

# IFT 1005

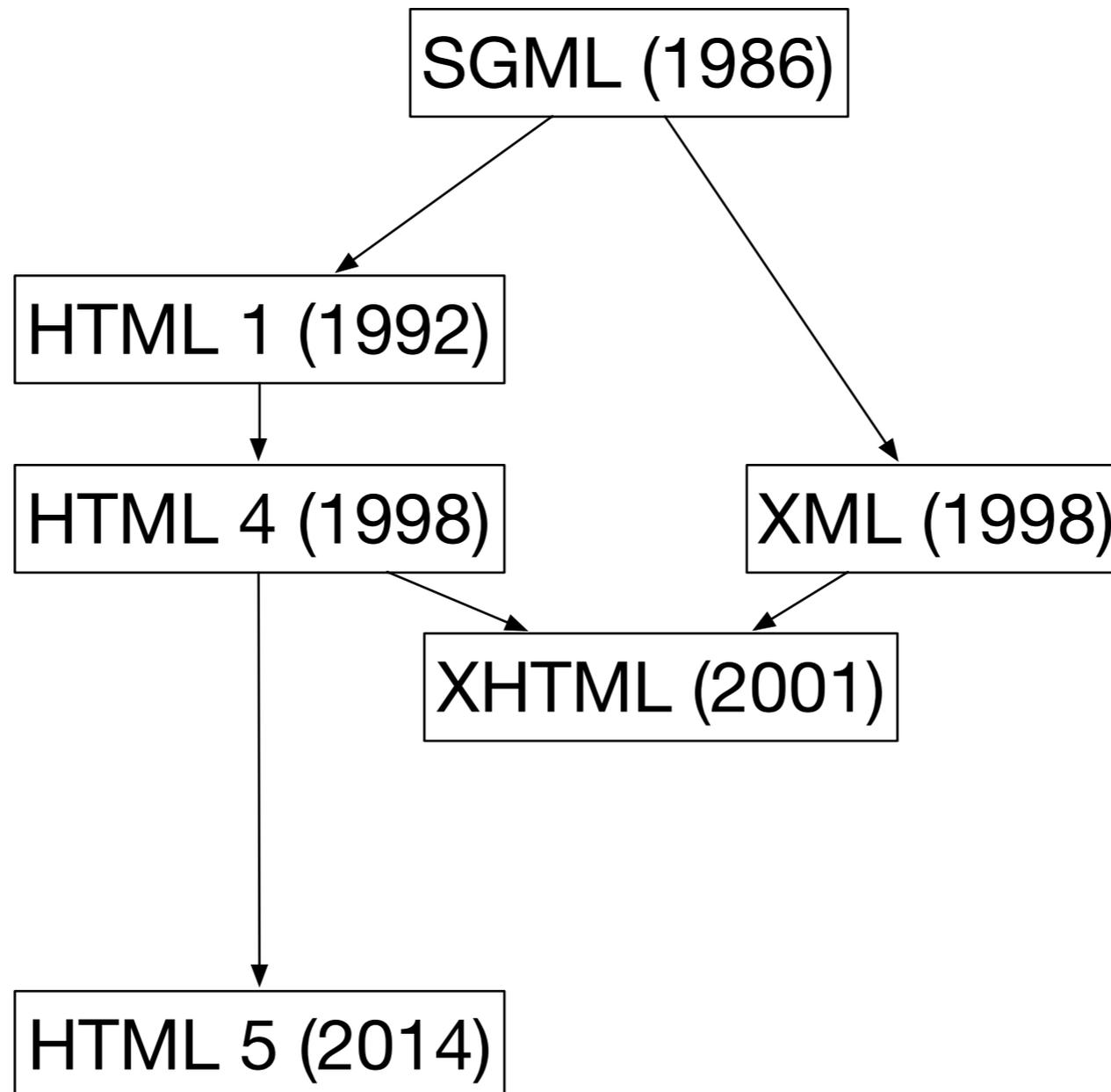
# Langages de balisage

Guy Lapalme  
Université de Montréal

# Balisage

- Enrichissement d'un document pour en indiquer
  - structure
  - présentation
  - traitement
- Via des balises (tags)
  - distinguables du texte
    - `<...>` en XML/HTML/MathML/SVG ...
    - `\...` en LaTeX
    - `{\...}` en RTF

# Évolution



# Balisage XML

- Information **à propos** du texte intégrée **dans** le texte
- Balise
  - entre < et >
  - nom de la balise et autres informations associées (attributs)

```
<auteur naissance="1802" mort="1885">  
    <nom>Hugo</nom>  
    <prenom>Victor</prenom>  
</auteur>
```

# Balisage sémantique

- balise est utilisée pour son **sens** pas son apparence
- noms arbitraires
- idéalement évocateurs
  
