

Méthodes de modélisation des connaissances

- ▶ Structuration des données en fonction des supports :
 - ▶ Base de données : SGBD
 - ▶ Fichiers : XML
- ▶ Langages de description de modèles : Merise, UML.

Méthodes de modélisation des connaissances

- ▶ Utilisation de types de données complexes.
- ▶ Nécessité de communiquer la composition de ces types :
 - ▶ Par le code : dépendance au langage de programmation, communication du source du programme, accès aux interfaces.
 - ▶ Par un schéma externe : problème de la maintenance et de la vérification du schéma.

Les Markup Languages

- ▶ Les Markup Languages (ML) sont des outils qui permettent d'exprimer des modèles de structuration des données.
- ▶ Un ensemble d'outils permet de lier plus ou moins strictement la description et le fichier de données.
- ▶ **Remarque** : ceci ne dispense nullement de réaliser le modèle de conception (avec UML ou un autre langage).

Bibliographie

- ▶ Comprendre XSLT par Bernd Amann et Philippe Rigaux - O'Reilly
- ▶ Schémas XML de Jean-Jacques Thomasson - Eyrolles
- ▶ Le site officiel : <http://www.w3.org/XML>
- ▶ Le site français : <http://xmlfr.org>
- ▶ Un tutoriel pour les débutants chez developpez.com : <http://xml.developpez.com/cours/>
- ▶ Tutoriel XPath : <http://jerome.developpez.com/xmlxsl/xpath/>

Le Langage XML

- ▶ XML pour *eXtensible Markup Language*.
- ▶ Ancêtre : GML (Generalised Mark-up Language) de Charles Goldfarb, lui-même inspiré par William Tunicliffe (1967) qui décrit le premier la séparation du contenu d'un document de sa présentation.
- ▶ En 1986 le standard SGML est établi, 3 ans avant la création du HTML et du Web.
- ▶ XML est développé en 1996 par Jon Bosak.
- ▶ Historiquement, l'apparition de XML dans la suite de SGML et HTML fait que les présentations de XML et HTML sont souvent liées mais dans les faits, il n'existe pas d'obligation de considérer XML dans le contexte du Web.

Différences entre XML et HTML

- ▶ Le XML est un langage descriptif.
- ▶ Originaire du SGML, il s'illustre lui aussi à l'aide de balise.
- ▶ La différence avec le HTML se situe au niveau de sa capacité à s'auto-structurer dans sa façon de décrire l'information. Alors que ce dernier se contente de formater une information pêle mêle.
- ▶ La balise XML décrit l'information qu'elle jalonne alors que le HTML détermine la façon de présenter l'information qu'il balise.

Format XML

- ▶ Un fichier XML est balisé par des TAGS ou mots clés qui fonctionnent comme un système de parenthésage mathématique.
- ▶ Ces mots clés peuvent être vus comme les éléments lexicaux et l'agencement ou composition que vous réalisez avec ces mots clés définit une sorte d'arbre syntaxique qui permet de reconstruire les instances des objets de votre programme.
- ▶ La notion d'arbre est ici essentielle, un document XML possède une racine, des branches de décomposition et des feuilles avec les valeurs des variables.

Les TAGS

- ▶ Un TAG est un mot - ou mot clé - encadré par les signes < > . L'ensemble des TAGS "*valides*" définit le langage de description.
- ▶ Chaque fichier XML doit commencer par un TAG entête :

```
<?xml version=" 1.0"  
      encoding=" ISO-8859-1" ?>
```

- ▶ TAG commentaire :

```
<!-- Description du reseau metabolique -->
```

Les TAGS

- ▶ TAG décrivant un noeud de l'arbre :

`<section>` Debut du noeud

`</section>` Fin du noeud

- ▶ TAG avec attribut :

`<Chapitre numero=" 1">`

NB : le TAG se referme sans citer l'attribut.

Les TAGS

- ▶ Si XML définit lui-même un ensemble de TAGS, il doit surtout être considéré comme un **meta langage**, c'est à dire un langage qui permet d'en définir d'autres.
- ▶ Les **attributs** :
 - ▶ Chaque TAG peut supporter des spécifications qui sont alors indiqués comme les valeurs des attributs de ces TAGS.
 - ▶ Les attributs autorisés pour chaque TAG doivent être précisés lors de la définition du langage.
- ▶ Les TAGS et les attributs ne sont pas prédéfinis, ce qui signifie que l'utilisateur est libre de créer les TAGS et les attributs qui lui sont nécessaires.

