DEA Informatique de Lyon 13 juillet 1999
PFE du Département Informatique de l'INSA de Lyon 13 septembre 1999

# Représentation et exploitation de graphes de description de documents multimédia dans un langage structuré

par
Elöd EGYED-ZSIGMOND

*Présentation du travail effectué pendant les stages DEA et de projet de fin d'étude au sein du LISI et du Département d'Informatique de l'INSA de LYON sous la direction de M. Jean-Marie PINON en collaboration avec Yannick Prié et Alain Mille*

**contacts :** Elod.Egyed@lisi.insa-lyon.fr, yannick.prie@insa-lyon.fr, am@cpe.fr,Jean-Marie.Pinon@lisi.insa-lyon.fr,



Plan

## ✔ 1) Introduction

2) Contexte de l'étude

3) Démarche et modèles proposés

4) Etude critique et validation expérimentale des modèles

5) Le prototype

6) Conclusion

## Introduction

Grandes quantités de documents multimédia numériques, qu'on veut exploiter de façon intelligente.

Il faut leur donner une **structure**

et un **cadre d'exploitation**.



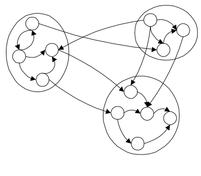
## Problématique

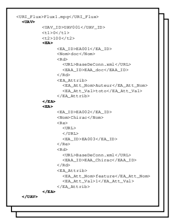
Un ***graphe réparti*** (STRATES-IA)

Des ***langages*** (XML et dérivés)

Un ***algorithme*** (Multipropagation)

⇒ Chercher un modèle de représentation du ***graphe*** dans les ***langages***, sur lequel l '***algorithme*** s'applique de façon optimale





• Représenter le graphe en XML pour partager et exploiter son contenu

Plan

*1) Introduction*

**✔ 2) Contexte de l'étude**

3) Démarche et modèles proposés

4) Etude critique et validation expérimentale des modèles

5) Le prototype

6) Conclusion

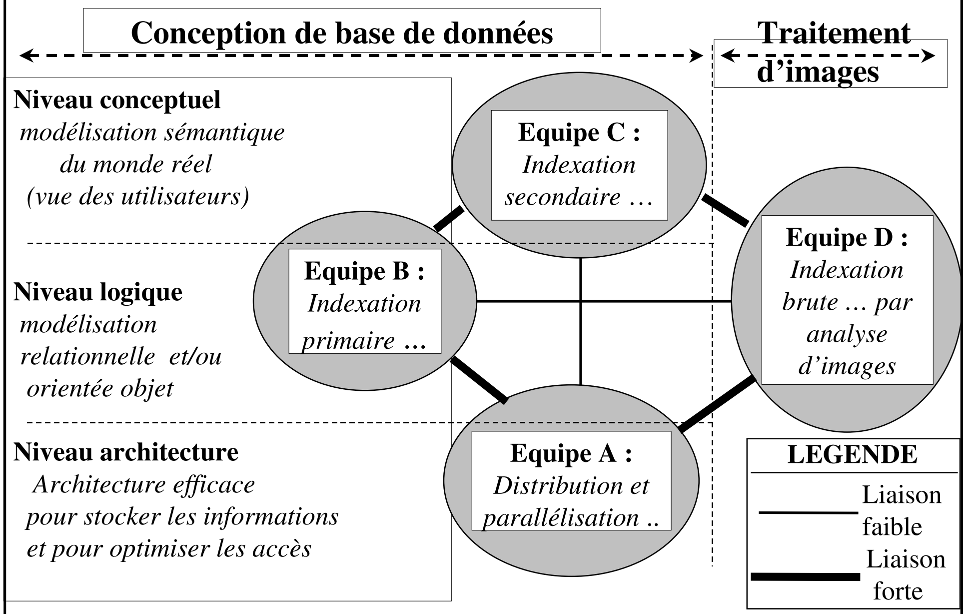


*Projet :*

Système
d 'Exploration
de Séquences
Audiovisuelles et
Multimédias enrichi par
l 'Expérience

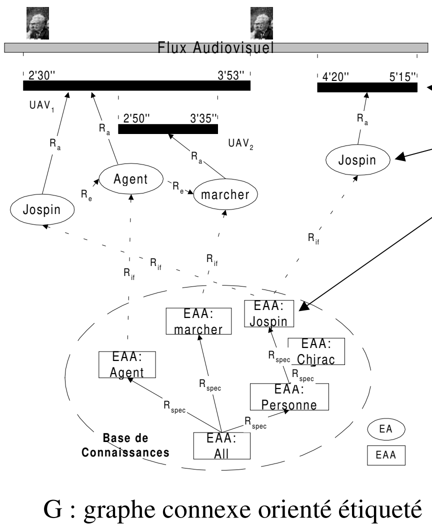
***Partenaires***

**CCETT - CNET**
**CPE**
**ENS-Lyon**
**INSA-Lyon**
**INA-Lyon**
**France-3-RAA**




Elöd EGYED-ZS: Représentation et exploitation de graphes de description de documents multimédia dans un langage structuré 7