- Orienté données ou document

# XML orienté données

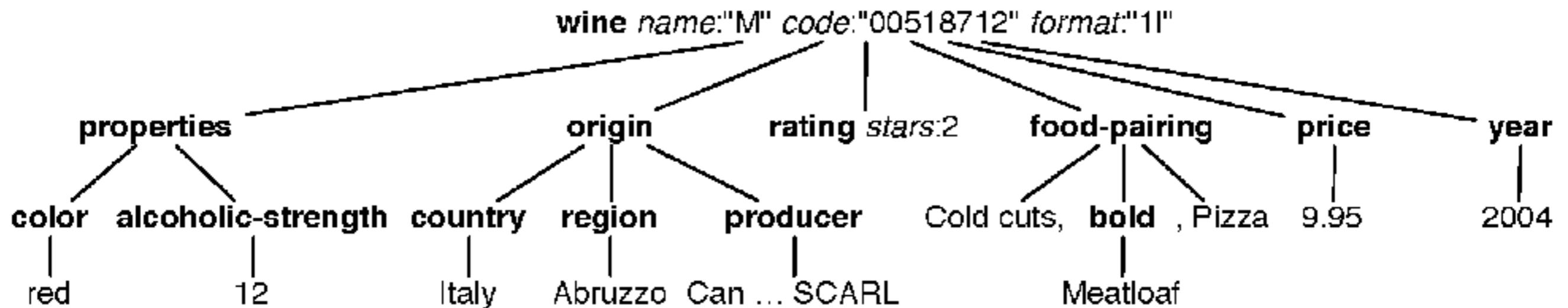
## *data oriented*

- Chaque information est étiquetée
- Ignore les blancs et espaces entre les balises

```

<?xml version="1.0" encoding="UTF-8"?>
<wine name="M" code="00518712" format="11">
  <properties>
    <color>red</color>
    <alcoholic-strength>12</alcoholic-strength>
  </properties>
  <origin>
    <country>Italy</country>
    <region>Abruzzo</region>
    <producer>Cantina Miglianico SCARL</producer>
  </origin>
  <rating stars="2" />
  <food-pairing>Cold cuts, <b>Meatloaf</b>, Pizza</food-pairing>
  <price>9.95</price>
  <year>2004</year>
</wine>

```



# XML orienté document

- Chronologie d'auteurs
  - Organisation de textes d'auteurs par période
  - Extraits de textes libres de droit
    - prose
      - suite de paragraphes
    - poésie
      - groupes de lignes imbriqués

## Exemple

```
<?xml version="1.0" encoding="UTF-8"?>
<auteur id="esope">
  <nom>Ésope</nom>
  <mort>-0560</mort>
  <prose>
    <head>La Cigale et les Fourmis</head>
    <p>C'était l'hiver [...] <q>Pourquoi</q> n'as-tu pas
      fait de provisions pendant l'été ? </p>
    <p>Je n'étais pas oisive, dit-elle, je chantais en artiste.</p>
    <p>Ah l'été, tu étais musicienne repartirent les fourmis en riant ;
      en hiver fais-toi danseuse.</p>
  </prose>
</auteur>
```

## Complétez l'arbre!

# XML : Éléments

- Balise d'ouverture doit avoir une balise de fermeture correspondante
  - `<elem>...</elem>`
  - `<elem/> = <elem></elem>`
- Éléments bien imbriqués
  - ~~`<a><b></a></b>`~~
- Un seul élément racine
- Nom d'un élément
  - débute avec une lettre ou `_`
  - peut contenir aussi chiffres `- . :`
  - peut pas commencer par `xml`
  - noms dépendent de la casse

# XML : Attributs

- Nom: même règles que pour les éléments
- dans le tag d'ouverture
- pas deux attributs de même nom
- valeurs entre ' ou "

# XML : Données

- Caractères réservés

<	:	&lt;
>	:	&gt;
"	:	&quot;
'	:	&apos;
&	:	&amp;

# XML :Autres

- Déclaration xml
  - `<?xml version="1.0" encoding="utf-8"?>`
  - mettre `.xml` comme extension de fichier
- Processing instructions
  - `<?xml-stylesheet type="text/css" href="..."?>`
- Commentaires
  - `<!-- ... -->`
  - pas de `--` dans les commentaires

# Vérifications d'un fichier XML

- Bonne formation
  - correspond à une structure d'arbre
- Validation
  - selon un schéma (grammaire d'arbre)
  - imbrication correcte des balises
  - contenu texte correspondant au bon type

# Pourquoi XML ?

- Ancêtre de HTML
- Remplacé par JSON dans certaines applications *orientées données*
- Toujours utilisé pour baliser les données textuelles
- **Validation** très importante