Définir une arborescence

- ▶ La structure d'un document se décompose en un préambule (entête) et un corps.
- ▶ Ce corps commence par le TAG racine de l'arbre et se terminera par le même TAG dans sa position *fermée*.
- ▶ Exemple :

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<journal nom="Linux à Gogo">
  <!-- Description du numero' ' -->
  <Chapitre numero="1">
    .....
  </Chapitre>
</journal>
```

Visualiser un fichier XML

- ▶ La plupart des navigateurs Web peuvent afficher correctement les données d'un fichier XML sous forme d'arbre.
- ▶ **Attention** : à l'heure actuelle le traitement des fichiers XML par les navigateurs n'est pas normalisé. Il est extrêmement fréquent d'obtenir des résultats différents suivant le navigateur utilisé.
- ▶ Exemple :

Déclaration de la structure du document

- ▶ Le fait que la définition de la structure du document et les données soient placées ensemble dans le fichier XML est un des reproches majeurs qui est fait à XML.
- ▶ DTD et Schémas
 - ▶ Il est possible de séparer ces deux informations et de spécifier la structure dans un *Document Type Definition* **DTD** ou bien dans un *Schéma* **xsd**.

Caractéristiques des DTD

- ▶ Un DTD décrit la structure logique du document.
- ▶ C'est un ensemble de règles ou contraintes que tout document qui déclare ce DTD doit respecter afin d'être considéré comme bien formé.
- ▶ **Remarque** : un document qui ne déclare pas de DTD est considéré bien formé par défaut.
- ▶ La présence d'un DTD permet d'exporter la structure d'un document à l'extérieur de celui-ci.
- ▶ Lorsque deux documents déclarent le même DTD il est possible de garantir qu'ils respectent la même syntaxe.

Caractéristiques des DTD

- ▶ Déclaration d'un fichier DTD :

```
<!DOCTYPE mesDTD 'mesDTD.dtd'>
```

- ▶ La vérification des règles du DTD se fait au moyen d'un *parser* XML.
- ▶ Un parser a un rôle assez semblable à celui d'un compilateur :
 - ▶ Analyse syntaxique du programme et signalement des erreurs.
 - ▶ Traduction du programme en langage machine - remplacement des entités par leur valeur.
 - ▶ Construction de l'arbre syntaxique - mais pas de traitement dans le cas du parser.

Caractéristiques des DTD

- ▶ Les DTD ont leur propre langage :

- ▶ Déclarer un élément :

```
<!ELEMENT nom du champs (type du champs)>
```

- ▶ Exemple:

```
<!ELEMENT journal (#PCDATA)>
```

- ▶ Fichier XML de données correspondant :

```
<journal> Linux a Gogo </journal>
```

Caractéristiques des DTD

- ▶ Il est possible de déclarer une liste d'éléments composites comme suit :

```
<!ELEMENT journal (nom, adresse , numero)>
```

- ▶ Liste d'attributs :

```
<!ATTLIST champs type_attribut valeur>
```

```
<!ATTLIST journal type_journal #PCDATA>
```

- ▶ Fichier XML correspondant :

```
<journal type_journal=' 'mensuel' '>
```

```
  <nom> Linux à Gogo </nom>
```

```
    <adresse> Paris </adresse>
```

```
    <numero> 25 </numero>
```

```
</journal>
```

Caractéristiques des DTD

- ▶ Déclaration des entités : une entité est un type de données qui permet de faire référence à un autre élément du DTD en tant que type de l'élément du TAG.

```
<!ENTITY nom valeur>
```

```
<!ENTITY MENS 'Mensuel'>
```

```
<!ATTLIST journal type_journal ENTITY #REQUIRED>
```

- ▶ Fichier XML correspondant :

```
<journal type_journal='MENS'>
```

```
  <nom> Linux a Gogo </nom>
```

```
</journal>
```

- ▶ Il est possible de déclarer des entités externes c.-à-d. décrites dans un autre fichier.

