathcal{R}$ refers to all relations of a complete spatio-temporal representation.

The spatio-temporal and hypermedia relation graph of the multimedia document illustrated in Figure 2 is presented in Figure 7. It is composed of 6 objects, i.e., 4 multimedia objects and 2 hypermedia links. Each edge is labeled by a qualitative spatio-temporal relation, e.g., the relation $\langle e, DC \rangle$ between *Acropolis* and *Agora* means that these objects are played strictly simultaneously and are disconnected spatially.

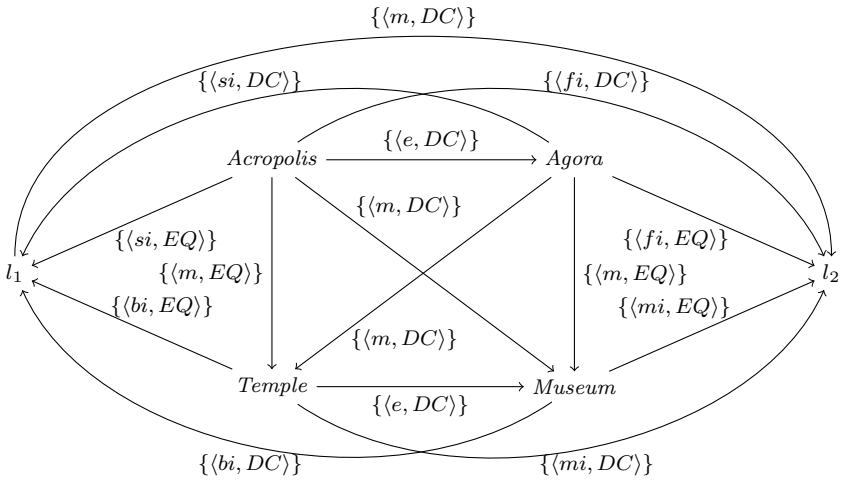


Fig. 7 The spatio-temporal and hypermedia relation graph of the multimedia document illustrated in Figure 2.

According to the profile $\mathcal{P}$ described in Section 3.1, several qualitative relations in Figure 7 are forbidden. For instance, the relation $\langle e, DC\rangle$ between *Acropolis* and *Agora* violates the constraints ⓐ and ⓑ of the profile $\mathcal{P}$. Actually, the temporal relation e (equals) is not compatible with the former constraint ⓐ because the two multimedia objects are played strictly at the same time, while ⓐ indicates that it is forbidden. Moreover, the spatial relation DC (disconnected) does not satisfy the latter constraint ⓑ because the two images are presented side-by-side, while ⓑ prohibits it.

In this context, the abstract multimedia document specification must be adapted, i.e., the forbidden relations must be replaced by other relations which comply with the constraints specified in the profile. In the following section, we show how to achieve this task by transforming the relation graph in a minimal way.

5 Adaptation of a Multimedia Document Specification

The semantic adaptation framework requires exploring potential executions to find those satisfying the adaptation constraints, and, if none is found to generate close specifications satisfying the constraints. For that purpose, we first define the notion of closeness through a specific distance (§5.1), then we describe our algorithm (§5.2).

5.1 Defining a Distance Between Abstract Multimedia Document Specifications

To provide an adapted relation graph, i.e., in which all relations comply with a profile, which is as close as possible from an initial relation graph, we define a distance between two relation graphs that depends on the proximity between relations beared by the same arc in both graphs.

Definition 4 (Distance between two relation graphs)

$$d(\lambda, \lambda') = \sum_{n,n' \in O} \text{Min}_{r \in \lambda(\langle n,n' \rangle),\ r' \in \lambda'(\langle n,n' \rangle)} \delta(r, r')$$

We consider that the proximity between two relations relies on the conceptual neighborhood between these relations and is measured by the shortest path distance in the corresponding neighborhood graph with δ.

Definition 5 (Conceptual neighborhood) Let x and y be two objects and $r(x, y)$ a relation between x and y. The conceptual neighborhood relation is a binary relation $N_{\mathcal{R}}^X$ between elements of a set of relations $\mathcal{R}$ such that $N_{\mathcal{R}}^X(r, r')$ iff the continuous transformation X can transform $r(x, y)$ into $r'(x, y)$ without transiting by a third relation $r''(x, y)$.

