

# INTERFACE D'ÉDITION DE DOCUMENTS STRUCTURÉS MULTIMÉDIA

Nabil Layaida et Jean-Yves Vion-Dury

Unité mixte Bull-IMAG/Systèmes.

Z.I Mayencin, 2 Av. de Vignate, 38610 Gières.

E-mail: {Nabil.Layaida, Jean-Yves.Vion-Dury}@imag.fr

**Résumé :** L'édition multimédia est un processus créatif complexe, se devant d'être souple et interactif d'une part, et intelligent d'autre part si l'on considère que le système d'édition doit décharger au maximum l'utilisateur des tâches de bas niveau et abstraire la complexité temporelle. Nous proposons une approche synthétique prenant en compte ces deux aspects en s'appuyant d'une part sur un modèle de document multimédia structuré permettant des spécifications temporelles algébriques de haut niveau, et d'autre part sur l'utilisation de représentations visuelles interactives multiples. Celles-ci sont tridimensionnelles pour répondre à la complexité qualitative et quantitative du processus. Les différentes vues sont maintenues en cohérence forte au cours du processus d'édition grâce à une représentation interne unique du document. Cette dernière peut être construite automatiquement par application des opérateurs algébriques sur la structure du document que notre modèle permet de spécifier.

**Mot clés :** édition multimédia, interface d'édition, algèbre temporelle, images virtuelles, vues multiples.

## 1 Édition multimédia : état de l'art

La grande majorité des systèmes d'édition actuels proposent de créer des documents multimédia au moyen d'un langage de programmation, ou *script*, interprété ou compilé, permettant une description précise des contraintes spatiales et temporelles. Cependant, en se préoccupant essentiellement des contraintes de bas niveau, la plupart de ces éditeurs proposent une vision du document complexe, difficile à modifier, à maintenir et à percevoir. De plus, l'approche "programmation", en s'éloignant de la structure fondamentale logique du document, rend difficile l'utilisation des techniques de formatage automatique, déjà présentes dans les éditeurs textuels avancés [Furuta88], et généralisables aux différents médias. De fait, la majorité de ces systèmes d'édition mélangent implicitement instructions de formatage et structures logiques, accroissant considérablement la complexité du document et sa compréhension.

Certains auteurs prennent en considération l'importance de l'interactivité dans le processus d'édition, et proposent des représentation graphiques qui sont essentiellement de type linéaire (*time line*) [Drapeau91], où les entités multimédia sont disposées sur un plan dont l'abscisse représente le temps. Les contraintes temporelles sont alors exprimées de manière absolue, en fixant par exemple la date de commencement et de fin de l'objet sur l'échelle horizontale.

## 2 Vers un modèle structuré de documents multimédia

La modélisation des documents multimédia doit répondre à un ensemble de contraintes : d'une part, fournir une représentation uniforme des données multimédia pour simplifier l'interface de l'utilisateur, et d'autre part offrir une représentation du document suffisamment riche et structurée pour permettre une manipulation efficace. Les documents multimédia sont perçus comme un ensemble d'objets liés dans le temps et dans l'espace, chaque objet possédant une dimension dans un ensemble de médias (dimension spatiale), cette dimension étant caractérisée par une date d'occurrence et une durée de présentation (dimension temporelle).

La fonction d'un éditeur multimédia est de prendre les spécifications des relations spatiales et temporelles entre les objets afin de les projeter sur chacune de ces dimensions. L'utilisateur n'intervient de manière détaillée que lorsqu'il doit corriger des incohérences de spécification ou lorsqu'il veut affiner certains détails de présentation. Ce modèle de l'édition repose sur la représentation des données multimédia et surtout du temps, la dimension temporelle étant celle qui fédère l'ensemble des médias.

### 2.1 Modélisation des objets multimédia

Un objet multimédia est caractérisé par son nom symbolique, son type (média) et un ensemble d'attributs de présentation (taille, durée, ressources nécessaire à sa présentation, etc.). Le type de l'objet et l'ensemble de ses attributs permettent de définir les paramètres de présentation, les ressources dont il a besoin, et éventuellement les opérations de transformation qui peuvent lui être appliquées, et dans quelles limites pour ne pas altérer son contenu (étirement dans le temps, dans l'espace, etc.).

Nous pouvons distinguer trois types d'objets multimédia par rapport à leurs caractéristiques temporelles :

1. **Les objets discrets** : ils ne possèdent pas de dimension temporelle intrinsèque, par exemple le texte et les images fixes.
2. **Les objets continus** : leur présentation est soumise à une contrainte de débit. La durée de présentation de ces données peut donc être calculée à l'avance. C'est le cas du son et de la vidéo.
3. **Les objets indéterministes (discrets ou continus)** : leur durée de présentation est imprévisible. D'une manière générale, ce sont des objets retournés à l'éditeur par appel externe de programme (calcul d'une image de synthèse par exemple).

