

Philippe GENOUD (LIG-STeamer)
Philippe.Genoud@imag.fr

M2CCI 2023-2024
cours PLAI-TW (Technologies du Web)

Introduction à HTML

dernière modification : 28/09/2023 10:19



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Qu'est ce que HTML ?

- **H**yper **T**ext **M**arkup **L**anguage
 - *markup langage* : langage de balisage
- HTML **N'EST PAS** un langage de programmation
- Langage pour créer des pages web
- Brique de base du Web

Qu'est-ce qu'une page web ? *

* d'après cours M1105 (Web et documents numériques) DUT Informatique UGA – Sylvie Pesty

- une page web est un document numérique
 - composé de
 - textes, sons, vidéos, ...
 - liens hypertextes vers d'autres documents permettant de passer du document à un autre
 - consultable à l'aide d'un navigateur (browser)
 - firefox, chrome, safari, edge,

World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an [executive summary](#) of the project, [Mailing lists](#), [Policy](#), November's [W3 news](#), [Frequently Asked Questions](#).

[What's out there?](#)

Pointers to the world's online information, [subjects](#), [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#), [X11 Viola](#), [NeXTStep](#), [Servers](#), [Tools](#), [Mail robot](#), [Library](#).)

[Technical](#)

Details of protocols, formats, program internals etc

[Bibliography](#)

Paper documentation on W3 and references.

[People](#)

A list of some people involved in the project.

[History](#)

A summary of the history of the project.

[How can I help ?](#)

If you would like to support the web..

[Getting code](#)

Getting the code by [anonymous FTP](#), etc.



<http://info.cern.ch/hypertext/WWW/TheProject.html>

Page d'accueil du 1^{er} site (ensemble de pages) web mis en ligne par le CERN en décembre 1990

source : The birth of the web

<https://home.cern/science/computing/birth-web>

Rôle de HTML

- Dans un document numérique, et donc une page web, il faut bien séparer * :
 - l'aspect **information**, relatif au **contenu** (les textes, images, sons, vidéos, liens) et à la **structuration** (l'organisation) de cette information
 - l'aspect **présentation** de l'information, relatif à la **forme**, car plusieurs «formes» sont possibles (sur grand écran, smartphone, imprimante,...)

* cours M1105 (Web et documents numériques) DUT Informatique UGA – Sylvie Pesty

- HTML le langage de base pour décrire des pages web
 - en fait plusieurs langages peuvent être présents dans les pages web



HyperText Markup Language

Pour décrire et structurer les informations de la page



Cascading Style Sheets

Pour présenter les informations de la page (couleurs, encadrés, taille du texte, positionnement des divisions, des images,...)

© UGA-2023 Philippe GENOUD









Javascript

Langage de programmation qui permet d'ajouter un comportement (animation, gestion de l'interaction...) aux pages web





Créer une première page web

- de quoi avons nous besoin ?

Un éditeur de texte (n'importe lequel)

- Notepad++ (Windows) 
- TextMate (Mac) 
- Sublime Text 
- Atom.io 
- Brackets 
- **Visual Studio Code** 
- ...

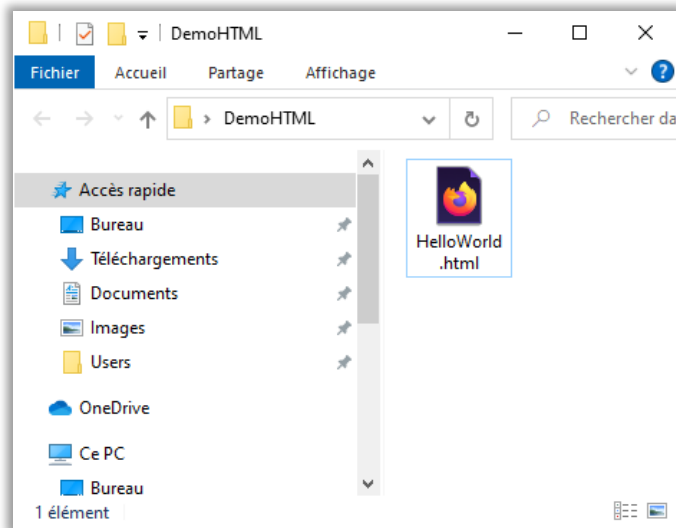
Un navigateur (n'importe lequel)

- **Mozilla Firefox** 
- **Google Chrome** 
- Safari 
- Edge (mais pas IE !) 
- ...

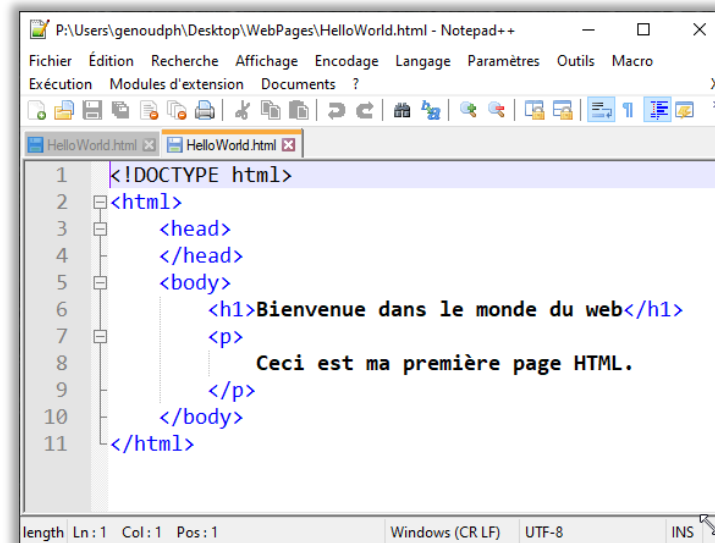
<https://survey.stackoverflow.co/2023/#section-most-popular-technologies-integrated-development-environment>

Créer une première page web

Permet d'actualiser (recharger)
la page courante(Ctrl + R)