Figure 8 presents a conceptual neighborhood for the Allen temporal relations [12] and Figure 9 presents another conceptual neighborhood for the RCC8 spatial relations [32][3]. These neighborhood graphs allow for computing the distance δ between two temporal or spatial relations.

Definition 6 (Conceptual distance between relations) The conceptual distance δ between two relations is the length of the shortest path between r and r' in the graph of $N_{\mathcal{R}}^X$.

For example, the distance δ between the *before* (b) and the *overlaps* (o) temporal relations is 2 in Figure 8. The distance δ between the *disconnected* (DC) and the *equals* (EQ) spatial relations is 4 in Figure 9.

As we consider in this paper spatio-temporal relations, i.e., $r = \langle r_t, r_s \rangle$, it is necessary to define the distance on such relations. We propose to simply sum the two distances, like in Manhattan distances. It is possible to give more importance to one or another component by choosing a different distance.

Definition 7 (Spatio-temporal distance between relations) Let $r_1 = \langle r_t^1, r_s^1 \rangle$ and $r_2 = \langle r_t^2, r_s^2 \rangle$ be two spatio-temporal relations, and N_t^X and $N_s^{X'}$ a temporal and a spatial neighborhood graph. The spatio-temporal distance $\delta(r_1, r_2) = \delta_t(r_t^1, r_t^2) + \delta_s(r_s^1, r_s^2)$ where δ_t is computed from N_t^X and δ_s is computed from $N_s^{X'}$.

For instance, $\delta(\langle e, DC \rangle, \langle m, EQ \rangle) = \delta_t(e, m) + \delta_s(DC, EQ) = 3 + 4 = 7$ with δ_t and δ_s computed from the two neighborhood graphs presented in Figure 8 and 9.

Other metrics for computing δ can be used together with other neighborhood graphs. However, they must reflect the semantic proximity between two spatio-temporal relations and thus two spatio-temporal and hypermedia specifications.

[3] The used continuous transformations are defined in [12] for the Allen relations and in [32] for the RCC8 relations.

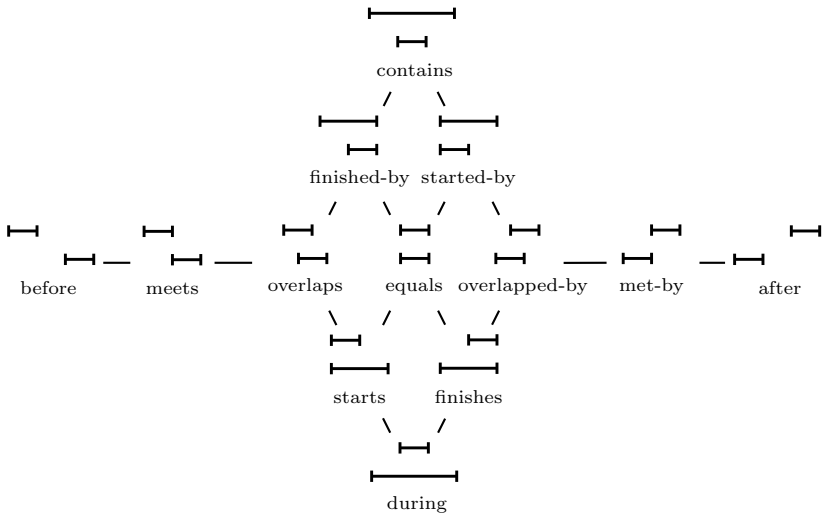


Fig. 8 A neighborhood graph for Allen temporal relations ($N^A_{A_{13}}$).