Un document est obtenu en organisant les objets multimédia au sein de deux structures distinctes :

1. **Les structures logiques** permettent l'organisation hiérarchique des objets d'un même média, par exemple un objet texte peut être représenté par un titre et une suite de paragraphes ; une vidéo, par un générique et une suite de scènes, etc.
2. **Les structures temporelles** : basées sur une algèbre utilisant des opérateurs présentée dans la sous-section 2.2.

Cette structuration des objets multimédia joue un rôle important au niveau de l'interface, car elle permet une réduction progressive de la complexité d'édition d'un document en permettant la manipulation de groupes d'objets (objets composites). Un objet composite est alors défini comme étant un ensemble d'objets multimédia et de leurs relations spatiales et temporelles.

## 2.2 Modélisation du temps et des documents

Considérant deux intervalles de temps distincts, il existe treize manières de les composer entre eux. Cette approche permet de définir autant d'opérateurs binaires permettant d'exprimer algébriquement les relations temporelles entre deux objets, et de construire des expressions complexes [Ghafoor92]. La mise en œuvre de ces opérateurs, qui sont à la base de l'algèbre multimédia, nécessite la prise en compte du temps absolu (spécification comportant une mesure directe du temps) ou d'un temps logique (spécification de relations causales).

Le langage basé sur ces opérateurs inclut des variables (noms d'objets multimédia de base), des opérations et des expressions algébriques facilement représentable graphiquement sous forme d'arbre. Les fonctions de l'algèbre sont classées en trois catégories :

1/ **Fonctions de désignation** : ce sont les fonctions permettant l'accès aux objets. Souvent ces fonctions sont couplées à des systèmes de recherches d'objets par le contenu et l'usage de noms permet entre autres la réutilisation des données multimédia qui occupent des volumes de stockage importants.

2/ **Fonctions de composition** : définissent les opérateurs temporels appliqués aux objets et aux expressions (opérateur *equals*, *meets*, *overlaps*, *before*, etc.). L'ensemble des opérateurs de l'algèbre est décrit dans [Allen83].

3/ **Fonctions de présentation** : définissent les directives de présentation appliqués aux objets par l'intermédiaire de leur attributs (fontes du texte, couleur, durée de présentation, etc.) et éventuellement les opérations de transformation (étirement dans le temps ou dans l'espace).

Les opérateurs algébriques ne suffisent pas pour décrire certaines contraintes temporelles de bas niveau. A cette fin, nous avons prévu d'appliquer les contraintes à deux niveaux : entre intervalles où les objets sont considérées comme l'élément de base de la composition, entre événements où chaque objet est représenté par deux marqueurs début-fin.

Dans tous les cas la représentation temporelle interne d'un document est exprimée en terme d'événements, car quelque soit le niveau d'expression des opérateurs temporels, il est toujours possible d'extraire cette représentation (par compilation ou interprétation). Sous cette forme, un document défini *un graphe de dépendances temporelles* formé par un ensemble de nœuds (les événements) reliés par des arcs (les relations de précedence temporelle) [Rossum93]. Ce graphe est la représentation fondamentale d'un document, toutes les opérations de l'éditeur (formatage temporel et spatial, pré-allocation des ressources et vérification de la cohérence temporelle) l'utilisent comme donnée de base. Il est notamment utilisé lors de la phase finale de présentation du document au lecteur pour orienter l'enchaînement des événement de présentation.

### 3 Interface d'édition

Le rôle de l'interface d'édition est de favoriser la créativité de l'utilisateur en répondant correctement à la complexité de la tâche. Complexité qualitative de part la multiplicité des concepts mis en œuvre (contraintes temporelles multiples et de nature différentes, hétérogénéité des canaux physiques impliqués dans l'édition, structure logique et sémantique du document), mais aussi quantitative si l'on considère la conception de documents volumineux.

Nous proposons d'aborder cette complexité en suivant trois principes :

1/ *Utilisation d'images virtuelles* : Les images virtuelles sont des représentations visuelles tridimensionnelles interactives et animées [VionDury94]. En utilisant certaines propriétés que nous ne détaillerons pas ici, elles permettent de visualiser des ensembles d'objets à forte cardinalité, ainsi que leurs relations logiques. De part leur forte interactivité, leur indépendance vis à vis de tout matériel spécialisé et leur aptitude à exprimer des relations complexes, elles sont très adaptées selon nous à la manipulation d'ensembles d'objets multimédia.