Strates Interconnectées par les Annotations (STRATES-IA)




G : graphe connexe orienté étiqueté

Eléments (sommets)
- **UAV** unité audiovisuelle
- **EA** élément d'annotation
- **EAA** élément d'annotation abstrait

Relations (arcs)
- $\mathbf{R_a}$ relation d'annotation
- $\mathbf{R_e}$ relation élémentaire
- $\mathbf{R_{if}}$ relation d'inscription dans le flux
- $\mathbf{R_{spec}}$ relation de spécialisation

…

Exploitation :

requêtes : sous graphes potentiels
solutions : sous-graphes de G
méthode : recherche d'isomorphismes de sous-graphes partiels

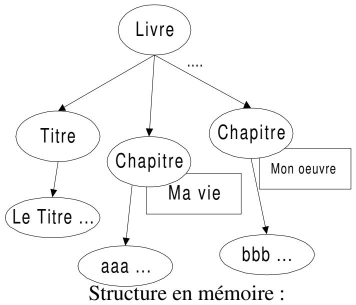
Elöd EGYED-ZS: Représentation et exploitation de graphes de description de documents multimédia dans un langage structuré 8

## XML : un langage de structuration de documents fondé sur le concept de balisage

- Créé pour être utilisé sur l’Internet
- Méta-langage

```
<!ELEMENT Titre (#PCDATA)
<!ELEMENT Chapitre (#PCDATA)*>
     <!ATTLIST Chapitre
              Soustitre CDATA>
<!ELEMENT Livre (Titre,Chapitre+)>
```

Modèle du document **DTD** (Document Type Definition)




Structure en mémoire :
arbre DOM (Document Object Model)

```
<Livre>
        <Titre>Le titre du livre</Titre>
        <Chapitre Soustitre= »Ma vie">
                aaa… contenu du chapitre 1
                …
        </Chapitre>
        <Chapitre Soustitre= »Mon oeuvre">
                bbb… contenu du chapitre 2
                …
        </Chapitre>
</Livre>
```

Instances du modèle (documents)



## Langages dérivés et complémentaires de XML

- Langages dérivés :
  - Schema for Objectoriented XML (SOX),
  - Resource Description Framework (RDF), ...
    (orienté objet, typage du contenu)

- Langages complémentaires
  - eXtended Link Language (XLL),
  - XML Stylesheet Language (XSL), …
    (transformation, liens)

- Langages de requêtes
  - XML-QL, XML-GL, ...

Représentation et exploitation des graphes en XML: état de l'art

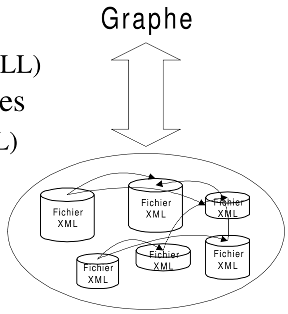
- Représentation
  - Représenter des sites Web (RDF)
  - Descripteurs de documents multimédia (DDL MPEG7)

    (*1 graphe dans 1 document XML*)

- Exploitation
  - Langages complémentaires (XSL, XLL)
  - Propositions de systèmes de requêtes

    (SGML-QL, XML-QL, XML-GL)

    (*requêtes sur un document*)






Plan

*1) Introduction*

*2) Contexte de l'étude*

## ✔ 3) Démarche et modèles proposés

4) Etude critique et validation expérimentale des modèles

5) Le prototype

6) Conclusion

## Démarche de l'étude

• Etude des syntaxes possibles

• Définition de critères de construction

• Proposition de modèles

• Définition de critères de comparaison

• Etude et validation des modèles proposés



## Critères de construction

• Syntaxe choisie
  (XML sans/avec attributs, XLL, RDF)

• Hiérarchie des éléments
  (encapsuler/mettre au même niveau des éléments)

• Représentation des liens
  (liens inclus/exclus)

⇨ Modèles créés en combinant les critères de construction

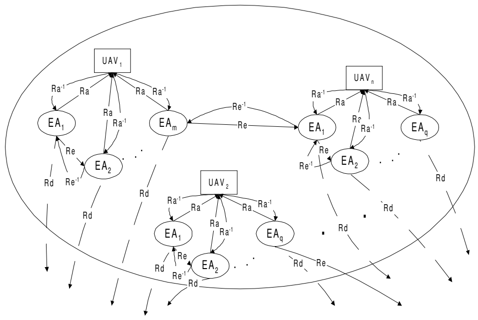
## M1: Modèle de base (définition)