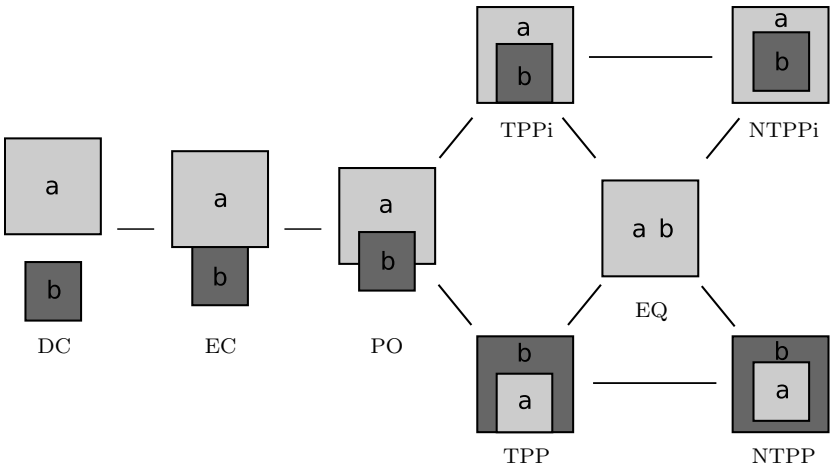


Fig. 9 A neighborhood graph for RCC8 spatial relations (N^T_{RCC8}).

5.2 Adaptation Algorithm

From an initial relation graph and a given profile, we compute adapted relation graph solutions which are close from the initial one and such that all of their relations satisfy the profile. For that purpose, we consider all possible relation graphs which satisfy a given profile and for each of them we select those which are at the minimal distance from the initial relation graph.

This process has been implemented through the *Adaptation* algorithm (Algorithm 1). This algorithm extends the Nebel's backtracking algorithm [29], which enumerates consistent relation graphs. This faster extension uses a branch-and-bound search. The problem of finding consistent relation graphs is known to be NP-complete [13], however many optimizations and heuristics have been proposed, e.g., [37] treats matrix elements and relations in privileged orderings, [29] groups predefined set of relations and [10] optimizes the backtracking search with a forward-checking scheme. These algorithm variants can be used to enhance the execution performance of the *Adaptation* algorithm.

Algorithm 1: Adaptation

```
Input: An initial matrix I corresponding to the initial relation graph and a matrix P
       corresponding to relations that satisfy all constraints specified in a profile.
Data: Min is a current minimum computed distance (initialized with a high value).
Output: S is a set of adapted relation graphs.
pathConsistency(P);
if P does not contain an empty relation then
    Choose an unprocessed label P_{x,y} and split P_{x,y} into r_z = r_1, ..., r_n;
    if no label can be split then
        tmp ← d(I, P);
        if tmp < Min then
            Min ← tmp;
            S ← {P};
        end
        if tmp = Min then
            S ← S ∪ {P};
        end
    end
    else
        for all labels r_z (1 ≤ z ≤ n) do
            P_{x,y} ← r_z;
            if d(I, P) ≤ Min then
                Adaptation(I,P);
            end
        end
    end
end
```

The algorithm inputs are two relation graphs: one corresponding to the initial document specification and another one corresponding to possible relations which satisfy a profile. Both specifications are encoded into two matrix I and P, respectively. Before the $Adaptation(I, P)$ call, we first sort each element $P_{x,y}$ according to the distance δ (Definition 7) from each element $I_{x,y}$.

Algorithm 1 computes the consistent possible relation graphs with the help of the *pathConsistency* function[4]. When the computed relation graphs are consistent, we select the minimal ones with the distance d defined in Definition 4, and store them in $\mathcal{S}$. When Algorithm 1 terminates, the set $\mathcal{S}$ contains all adapted relation graph solutions and the variable Min is the minimal distance from the initial document specification.

[4] A path consistency function was preliminary defined for qualitative Allen temporal relations in [1].

On the multimedia document of Figure 2, from the initial relation graph presented in Figure 7 and the profile $\mathcal{P}$ presented in Section 3.1, Algorithm 1 computes the adapted relation graph illustrated in Figure 10. This adapted relation graph contains only spatio-temporal relations that satisfy all constraints specified in the profile $\mathcal{P}$. Moreover, the distance d between the initial and this adapted relation graph is equal to 52 (this is the minimal distance for satisfying the profile $\mathcal{P}$).