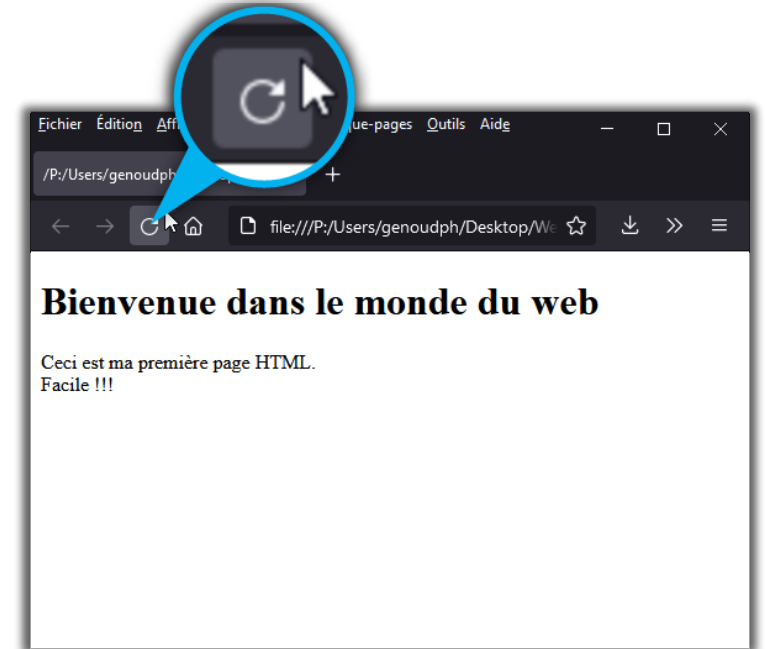


1) Créer un fichier texte
avec l'extension **.html**



2) le modifier avec
l'éditeur de votre choix

Recommandation : utilisez VS Code
pour plus d'information sur l'édition
de code HTML avec VS Code voir
[cette page Web](#)



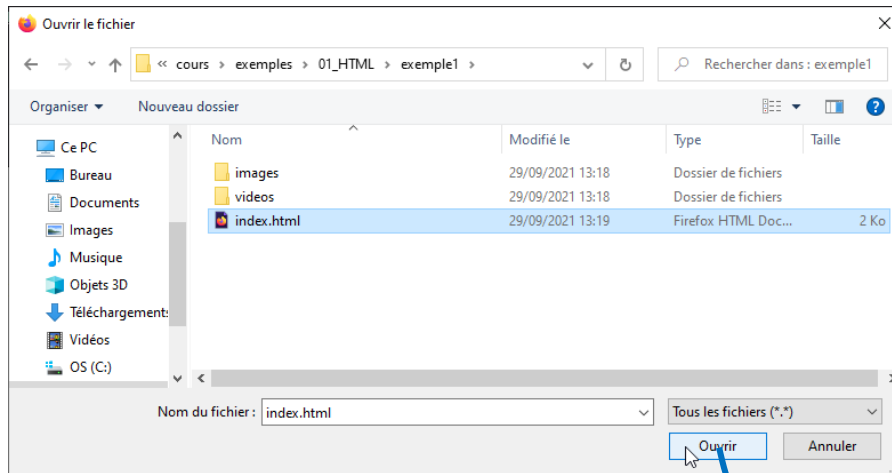
3) le visualiser avec le
navigateur de votre choix

Recommandation : si vous utilisez VS Code,
l'extension [Live Server](#) permet de rafraichir
automatiquement la page dans le navigateur
dès que le fichier HTML source est
sauvegardé.

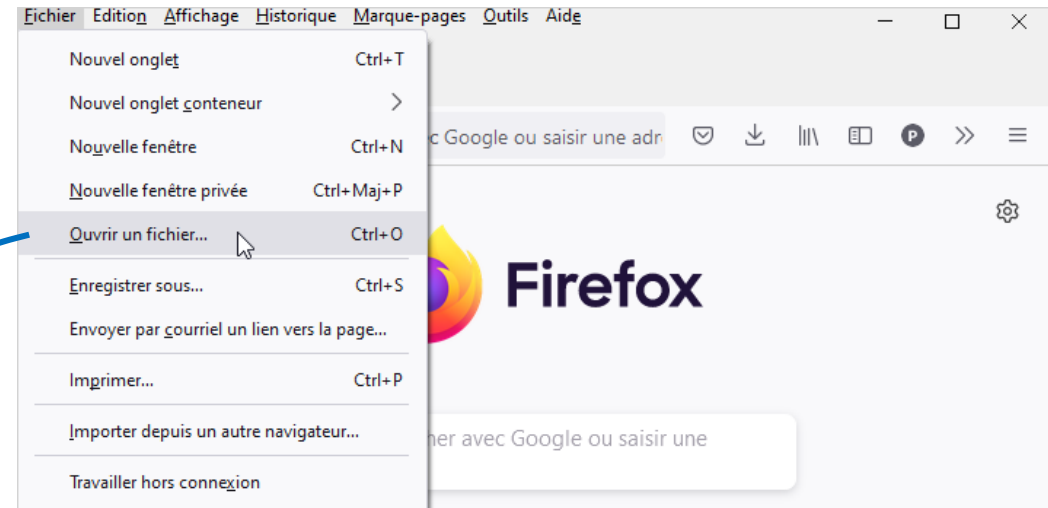


Consulter une page web

- consultation locale :
 - la page est dans un répertoire de votre PC : ouvrir la page depuis le navigateur