DTD :

```
<!ELEMENT eBDA (URI_Flux,(UAV,EA+)+)>   avec :
<!ELEMENT UAV (UAV_ID,t1,t2,EA_ID+)>
<!ELEMENT EA (EA_ID,Nom,UAV_ID,Re*,Re-1*,Rd,EA_Attrib*)>

<!ELEMENT BaseDeConnaissance (EAA)+>  avec :
<!ELEMENT EAA (EAA_ID,Ri+,Ri-1*,EAA_Att_Def*,Rif*,Valence*)>
```

- XML standard
- Pas d'attribut
- Eléments non encapsulés
- Liens inclus

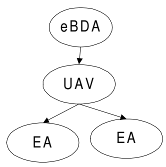




## M2, M3, M4 : Variantes du Modèle de base (M1)

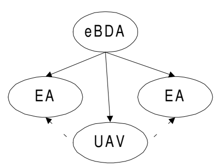
M2 : Sans attribut
éléments encapsulés

```
….
<!ELEMENT UAV (UAV_ID,t1,t2,EA+)>
<!ELEMENT eBDA (URI_Flux,UAV+)>
…
```




M3 : Avec attributs
éléments non encapsulés

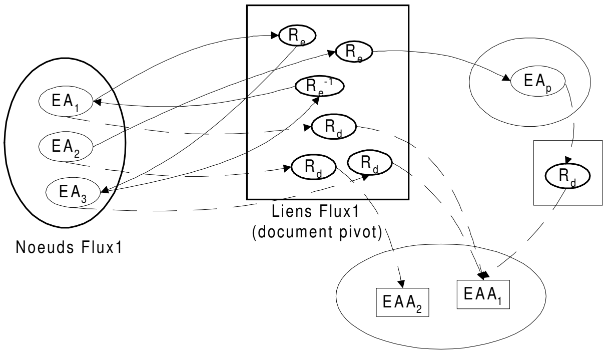
```
….
<!ELEMENT UAV (t1,t2)>
<!ATTLIST UAV
        id_uav ID #REQUIRED
        ref_EA IDREFS #REQUIRED>
<!ELEMENT eBDA (URI_Flux,(UAV,EA+)+)>
….
```




M4 : Avec attributs
éléments encapsulés

```
….
<!ELEMENT UAV (t1,t2,(EA+))>
<!ATTLIST UAV
        id ID #REQUIRED>
….
```

## M5, M6: modèles avec liens exclus




- M5 : XML standard
- M6 : syntaxe XLink



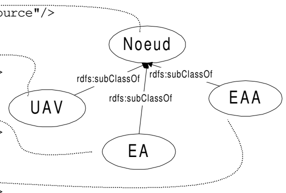
## M7: modèle en RDF

```
<rdfs:Class ID="Noeud">
     <rdfs:subClassOf rdf:resource="rdfs:Resource"/>
</rdfs:Class>

<rdfs:Class ID="UAV">
     <rdfs:subClassOf rdf:resource="#Noeud"/>
</rdfs:Class>

<rdfs:Class ID="EA">
     <rdfs:subClassOf rdf:resource="#Noeud"/>
</rdfs:Class>

<rdfs:Class ID="EAA">
     <rdfs:subClassOf rdf:resource="#Noeud"/>
</rdfs:Class>
```



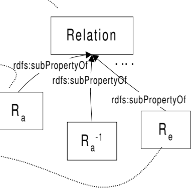

- Orienté Objet
- Variables typées

```
<rdf:Property ID="Relation">
     <rdfs:range rdf:resource="#Noeud"/>
     <rdfs:domain rdf:resource="#Noeud"/>
</rdf:Property>

<rdf:Property ID="Ra">
     <rdfs:subPropertyOf rdf:resource="#Relation"/>
     <rdfs:range rdf:resource="#EA"/>
     <rdfs:domain rdf:resource="#UAV"/>
</rdf:Property>

<rdf:Property ID="Re">
     <rdfs:subPropertyOf rdf:resource="#Relation"/>
     <rdfs:range resource="#EA"/>
     <rdfs:domain resource="#EA"/>
</rdf:Property>

…
```