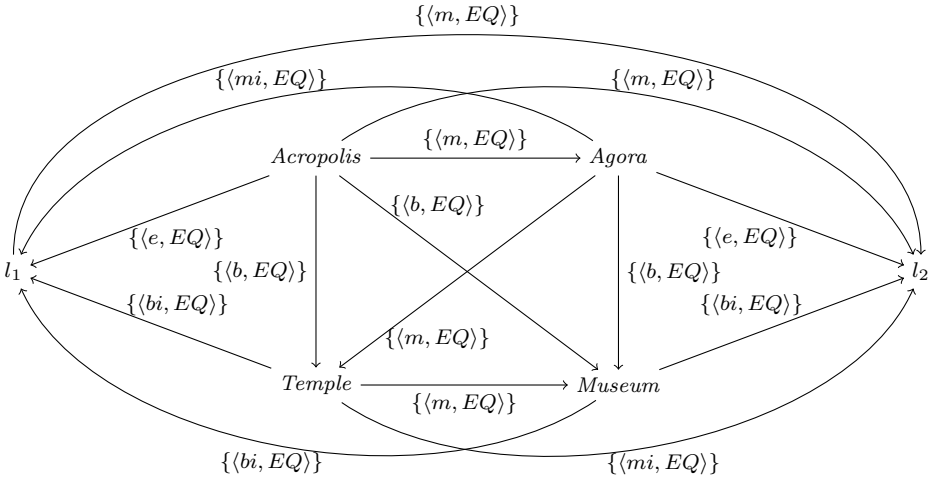


Fig. 10 An adapted relation graph in which all relations comply with the profile $\mathcal{P}$.

Figure 11 illustrates the execution on a PDA corresponding to the adapted relation graph presented in Figure 10. All constraints specified in the profile $\mathcal{P}$ are now satisfied. Additionally, the produced adapted document is close to the initial one because the presentation ordering of multimedia objects has been preserved.

We have shown in this section that the semantic framework proposed in [11] can be extended to combine the spatio-temporal and hypermedia dimensions of abstract multimedia document specifications.

6 Application to Standard Multimedia Documents

This semantic adaptation framework has to be applied on SMIL documents [4]. SMIL is a W3C recommendation for Synchronized Multimedia documents on the web. It is an XML-based language which enables simple authoring of interactive audiovisual presentations. Moreover, it is the main format used on mobile phone to encode Multimedia Messages (MMS) together with online multimedia services [30]. Many applications and players support SMIL documents, e.g., Real Player[5], Ambulant[6], and PocketSMIL[7] and virtually every mobile phone supporting MMS.

[5] `http://www.real.com/international/`

[6] `http://www.cwi.nl/projects/Ambulant/`

[7] `http://wam.inrialpes.fr/software/pocketsmil/`

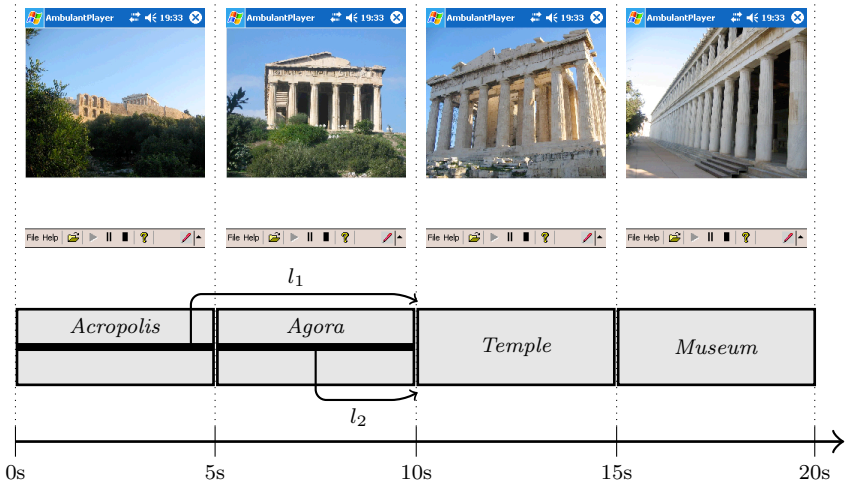


Fig. 11 An execution, played on a PDA, which corresponds to the adapted relation graph presented in Figure 10.

The temporal and spatial behaviors of a SMIL document are not expressed with qualitative relations, as used in the previous sections, but mainly with quantitative values (especially, for the specification of the spatial dimension).

Consequently, the natural way of using the adaptation framework presented in the previous section for adapting actual documents, consists of taking the equivalent qualitative representation of the document ($\hat{\alpha}$), adapting it and translating it back to an adapted document ($\check{\alpha}$) as illustrated in Figure 1. The pair of functions $\langle\hat{\alpha}, \check{\alpha}\rangle$ have to satisfy natural constraints like: $\check{\alpha} \cdot \hat{\alpha} = Id$.