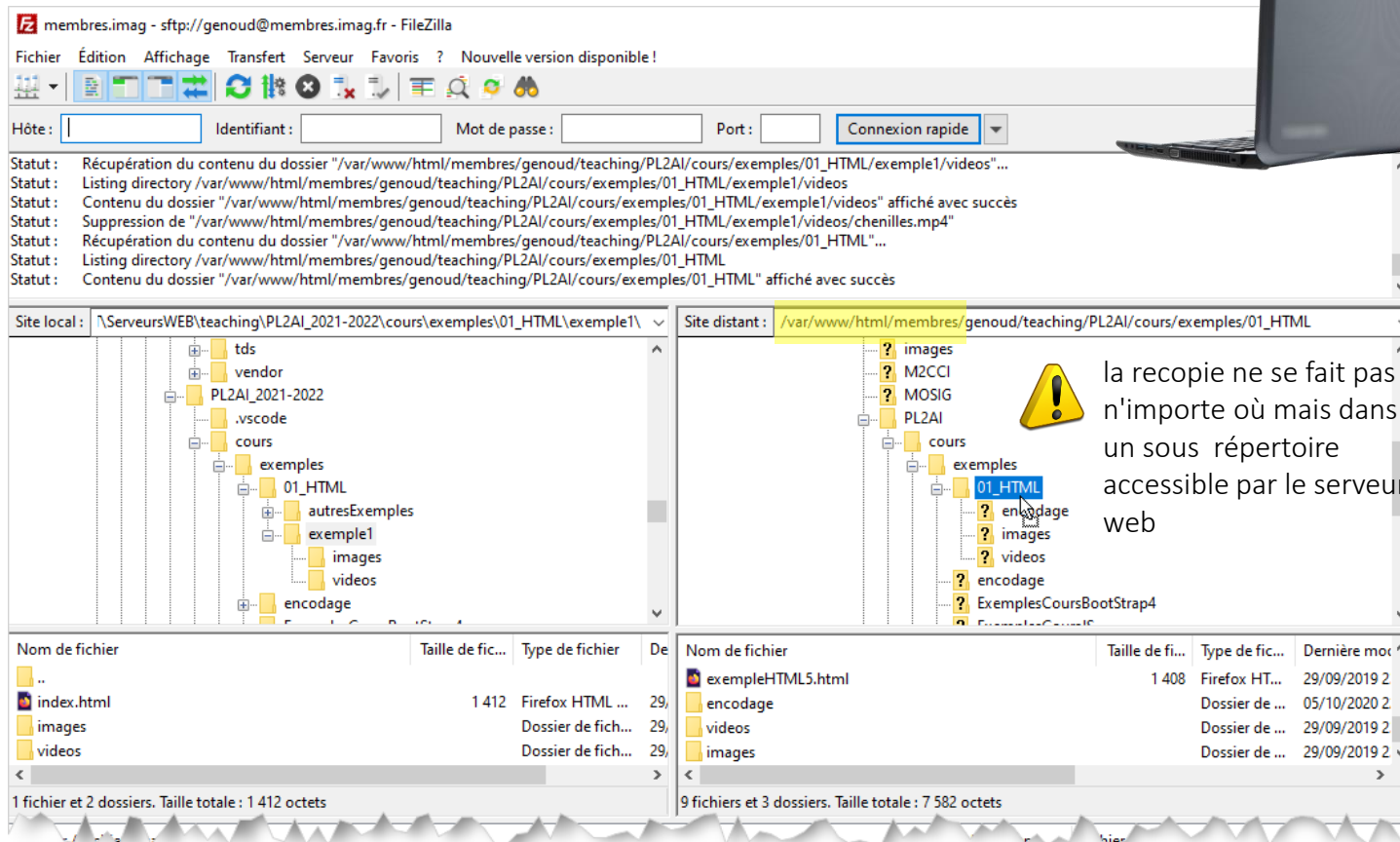
Le fichier index.html et les éventuelles ressources qu'il utilise (images, vidéo, feuilles de styles css, code javascript) sont chargés et affichés dans la fenêtre du navigateur



Consulter une page web

- consultation à distance
 - la page est quelque part sur le web

machine locale où le site a été créé



recopie de la page et des ressources associées (FTP, scp, WebDav ...)

machine *serveur*
ex : lig-membres.imag.fr
+ logiciel "*serveur web*"
ex : Apache, IIS, ...



Consulter une page web

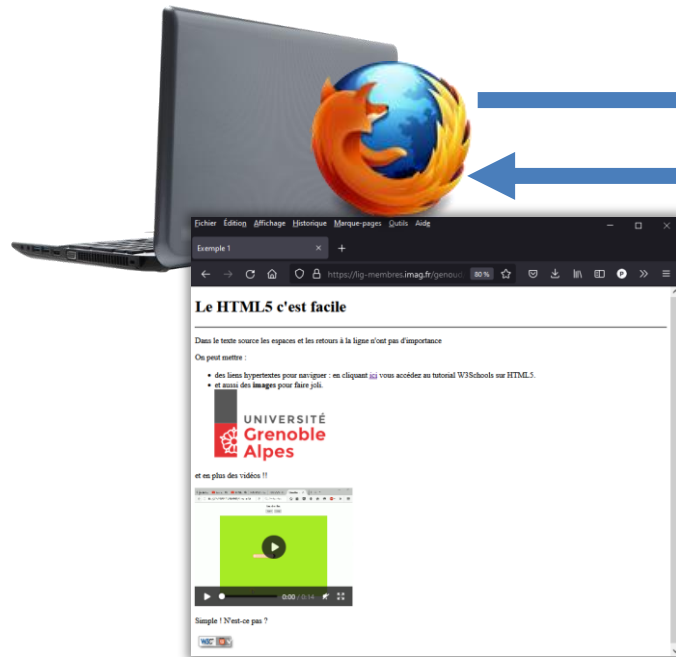
- consultation à distance
 - la page est quelque part sur le web

machine *client*

ex : votre PC

+ logiciel "*client web*"

ex : Firefox, Chrome, Edge ...



<http://lig-membres...>
(requête)

Protocole HTTP

Page web
(réponse)

Réseau Internet

machine *serveur*

ex : lig-membres.imag.fr

+ logiciel "*serveur web*"

ex : Apache, IIS, ...



* illustration inspirée cours M1105 (Web et documents numériques) DUT Informatique UGA – Sylvie Pesty

https://lig-membres.imag.fr/genoud/teaching/PL2AI/cours/exemples/01_HTML/exemple1/

Consulter une page web

- consultation à distance
 - la page est quelque part sur le web et est accédée via son *URL* (Uniform Resource Locator)