```
<!ENTITY MENS SYSTEM 'entites.xml'>
```

Limites et Inconvénients des DTD

- ▶ Les DTD sont écrits dans une syntaxe particulière différente de la syntaxe XML.
- ▶ Un seul type primitif PCDATA défini.
- ▶ Pas de possibilités de poser des contraintes sur les données ; nombre précis d'occurrences (seule la forme * ou 0..n existe), format (date, longueur de chaîne).
- ▶ Depuis mai 2001 le W3C préconise donc de remplacer les DTD par des **schémas** XML.

Les Schémas

- ▶ Contrairement aux DTD, les schémas permettent de décrire l'imbrication et l'ordre d'apparition des éléments et de leurs attributs soit une **grammaire** d'un langage particulier.
- ▶ Un **schéma** XML est un fichier écrit en XML.
- ▶ Si un **schéma** existe le fichier de données devra contenir la déclaration de ce *schéma*.

Exemple :

```
<journal
  xmlns:xsi=
    '' http://www.w3.org/2001/XMLSchema-instance ''
  xsi:noNamespaceSchemaLocation=''journal.xsd''>
  <nom> Linux à Gogo </nom>
</journal>
```

Description d'un Schéma

- ▶ Les descriptions utilisent un nouveau TAG `xsd`.
- ▶ L'élément racine du *schéma* est de type `xsd:schema`.
- ▶ Pour décrire un élément on donne son *nom*, son *type*, et si nécessaire les liens ou contraintes qui s'appliquent.
- ▶ Pour décrire un élément on donne son *nom*, son *type*, et si nécessaire les liens ou contraintes qui s'appliquent.

Exemple

```
<xsd:element name=' 'journal ' '>
  <xsd:complexType>
    <xsd:all>
      <xsd:element name=' 'nom ' '
                    type=' 'xsd:string ' '/>
      <xsd:element name=' 'adresse ' '
                    type=' 'xsd:string ' '/>
      <xsd:element name=' 'numéro ' '
                    type=' 'xsd:positiveInteger ' '/>
      <xsd:element name=' 'type_journal ' '
                    type=' 'type-j ' '/>
    </xsd:all>
  </xsd:complexType>
</xsd:element>
```

Exemple

```
<xsd:simpleType name=' 'type_j' '/>  
  <xsd:enumeration value=' 'mensuel' '/>  
  <xsd:enumeration value=' 'hebdomadaire' '/>  
  <xsd:enumeration value=' 'journalier' '/>  
</xsd:simpleType>
```

Les Markup Languages dédiés

- ▶ Pour les mathématiques : MathML

- ▶ Documentation : <http://www.w3.org/TR/REC-MathML/>
- ▶ Exemple : codage de $a - b$

```
<apply>  
  <minus/>  
  <ci>a</ci>  
  <ci>b</ci>  
</apply>
```

Exemple MathML

- Plus complexe : $x^2 + 4x + 4 = 0$

<mrow>

<mrow>

<msup>

<mi>x</mi>

<mn>2</mn>

</msup>

<mo>+</mo>

<mrow>

<mn>4</mn>

<mo>⁢</mo>

<mi>x</mi>

</mrow>

<mo>+</mo>

<mn>4</mn>

</mrow>

<mo>=</mo>

<mn>0</mn>

</mrow>

Les Markup Languages dédiés

- ▶ Echange d'Objets : XMI - XML Metadata Interchange Specification
 - ▶ Documentation :
http://www.omg.org/news/pr99/xmi_overview.html
 - ▶ Ce format est utilisé pour décrire des objets - par exemple des diagrammes de classes UML.
 - ▶ Les logiciels ArgoUML et Poseidon peuvent utiliser le format XMI.
- ▶ Les graphiques : Precision Graphics Markup Language PGML et Scalable Vector Graphics - SVG

Les Markup Languages dédiés

- ▶ Exemple PGML:

```
<?xml version="1.0"?>
<!DOCTYPE pgml SYSTEM "pgml.dtd">
<pgml>
  <group fillcolor="red">
    <rectangle x="100" y="100" width="100" height="100" />
    <rectangle x="300" y="100" width="100" height="100" />
  </group>
  <group fillcolor="blue">
    <rectangle x="100" y="300" width="100" height="100" />
    <rectangle x="300" y="300" width="100" height="100" />
  </group>
</pgml>
```

Les Markup Languages dédiés

- ▶ Exemple SVG

<http://www.carto.net/papers/svg/samples/>