This strategy can be applied to different multimedia specification languages or formats. Actually, the translation functions $\hat{\alpha}$ and $\check{\alpha}$ must be defined for each particular language. For instance, a SMIL document can be adapted and converted into another multimedia document language, like MPEG-4 BIFS [36].

With respect to our proposed strategy, the goal of the $\hat{\alpha}$ function is the extraction of the relational structure involving the document objects and its encoding as a relation graph on which to apply the above defined adaptation operations. This extraction is obtained by the following procedure:

1. extract all multimedia objects and hypermedia links, and make them a node in the initial relation graph;
2. add an edge between each pair of nodes (they are labeled by a set of spatio-temporal relations);
3. extract the relations implied by the spatio-temporal and hypermedia dimensions.

Simultaneously, a target device profile is interpreted, like in [28], in order to identify qualitative and quantitative presentation constraints, such as the screen size or the maximum number of objects that can be played in parallel. Thanks to the profile and the qualitative representation built from $\hat{\alpha}$, it is possible to adapt the document representation with the adaptation framework described in the previous section.

Thereafter, we need to inject the adapted information into the document. For that purpose, the $\check{\alpha}$ function can be defined in a straightforward manner:

1. compute a quantitative solution from the qualitative adapted representation. For that purpose, we consider the initial document quantitative values and the adapted computed qualitative relations between objects. The Cassowary Solver [5], a quadratic optimizer, is then used for computing an adapted quantitative solution. Actually, it computes a quantitative solution close to the initial values that complies with all the given qualitative relations. Quantitative constraints specified in the profile, like the screen size, are also sent to the solver in the form of linear inequalities in order to produce a suitable and playable document;
2. propagate the quantitative solution information into the document, e.g., update the multimedia objects synchronization or the display layout.

The presented adaptation framework has been implemented in an interactive adaptation prototype. The prototype architecture is depicted on Figure 1 and uses Algorithm 1 to produce adapted solutions. Figure 12 presents a screenshot of our prototype. A video screencast is also available at [22].

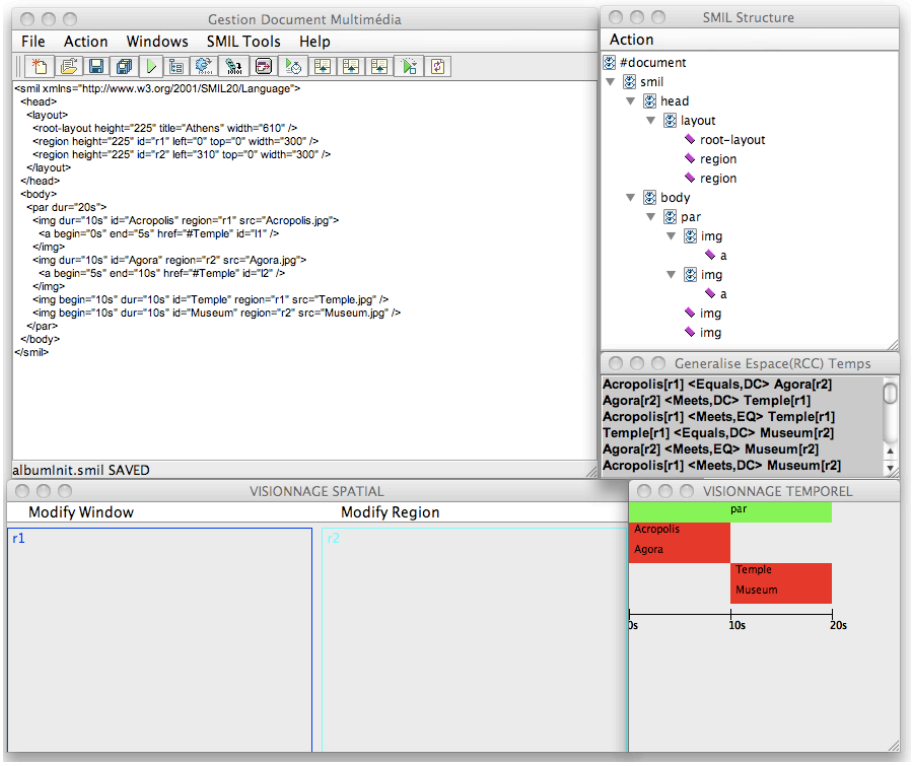


Fig. 12 Screenshot of our SMIL document adaptation prototype.