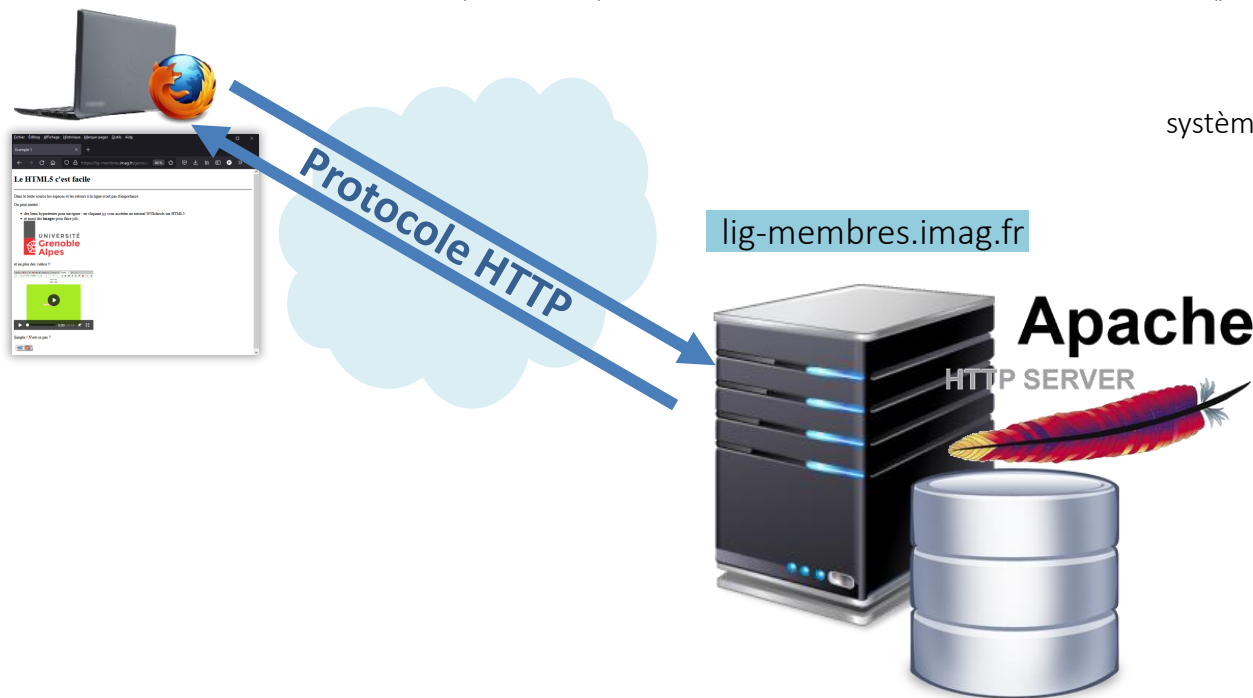
`https://lig-membres.imag.fr/genoud/teaching/PL2AI/cours/exemples/01_HTML/exemple1/index.html`

machine hôte
(*hostname*)

chemin d'accès à la ressource sur le logiciel serveur web (Apache)
(*path*)

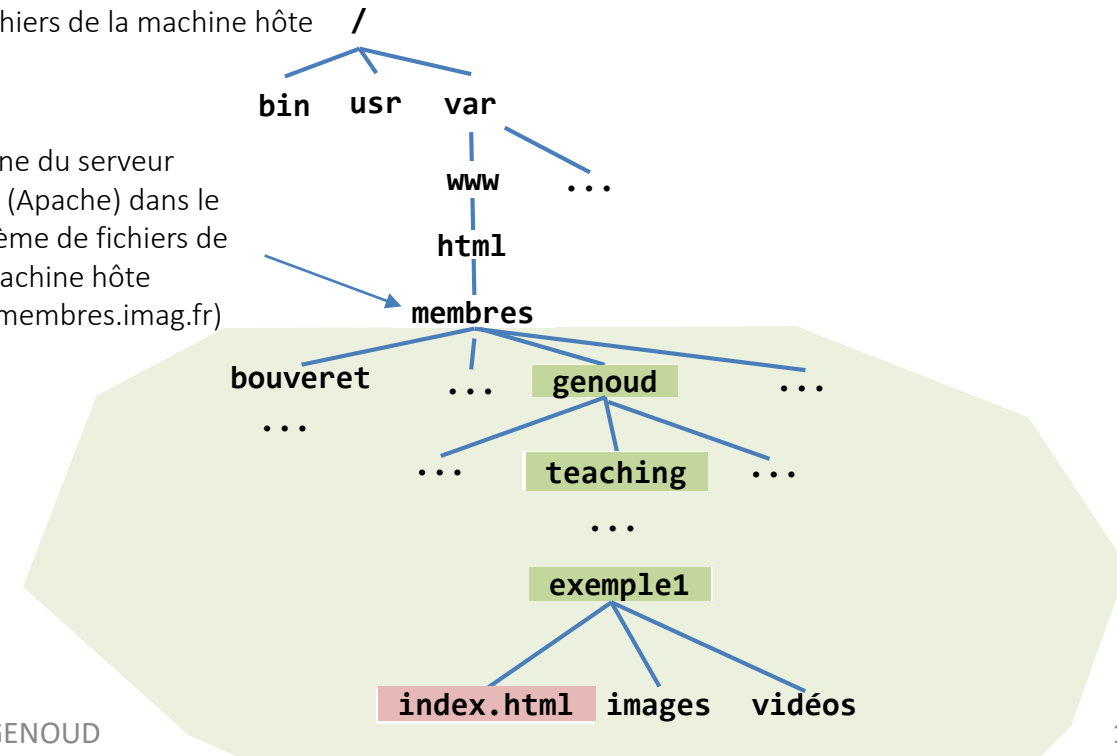
nom de la ressource
(*filename*)

si le fichier s'appelle `index.html`
son nom peut être omis



système de fichiers de la machine hôte /

Racine du serveur
web (Apache) dans le
système de fichiers de
la machine hôte
(lig-membres.imag.fr)



(URL : un cas particulier d'URIs (Uniform Resource Identifier))

- URI : generic syntax

scheme ":" ["//" authority "/"] [path] ["?" query] ["#" fragment]

- **scheme:** http, ftp, mailto, ...
- **authority:** [userinfo@]host[:port]
 - **userinfo:** authentication section e.g: username:password
 - **host:** domain name, IP address
 - **port:** port number, ex: 80 for HTTP standard port
- **path:** a sequence of segments separated by slashes, e.g. : a path in the hierarchical file system of the HTTP server.
- **query:** a query string of non-hierarchical data. (e.g: a sequence of attribute–value pairs separated by a delimiter (&) for HTTP requests)
- **fragment:** a fragment identifier providing direction to a secondary resource (e.g.: anchor id in a HTML document)