```
<svg width="500" height="300" >
<clipPath id="myClip">
  <path d=' 'M 0.0 112 L 20 124 L 40 129
        L 60 126 L 80 120 L 100 111
        ...
        L 61 179 L 42 171 L 30 159
        L 13 140 L 00 112 Z' '/>
</clipPath>
```

Les Markup Languages dédiés

```
<g id="grid"  
  style="stroke:none; clip-path:url(#myClip)">  
  <rect style="fill:red" x="0" y="0"  
    width="500" height="20" />  
  ...  
  <rect style="fill:yellow" x="0" y="260"  
    width="500" height="20" />  
</g>  
</svg>
```

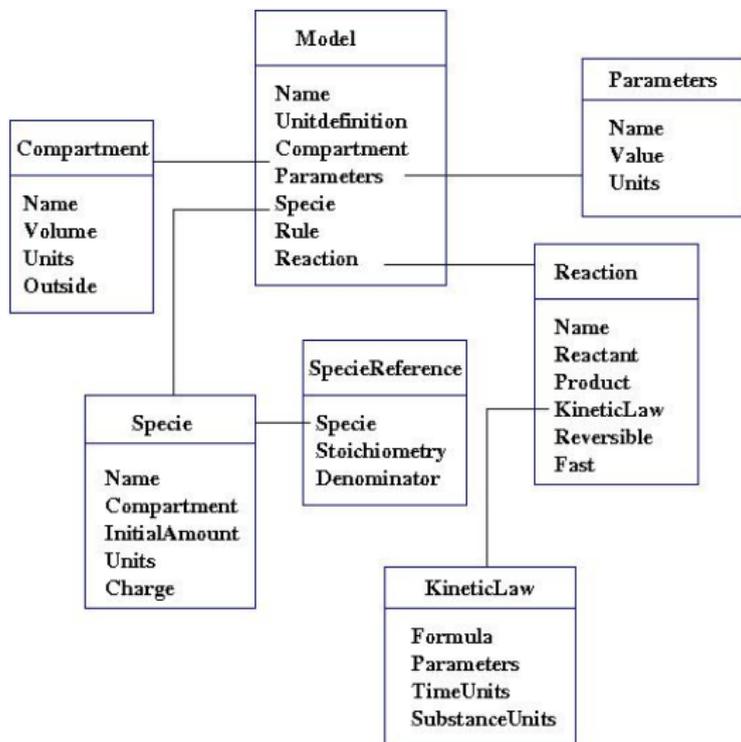
Les Markup Languages dédiés

- ▶ Pour la biologie :
 - ▶ Systems Biology Markup Language - SBML - maintenu par *caltech*
 - ▶ Documentation : <http://www.sbml.org>
 - ▶ Schema xsd du niveau 2 sur le site de sourceforge
 - ▶ CellML - maintenu par l'Université d'Auckland
 - ▶ Site web : <http://cellml.sourceforge.net/>

SBML

- ▶ Dédié à la modélisation - principalement en biochimie.
- ▶ Approche un modèle Objet.
- ▶ Permet la saisie d'équation, d'unité.
- ▶ Nombreuses connexions avec des bases de données -
Biomodels - et logiciels - copasi, cytoscape, cell designer.

SBML - les briques de base



SMBL et BioPaX, ontologies ...

- ▶ BioPaX couvre plus largement la structuration de connaissances en biologie.
- ▶ Les projets BioCyC, Reactom ... proposent un nomenclature plus proche des ontologies.
- ▶ Mapping possible entre les différents langages.

Une application pour la gestion de projet

- ▶ ANT :
 - ▶ Projet Apache-Jakarta
<http://jakarta.apache.org/ant/index.html>
 - ▶ Fichier de configuration en XML.
 - ▶ Fourni un certains nombres de tâches écrites en Java qui de se fait sont indépendantes du système.
 - ▶ Il est possible d'y ajouter ses propres tâches.
- ▶ `ant` recherche par défaut un fichier `build.xml` dans le répertoire courant.
- ▶ Ce fichier contient un ensemble de cible, par exemple `clean`
- ▶ Le tag racine est le tag `project`.
- ▶ le fichier `build.xml` qui se trouve dans le repertoire web est un exemple d'un tel fichier.

Utiliser des fichiers XML

- ▶ Les fichiers XML ne sont pas destinés à être analysés *“manuellement”*, entre autre à cause de leur verbosité.
- ▶ Pour relire un fichier XML, le programme utilisé doit connaître la syntaxe et les règles du langage XML et s’il existe, le DTD ou le schéma spécifique défini pour l’application.
- ▶ Un parser est un programme qui sachant un grammaire interprète les différents éléments contenu dans un fichier de données.

XSL

- ▶ Standard du W3C - Version 1.0 - basé sur le standard DSSSL (Document Style Semantics and Specification Language), norme ISO 1996.
- ▶ XSL est un langage de transformation de document : trier un document, extraire uniquement certaines informations.
- ▶ 2 étapes : transformation par XSLT du document XML en un autre document XML, puis mise en forme.
- ▶ C'est dans ce 2ième temps que l'on peut choisir le mode de présentation, par exemple HTML.
- ▶ La transformation est opérée par un processeur **XSLT**

XSLT - XPath

- ▶ Version 1.0 - novembre 1999.
- ▶ Les navigateurs implémentent plus ou moins XSI/XSLT.
- ▶ Il est nécessaire d'implémenter XPath pour utiliser XSLT.
- ▶ XPath est un langage de *pattern matching* qui travaille à partir de l'arbre syntaxique obtenu par la transformation XSLT du fichier..

Syntaxe XSLT

- ▶ Balise de déclaration :