This implementation can adapt multimedia documents in a client-server architecture on the client side, on the server side or on intermediate entities, such

as proxies. More precisely, the adaptation mechanism can be fully automated and made transparent to the user. The interactive prototype might also be useful for guiding an author during the document creation for providing feedback on how adaptations may look on particular terminals. Actually, one may specify a profile and if an author creates a forbidden relation between two objects the system could provide 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.

7 Conclusion

Multimedia document adaptation is a very important feature for the continuity of the web experience by users. However, the multiplicity of contexts in which such documents have to be played prevents from leaving this adaptation task to document designers. Document adaptation must then be considered at a more abstract level and be independent from the various document formats.

We have shown how to fill the gap between an abstract semantic framework for multimedia document adaptation and its concrete application on the spatio-temporal and hypermedia dimensions of standard multimedia documents, such as SMIL documents. The originality of this approach is that the adaptation process is flexible and independent of any multimedia description language: only transformation rules have to be defined for dealing with a new concrete language. Moreover, the computed adapted solutions are guarantee to remain close to the initial multimedia document.

References

1. J. Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, 1983.
2. I. Amous, A. Jedidi, and F. Sèdes. A contribution to multimedia document modeling and querying. *Multimedia Tools and Applications*, 25(3):391–404, 2005.
3. 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.
4. J. Ayars, D. Bulterman, A. Cohen, K. Day, E. Hodge, P. Hoschka, E. Hyche, M. Jourdan, M. Kim, K. Kubota, R. Lanphier, N. Layaïda, T. Michel, D. Newman, J. van Ossenbruggen, L. Rutledge, B. Saccocio, P. Schmitz, and W. ten Kate. *Synchronized Multimedia Integration Language (SMIL 2.0) - [Second Edition].* W3C, 2005. `http://www.w3.org/TR/2005/REC-SMIL2-20050107/`.
5. G. J. Badros, A. Borning, and P. J. Stuckey. The cassowary linear arithmetic constraint solving algorithm. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(4):267–306, 2001.
6. I. M. Bilasco, J. Gensel, and M. Villanova-Oliver. STAMP: a model for generating adaptable multimedia presentations. *Multimedia Tools and Applications, Special Issue on Metadata and Adaptibility in Web-based Information Systems*, 25(3):361–375, 2005.
7. J. Bormans and K. Hill. *MPEG-21 Overview v.5.* ISO/IEC JTC1/SC29/WG11/N5231, 2002. `http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm`.
8. P. D. Bra, D. Smits, and N. Stash. The design of AHA! In U. K. Wiil, P. J. Nürnberg, and J. Rubart, editors, *Proceedings of the 17th ACM Conference on Hypertext and Hypermedia*, pages 133–134. ACM Press, 2006.