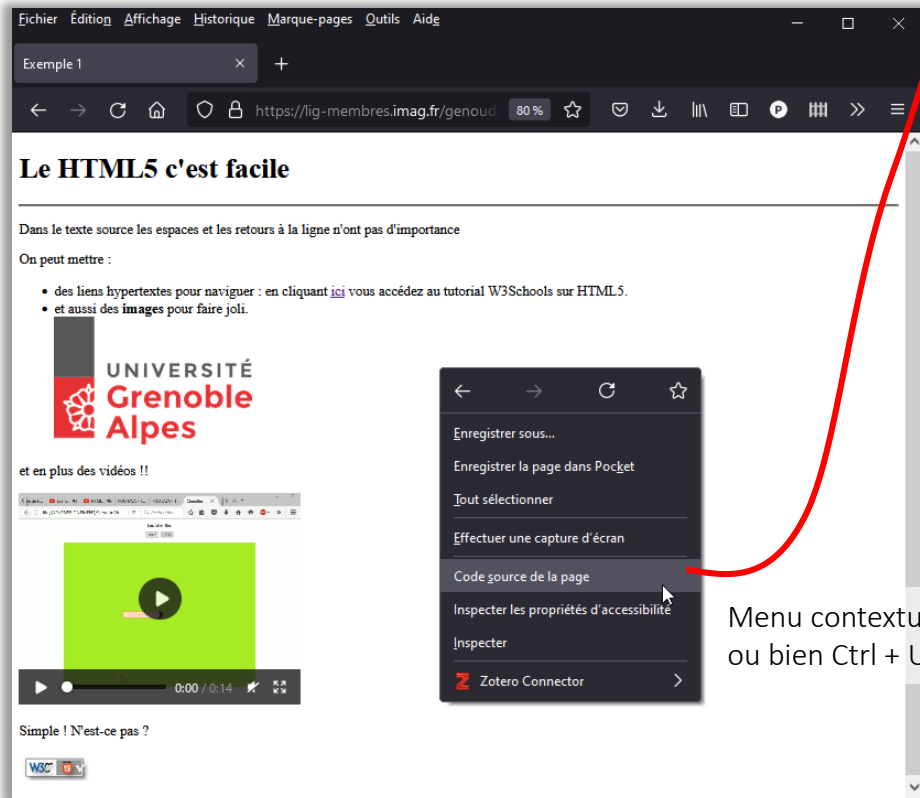


Internationalized Resource Identifier (RFC 3987): extension to support Universal Character Set (Unicode/ISO 10646)