2/ *Utilisation de vues multiples* : Cette approche s'impose lorsque que l'on veut représenter un ensemble d'informations fortement complexe. Les vues multiples permettent de projeter cette complexité sur des espace de moindre dimension, de filtrer et d'orienter l'information. Elles peuvent être de ce fait orthogonales ou complémentaires.

3/ *Maintien d'une cohérence forte entre les vues multiples* : Un des inconvénients majeurs des représentations visuelles multiples est lié à la difficulté de saisir la globalité de l'information, et de les maintenir en cohérence. Nous proposons d'utiliser le *graphe de dépendances temporelles* précédemment évoqué pour centraliser l'information et la diffuser à toutes les vues. Ainsi, une modification faite dans une vue quelconque se répercute sur le graphe central, produisant automatiquement une mise à jour dans l'ensemble des vues, où l'information apparaît de manière contextuelle.

Les trois vues que nous décrivons dans les sous-sections suivantes nous semblent fondamentales pour assurer efficacement le processus d'édition.

#### 3.1 La vue structurale

Cette vue, dont nous avons réalisé une maquette (Fig. 1) représente la structure temporelle du document sous la forme d'un arbre dont les nœuds sont des opérateurs temporels, et les feuilles des objets multimédia. Les opérateurs et les objets sont représentés par des polyèdres dont la forme et la couleur correspondent à leurs types (prismes blancs pour les opérateurs, polyèdres colorés pour les objets multimédia). L'expansion spatiale de l'arbre correspond à la nature des opérateurs : ses branches se développent dans le plan ( $Y=0$ ) pour les opérateurs parallèles et dans le plan ( $X=0$ ) pour les opérateurs séquentiels. Ainsi, l'utilisateur conserve une perception intuitive de la structure temporelle, puisque l'axe des Y exprime de cette manière un temps logique. L'axe des Z est réservé à l'expression de la hiérarchie. Outre l'interaction permettant d'explorer l'espace de travail, l'utilisateur peut agir sur la structure pour l'enrichir ou la modifier, "ouvrir" les polyèdres représentant des objets composites holophrastés, et ainsi, accéder à une description plus fine de leur contenu. Il peut également exprimer des contraintes temporelles non hiérarchiques, directement d'objet à objet, en les reliant par des segments typés.

Cette dernière possibilité nous paraît indispensable pour équilibrer l'approche purement algébrique en la rendant plus souple d'utilisation.

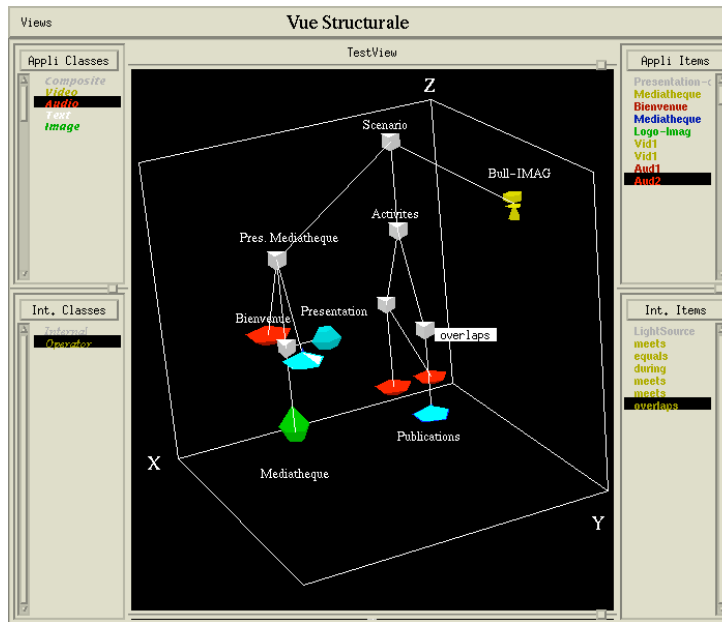


Fig. 1 : vue structurale d'un document multimédia

### 3.2 La vue ressources

La vue ressource, ou *timeline*, est très utilisée dans les éditeurs multimédia. Elle est formée par un ensemble d'axes temporels où chacun représente un canal d'un média particulier. Les objets multimédia sont représentés par des segments dont la longueur correspond à leur durée de présentation. L'utilisateur peut agir sur la taille des objets en modifiant leurs attributs temporels ou spatiaux. Cette représentation sert avant tout à allouer les ressources, prévenir les conflits et spécifier les dimensions des objets.