```
<?xml version='1.0' encoding='ISO-8859-1'?>  
  <xsl:stylesheet xmlns:xsl='http://www.w3.org/  
    1999/XSL/transform' version='1.0'>  
    <xsl:template match='/liste_clubs'>  
  
  </xsl:template>  
</xsl:stylesheet>
```

Les Feuilles de Style

- ▶ **XSL** (*eXtensible Style sheet Language*) : feuille de style pour XML.
- ▶ Objectifs :
 - ▶ factoriser les attributs d'affichage,
 - ▶ améliorer la lisibilité d'un document,
 - ▶ séparer l'aspect graphique du fichier qui contient les données.
 - ▶ possibilité d'adapter l'affichage aux périphériques : écran, imprimante etc
- ▶ Dans le principe très proche des CSS de HTML.

XPath

- ▶ Un langage différent de XML pour adresser des parties de document.
- ▶ Une syntaxe compacte à utiliser dans le XML.
- ▶ XPATH opère sur la structure logique du document pas sur la syntaxe.
- ▶ La documentation se trouve à <http://xmlfr.org/w3c/TR/xpath/>

XPath - Principes

Une expression est composée de 3 parties :

- ▶ Axe : `node-set` - ensemble non-ordonné de noeuds sans doublons, permet de se déplacer dans l'arbre des noeuds xml
Par exemple : `Child`, `parent`, `ancestor`, `self`....
- ▶ Test : Booléens : `true/false`, permet d'identifier le type du noeud recherché, il suit l'axe et doit être précédé de `::`
Par exemple : `parent::*`, `child::comment()` etc....
- ▶ Prédicat : condition sur les noeuds parcourus, il est encadré par `[]`
Le prédicat est facultatif.

XPath - exemple

```
<employees>
<!-- l'attribut num du noeud employe (type element)
      a comme valeur 1 -->
<employe num="1">
<nom>Dupont</nom>
<prenom>Jeanne</prenom>
<sexe>F</sexe>
<service>ressources humaines</service>
<tel>0645</tel>
<!--poste interne-->
<date_arrivée>03-04-2004</date_arrivée>
<!--arrivée dans la société-->
</employe>
```

XPath - les types d'objets

- ▶ Lorsqu'elle est évaluée toute expression XPath prend un des 4 types :
 1. Collections de noeuds : `node-set` - ensemble non-ordonné de noeuds sans doublons
 2. Booléens : `true/false`
 3. Réels : `number`
 4. Chaines de caractères : `string`

XPath - les types d'objets

- ▶ Exemple d'utilisation de XPath dans XSLT :
 - ▶ `child::*` résultat de type node-set
 - ▶ `child::nom=''` résultat de type booléen
 - ▶ `number(5 div 2)` résultat de type number
 - ▶ `string(child::nom)` résultat de type string - valeur du premier noeud fils nom du noeud courant.

Sélection

- ▶ Sélection à partir du nom d'un type d'élément :