http://fa.dbpedia.org/resource/ژرژ_برسنس

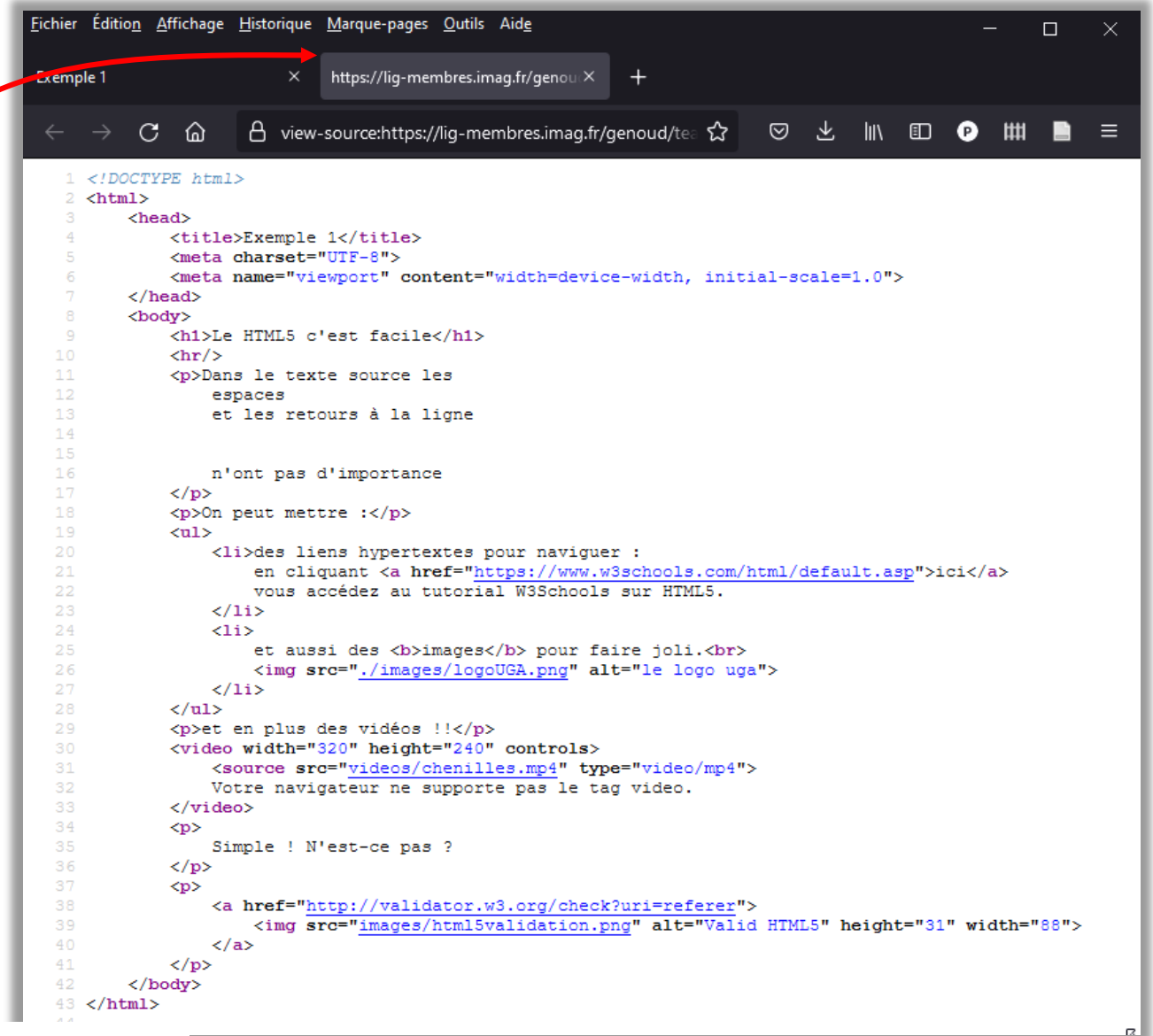
Consulter une page web

- Possibilité consulter le code HTML d'une page chargée dans le navigateur



Menu contextuel (clic droit)
ou bien Ctrl + U

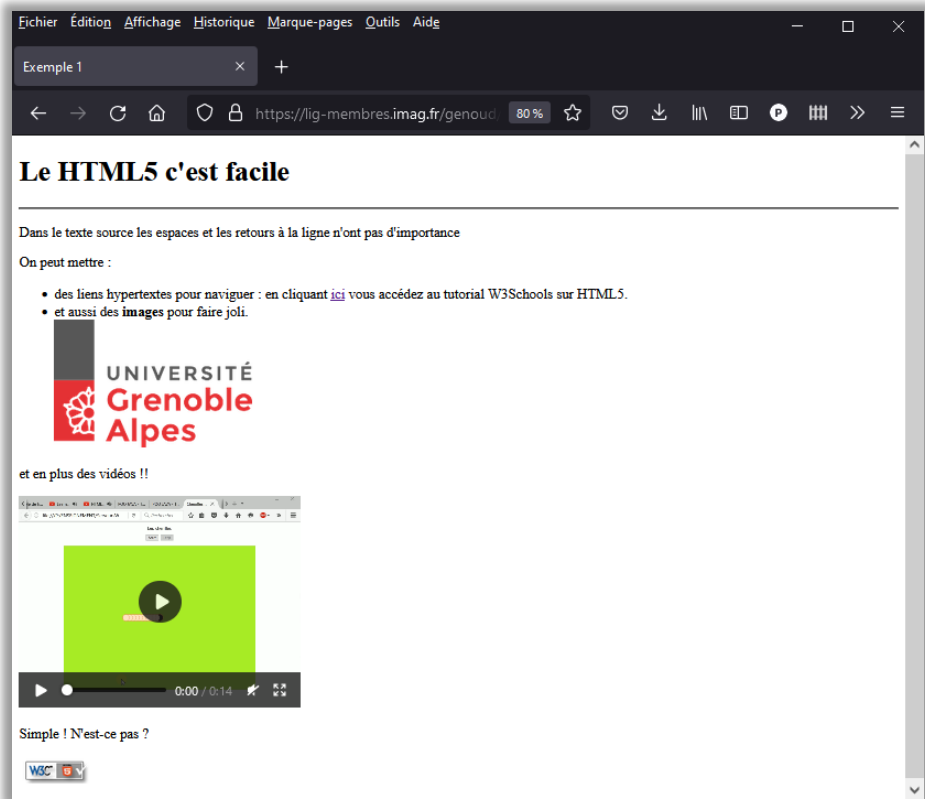
ouverture d'un onglet avec
le code HTML de la page



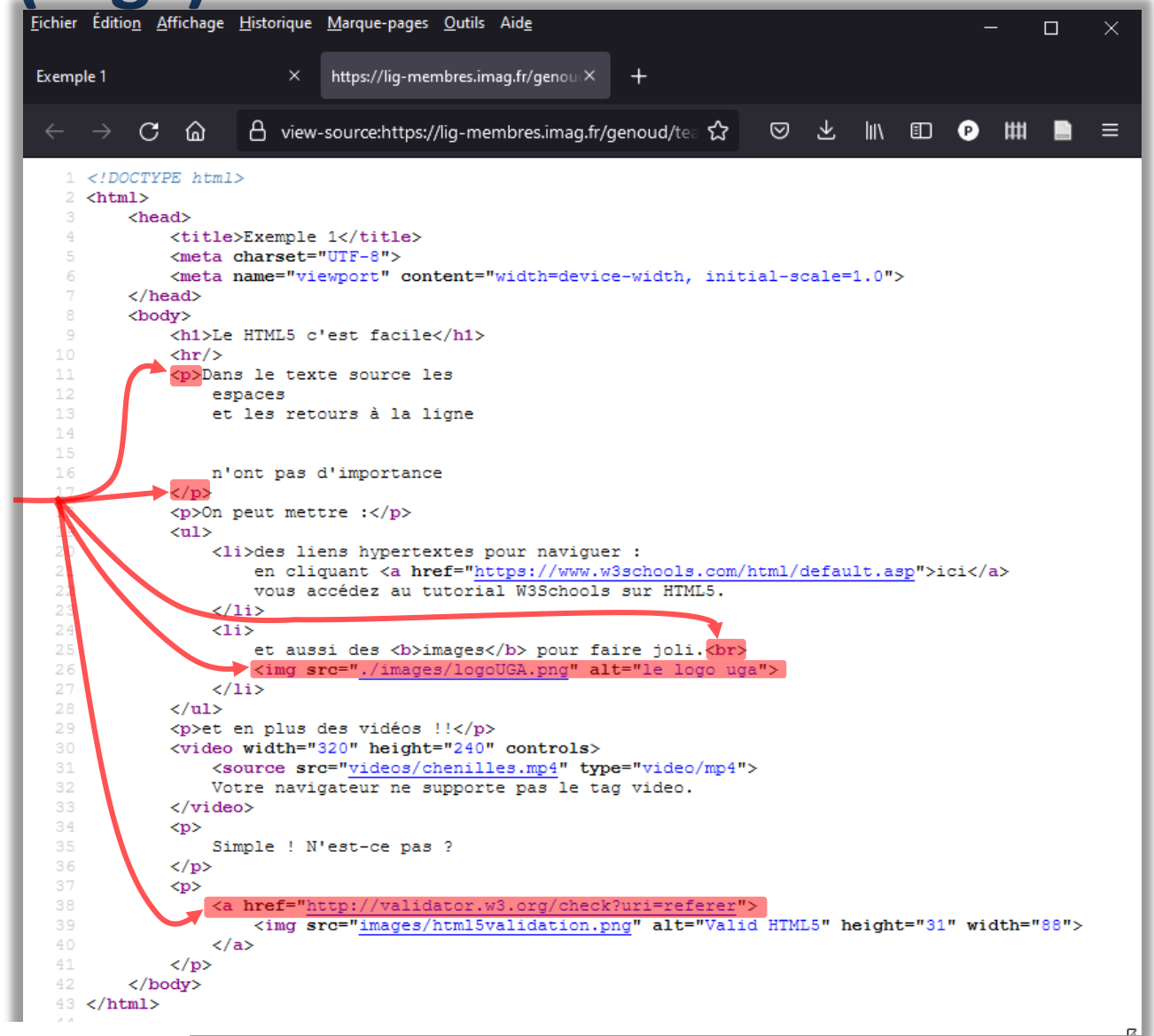
https://lig-membres.imag.fr/genoud/teaching/PL2AI/cours/exemples/01_HTML/exemple1/index.html

HTML : un langage à balises (tags)

- Les balises HTML servent principalement à :
 - structurer le texte : divisions, titres, paragraphes, tableaux, listes...
 - inclure des objets externes : images, sons, vidéos
 - réaliser des liens hypertextes



balises
(tags)



https://lig-membres.imag.fr/genoud/teaching/PL2AI/cours/exemples/01_HTML/exemple1/index.html