Nous proposons de représenter les canaux multimédia par des lignes parallèles à un axe temporel. Ces lignes sont disposées de manière cylindrique, s'enroulant autour de l'axe placé au centre, et sont considérées comme un support sur lequel l'utilisateur peut déplacer les objets par manipulation directe. La disposition globale, de nature linéaire, permet d'utiliser la perspective pour améliorer la perception d'ensemble du déroulement temporel. Les contraintes de temps, visualisées par des segments de liaison inscrits dans le volume intérieur du volume cylindrique, utilisent une métaphore de ressorts, dont la constante d'élasticité dépend du type de contrainte. Ainsi, lorsqu'on déplace un objet sur son axe temporel, on entraîne automatiquement tous les objets qui lui sont reliés par des contraintes de précedence fortes. Les objets eux-même sont représentés à l'aide du code de couleur utilisé dans la vue structurale par deux polyèdres de même forme correspondant aux marqueurs de début et fin de l'objet, et reliés par un ressort (les objets discrets sont confondus, et le ressort possède une constante d'élasticité nulle). La distance entre le marqueur de début et de fin de l'objet correspond intuitivement à sa profondeur temporelle. Les conflits de ressources sont visualisés par des chevauchements de marqueurs facilement identifiables. L'utilisateur configure l'information par manipulation directe des marqueurs, et par ajout ou retrait de ressorts et d'objets.

La vue ressources a été prise comme vue centrale du processus d'édition dans MAEStro [Drapeau91]. À notre avis, cette vue ne peut être efficace que si le scénario du document est connu à l'avance ; or en général le processus d'édition est un cycle rédaction–visualisation pour lequel cette vue est inadaptée car elle ne met pas en évidence les relations entre objets, ni la structure du document.

### 3.3 La vue présentation

Cette vue présente le document tel qu'il sera affiché à l'utilisateur final sur l'écran. Étant donné la dimension temporelle du document, celui-ci est représenté en trois dimensions (deux pour l'espace fenêtre et une troisième pour le temps). Le noyau fonctionnel de l'éditeur élabore, d'après la description interne du document, un premier formatage spatial. Cette opération est réalisée automatiquement au moyen d'un formateur classique, qui prend en charge le placement du texte, des images et de la vidéo dans la fenêtre de présentation. Suivant le même principe, les autres objets multimédia sont placés sur leurs dimensions respectives. La vue présentation est composée d'un ensemble de plans qui correspondent aux différentes scènes d'un document. Ces plans permettent au rédacteur de juger de la disposition spatiale des objets, et éventuellement de la corriger en forçant un objet à apparaître à un endroit particulier où dans un plan autre que celui choisi par le formateur. Ainsi, cette vue permet d'affiner le formatage effectué par l'éditeur, ce qui apporte une souplesse dans la présentation des documents dont le manque est souvent reproché aux éditeurs structurés.

## 4 Conclusion

A travers cette proposition, nous espérons avoir démontré l'importance d'un modèle de structuration temporelle des documents multimédia pour l'élaboration d'une interface visuelle adaptée à la complexité du processus d'édition.

Peu de travaux ont porté sur la définition d'une interface d'édition, et encore moins d'une interface de visualisation. Nous espérons élargir le débat auprès de la communauté de la communication homme–machine pour progresser dans cette voie.

## Bibliographie

- [Allen83] J.F. Allen "*Maintaining Knowledge about Temporal Intervals*" Communications of the ACM, pp.832–843, Vol 26, Numéro 11, Novembre 1983.
- [Drapeau91] G.D. Drapeau et H. Greenfield, "*MAEStro : A distributed Multimedia Authoring Environment*", Proceedings of USENIX summer 1991, Juin 1991.
- [Furuta88] R. Furuta, V. Quint et J. André, "*Interactively Editing Structured Documents*", Electronic Publishing : Origination, Dissemination and Design, pp.19–44, Avril 1988.
- [Rossum93] G.Van Rossum et al. "*CMIFed: A presentation environment for portable hypermedia documents*". In Proc. First International Conference on Multimedia, pp.183–188, Anaheim, CA, Août 1993.
- [Ghafoor92] T.D.C. Little et A. Ghafoor, "*Interval–Based Conceptual Models for Time–Dependent Multimedia Data*", Workshop on Multimedia Informations Systems, Tempe Arizona, 7 Fevrier 1992.
- [VionDury94] Jean–Yves Vion–Dury et Miguel Santana, "*Virtual Images: Interactive Visualization of Distributed Object–Oriented Systems*", OOPSLA'94 Conference Proceedings, Vol 29(10), pp.65–84, October 1994.