```
<xsl:value-of select="nom_element" />
```

- ▶ L'utilisation du caractère / permet de définir le chemin d'accès
- ▶ Utilisation du caractère * : exemple */paragraph sélectionne tous les noeuds paragraph quelque soit leur ascendant.
- ▶ L'expression section//paragraph permet la recherche de tous les noeuds paragraph descendants du noeud selection directement descendant du noeud courant.
- ▶ Les caractères . et .. sont utilisés comme d'habitude.

Expressions

- ▶ Sélection d'attribut :
 - ▶ `section[@titre]` sélectionne les éléments `section` qui ont un attribut `titre`.
 - ▶ `section[@titre="Introduction"]` sélectionne les éléments `section` dont l'attribut `titre` a pour valeur `Introduction`.
- ▶ Opérateurs arithmétiques : `+` `-` `*` *Div* *mod*
- ▶ Opérateurs logiques :
 - ▶ Comparaison : `=` `<=` `>=` `>` `<`
 - ▶ Les opérateurs `and`, `or` et `not` peuvent être utilisés : Exemple de sélection des noeuds `section` n'ayant pas d'attribut `titre`
`section [not (@titre)]`

Expressions

- ▶ Chemin de localisation des noeuds : définition de l'axe, du noeud et de 0 à n prédicats pour spécifier la recherche :
- ▶ Quelques **axes** :
 - ▶ **child** : contient les enfants directs du noeud contextuel.
 - ▶ **descendant** : contient les descendants du noeud contextuel.
Un descendant peut être un enfant, un petit-enfant...
 - ▶ **parent** : contient le parent du noeud contextuel, s'il y en a un.

Expressions

- ▶ Quelques **axes** (suite):
 - ▶ `ancestor` : contient les ancêtres du noeud contextuel. Cela comprend son père, le père de son père... Cet axe contient toujours le noeud racine, excepté dans le cas où le noeud contextuel serait lui-même le noeud racine.
 - ▶ `attribute` : contient les attributs du noeud contextuel ; l'axe est vide quand le noeud n'est pas un élément.
 - ▶ `self` : contient seulement le noeud contextuel.

Expressions

- ▶ Quelques **axes** (suite):
 - ▶ **preceding (following)** : tous les noeuds précédants (suivants) le noeud courant dans l'ordre de déclaration, quelque soit le degré d'éloignement de la racine.
 - ▶ **preceding-sibling (following-sibling)** : les noeuds frères du noeud courant le précédant (le suivant) dans l'ordre de déclaration.

Expressions

- ▶ Quelques exemples :
 - ▶ `child::*` sélectionne tous les éléments enfants du noeud contextuel.
 - ▶ `child::nom=''` sélectionne tous les éléments enfants dont le nom est vide.
 - ▶ `child::text()` : sélectionne tous les noeuds de type texte du noeud contextuel.
 - ▶ `attribute::name` : sélectionne tous les attributs name du noeud contextuel.

Fonctions usuelles

- ▶ `number()` : évalue une expression arithmétique et la convertit en `number`
- ▶ `round()` : arrondit un `number`
- ▶ `sum()` : retourne un `number` résultat de la somme de toutes les valeurs des noeuds d'un `node-set`.
- ▶ `count()` : retourne un `number` dont la valeur est le nombre de noeuds d'un `node-set`
- ▶ `position()` : retourne un `number` représentant la position du noeud courant dans le `node-set` courant.
- ▶ `last()` : retourne un `number` représentant la position du dernier noeud du `node-set` courant.

Fonctions usuelles

- ▶ Traitement des chaînes de caractères :
 - ▶ `string()` : évalue un objet XPath de n'importe quel type et le convertit en `string`.
 - ▶ `contains(arg1,arg2)` : retourne `true` s'il existe une occurrence du deuxième argument dans le premier, `false` sinon.

Fonctions usuelles

- ▶ Sélection des noeuds :
 - ▶ `comment ()` : sélectionne tous les noeuds commentaires fils du noeud courant.
 - ▶ `text()` : sélectionne tous les noeuds fils du noeud courant, ne contenant que du texte.
 - ▶ `node()` : sélectionne tous les noeuds fils du noeud courant.
 - ▶ `id('identifiant')` : sélectionne l'élément normalement unique, qui a un attribut de type ID valant "identifiant".