Syntaxe des balises

```
<nombalise>contenu</nombalise>
```

```
<h1>Ceci est un titre</h1>
```

```
<p>Ceci est un paragraphe</p>
```

```
<hr/> (balise auto-fermante)
```

```
<hr> (valide en HTML5)
```

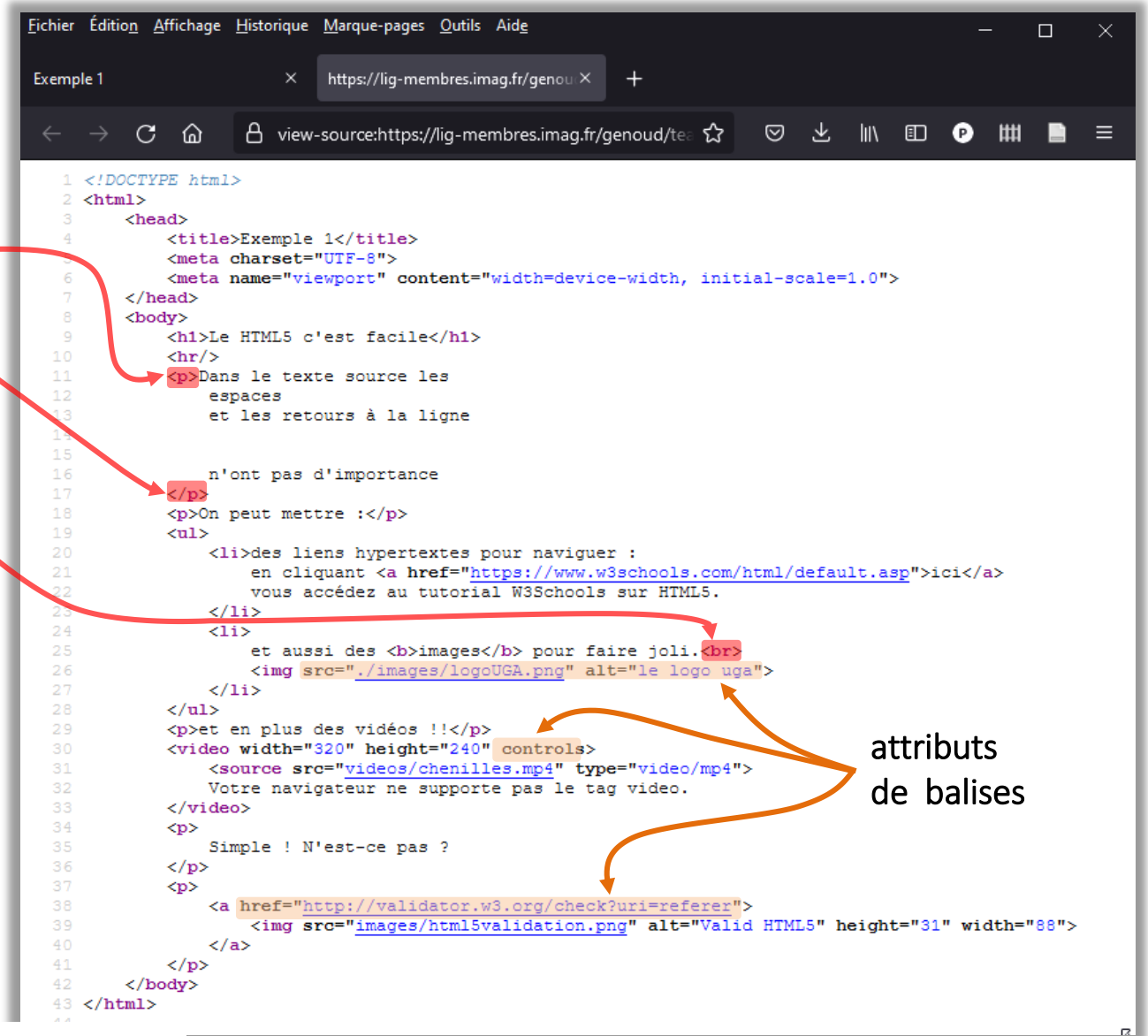
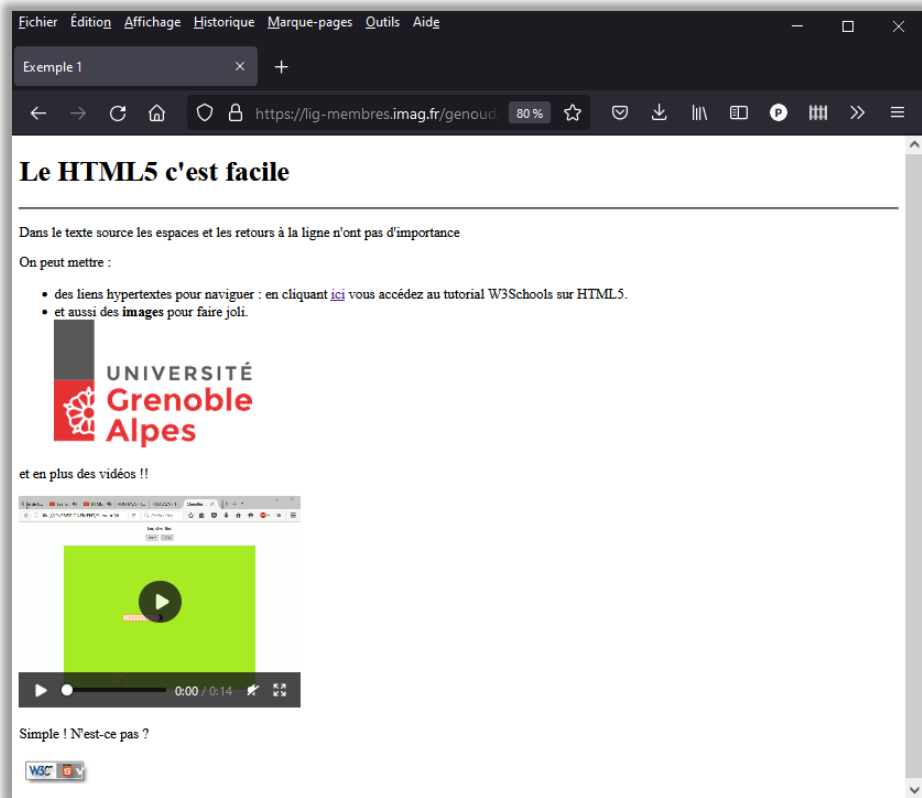
```