Plan

*1) Introduction*
*2) Contexte de l'étude*
*3) Démarche et modèles proposés*

**✔ 4) Étude critique et validation expérimentale des modèles**

5) Le prototype

6) Conclusion



Critères de comparaison des propositions

| | |
|---|---|
| • Richesse sémantique : | Expressivité intrinsèque de la syntaxe<br>Partage et évolutivité |
| • Taille et complexité structurelle : | Espace occupé par les documents<br>Niveau d'encapsulation des éléments |
| • Requêtes : | Mise à jour complexe<br>Exploitation temporellement critique<br>(recherche des voisins) |
| • Lisibilité : | Clarté de lecture pour l'homme |

Récapitulation des caractéristiques

| | Richesse sémantique | Taille fichier | Requêtes de MAJ | Recherche | Lisibilité |
|---|---|---|---|---|---|
| **M1 : Modèle de base** | 7 | 4 | 3 | 6 | 4 |
| **M2 : Variante avec EA encapsulé** | 6 | 2 | 3 | 4 | 2 |
| **M3 : Modèle de base avec attributs** | 5 | 3 | 3 | 6 | 3 |
| **M4 : Variante attr. + EA encapsulé** | 4 | 1 | 3 | 4 | 1 |
| **M5 : Liens exclus** | 3 | 6 | 1 | 2 | 7 |
| **M5' : Liens exclus, EA encapsulé** | 2 | 5 | 1 | 1 | 6 |
| **M6 : Liens exclus (XLink)** | 1 | 8 | 2 | 3 | 5 |
| **M7 : RDF** | 2 | 7 | 4 | 5 | 5 |

Classement des modèles selon les différents critères

• Les différents modèles se révèlent optimaux pour différents types d'exploitation

• L'utilisation des attributs est recommandée

• Minimiser les accès séquentiels : fragments de petite taille,
relations implicites, liens exclus

• Modèle optimal pour nous : M5'



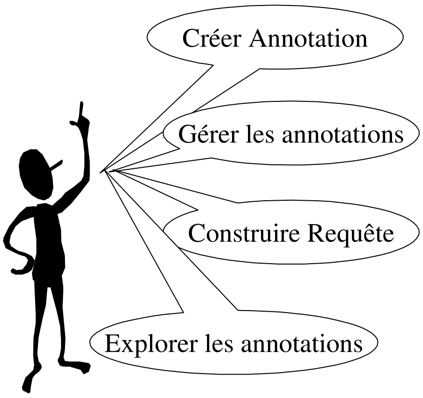
Plan

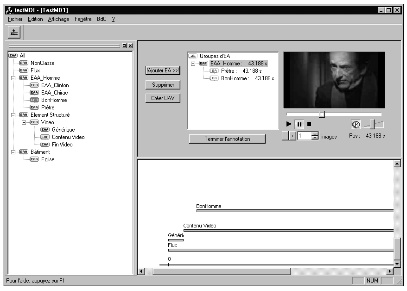
*1) Introduction*
*2) Contexte de l'étude*
*3) Démarche et modèles proposés*
*4) Étude critique et validation expérimentale des modèles*

**✔ 5) Le prototype**

6) Conclusion

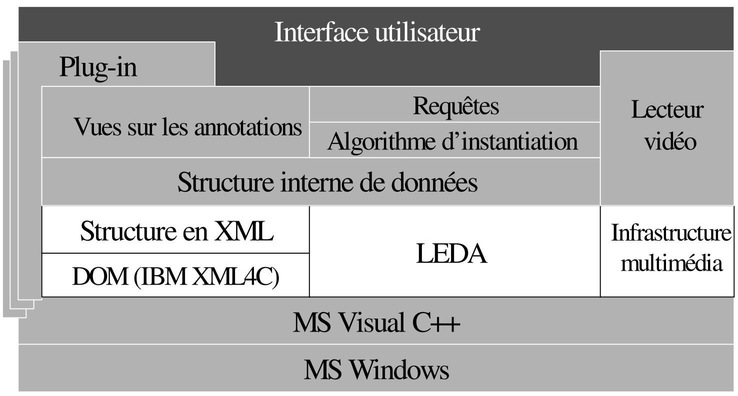
## Cas d'utilisation





## Architecture du prototype



Plan

*1) Introduction*
*2) Contexte de l'étude*
*3) Démarche et modèles proposés*
*4) Étude critique et validation expérimentale des modèles*
*5) Le prototype*

## ✔ 6) Conclusion



Conclusion

• Mise en place d'un cadre de construction et d'étude de modèles de représentation d'un graphe de connaissances dans un langage structuré

• Proposition et validation expérimentale de sept modèles de représentation répartie.

• Etude critique et comparative de ces modèles

• Soumission à la « Journée Jeune Chercheur » organisée par le GDR I3 (groupes «documents multimédias » et « écrit et documents »), Tours, 9 septembre 99

Perspectives : projet RNRT (RECIS)
MPEG7 DDL