9. P. Brusilovsky. Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3):87–129, 1996.
10. J.-F. Condotta, G. Ligozat, and M. Saade. Eligible and frozen constraints for solving temporal qualitative constraint networks. In *Proceedings of the 13th International Conference on Principles and Practice of Constraint Programming*, pages 806–814, 2007.
11. 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*, pages 31–36, 2003.
12. C. Freksa. Temporal reasoning based on semi-intervals. *Artificial Intelligence*, 54(1–2):199–227, 1992.
13. A. Gerevini and B. Nebel. Qualitative spatio-temporal reasoning with RCC-8 and Allen's interval calculus: Computational complexity. In *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 312–316, 2002.
14. J. Geurts, J. van Ossenbruggen, and L. Hardman. Application-specific constraints for multimedia presentation generation. In *Proceedings of the International Conference on Multimedia Modeling*, pages 247–266, 2001.
15. P. Grifoni. Multimodal fission. In P. Grifoni, editor, *Multimodal Human Computer Interaction and Pervasive Services*, chapter 6, pages 103–120. IGI Global, 2009.
16. L. Hardman, D. C. A. Bulterman, and G. van Rossum. The Amsterdam hypermedia model: adding time and context to the Dexter model. *Communications of the ACM*, 37(2):50–62, 1994.
17. J. He, T. Gao, W. Hao, I.-L. Yen, and F. Bastani. A flexible content adaptation system using a rule-based approach. *IEEE Transactions on Knowledge and Data Engineering*, 19(1):127–140, 2007.
18. D. Jannach, K. Leopold, C. Timmerer, and H. Hellwagner. A knowledge-based framework for multimedia adaptation. *Applied Intelligence*, 24(2):109–125, 2006.
19. E. Kirda and C. Kerer. DIWE: A Framework for Constructing Device-Independent Web Applications. In *Ubiquitous Mobile Information and Collaboration Systems*, pages 96–110. Springer, 2005. LNCS 3272.
20. 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/`.
21. S. Laborie. Spatio-temporal proximities for multimedia document adaptation. In *Proceedings of the 12th International Conference on Artificial Intelligence: Methodology, Systems, Applications*, pages 128–137. Springer, 2006. LNAI 4183.
22. S. Laborie. A screencast of our SMIL document adaptation prototype, 2009. `http://www.irit.fr/PERSONNEL/SIG/laborie/videos/ScreenCast-Adaptation.mov`.
23. 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.
24. S. Laborie, J. Euzenat, and N. Layaïda. A spatial algebra for multimedia document adaptation. In *Poster Proceedings of the First International Conference on Semantic and Digital Media Technologies*, pages 7–8, 2006.
25. N. Layaïda. *Madeus : système d'édition et de présentation de documents structurés multimédia*. PhD thesis, Université Joseph Fourier (Grenoble, France), 1997.
26. 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.
27. T. Lemlouma and N. Layaïda. The negotiation of multimedia content services in heterogeneous environments. In *Proceedings of the 8th International Conference on Multimedia Modeling*, pages 187–206, 2001.
28. T. Lemlouma and N. Layaïda. Context-aware adaptation for mobile devices. In *Proceedings of the 5th IEEE International Conference on Mobile Data Management*, pages 106–111, 2004.
29. B. Nebel. Solving hard qualitative temporal reasoning problems: Evaluating the efficiency of using the ORD-horn class. In *Proceedings of the 12th European Conference on Artificial Intelligence*, pages 38–42, 1996.
30. Open Mobile Alliance. *Multimedia Messaging Service Conformance Document*, January 2008. `http://www.openmobilealliance.org/Technical/release_program/docs/MMS/V1_3-20080128-C/OMA-TS-MMS-CONF-V1_3-20080128-C.pdf`.

31. J. Rabin and C. McCathieNevile. *Mobile Web Best Practices 1.0*. W3C, 2008. `http://www.w3.org/TR/mobile-bp/`.
32. D. A. Randell, Z. Cui, and A. Cohn. A spatial logic based on regions and connection. In B. Nebel, C. Rich, and W. Swartout, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176. Morgan Kaufmann, 1992.
33. C. Roisin. Authoring structured multimedia documents. In *Proceedings of the Conference on Current Trends in Theory and Practice of Informatics*, pages 222–239, 1998.
34. F. Rousseau, J. A. García-Marcías, J. V. de Lima, and A. Duda. User adaptable multimedia presentations for the world wide web. In *Proceedings of the 8th International Conference on World Wide Web*, pages 1273–1290. Elsevier North-Holland, Inc., 1999.
35. A. Scherp and S. Boll. MM4U - a framework for creating personalized multimedia content. In *Managing Multimedia Semantics*, chapter 11, pages 246–287. IRM Press, 2005.
36. B. Shao, L. M. Velazquez, N. Scaringella, N. Singh, and M. Mattavelli. SMIL to MPEG-4 BIFS Conversion. In *AXMEDIS '06: Proceedings of the Second International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution*, pages 77–84. IEEE Computer Society, 2006.
37. P. van Beek and D. W. Manchak. The design and experimental analysis of algorithms for temporal reasoning. *Journal of Artificial Intelligence Research*, 4:1–18, 1996.
38. A. Vetro, C. Christopoulos, and T. Ebrahimi, editors. *IEEE Signal Processing Magazine Special issue on Universal Multimedia Access*, volume 2 of *2*. IEEE Press, March 2003.
39. W3C. Mobile Web Initiative, 2005. `http://www.w3.org/2005/MWI/Activity.html`.
40. W3C. Ubiquitous Web Domain, 2010. `http://www.w3.org/UbiWeb/`.