```

- Balises (tag)
 - noms entourés de crochets (angle brackets) **<...>**
 - qualifient des portions de texte
- Fonctionnent (en général) par paires
 - **<nombalise>** : balise ouvrante (start tag)
 - **</nombalise>** balise fermante (end tag)
- Certaines balises sont "auto-fermantes"
 - ne délimitent qu'un point et non zone du document
 - **<nombalise/>** (/ facultatif en HTML5)
- balises ouvrantes et balises auto-fermantes peuvent être personnalisées à l'aide d'attributs
 - **<nombalise attr1="valeur1" attr2="valeur2" ...>**

Syntaxe des balises

- exemples de différents types de balises et d'attributs



balise ouvrante

balise fermante

balise vide (auto-fermante)

attributs de balises

https://lig-membres.imag.fr/genoud/teaching/PL2AI/cours/exemples/01_HTML/exemple1/index.html

Les différentes balises de HTML₅

- Plus d'une centaines de balises

- le but n'est pas de toutes les apprendre avec leur différents attributs !

```
<a> <abbr> <address> <area> <article> <aside> <audio> <b> <base> <bdi>  
  <bdo> <blockquote> <body> <br/> <button> <canvas> <caption> <cite>  
<code> <col> <colgroup> <command> <datalist> <dd> <del> <details> <dfn>  
<div> <dl> <dt> <em> <embed> <fieldset> <figcaption> <figure> <footer>  
<form> <h1> to <h6> <head> <header> <hgroup> <hr/> <html> <i> <iframe>  
<img/> <input> <ins> <kbd> <keygen> <label> <legend> <li> <link/> <map>  
<mark> <menu> <meta/> <meter> <nav> <noscript> <object> <ol> <optgroup>  
<option> <output> <p> <param> <pre> <progress> <q> <rp> <rt> <ruby> <s>  
  <samp> <script> <section> <select> <small> <source> <span> <strong>  
<style> <sub> <summary> <sup> <table> <tbody> <td> <textarea> <tfoot>  
  <th> <thead> <time> <title> <tr> <track> <ul> <var> <video> <wbr>
```

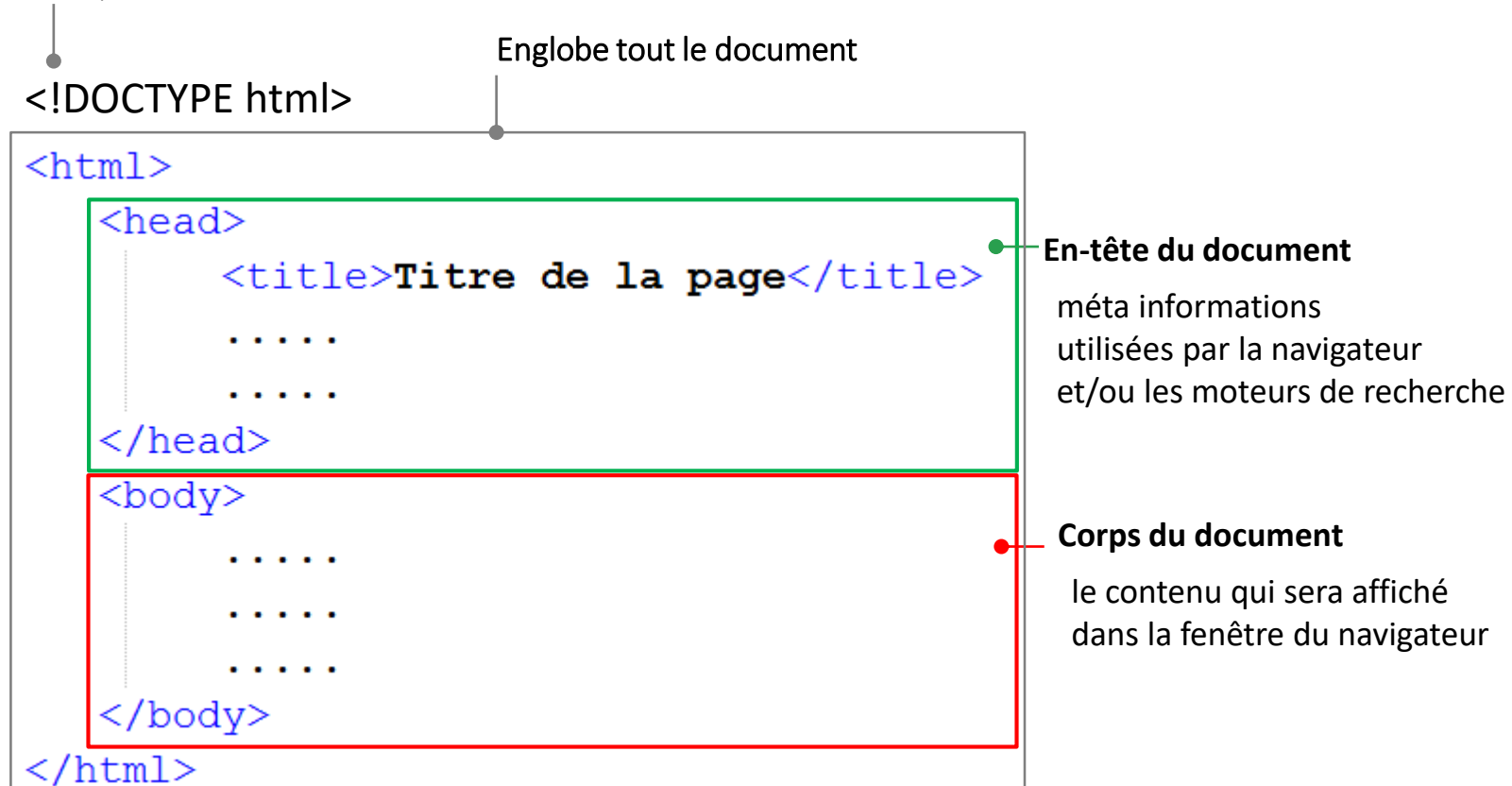
- Plus d'une centaines de balises

- se référer à des sources de documentation fiables

- tutoriaux de w3Schools : <http://www.w3schools.com/>
- Mozilla Developers Network (MDN) : <https://developer.mozilla.org/fr/docs/Web/HTML>

Structure d'un document HTML

La toute première chose dans le document. Ce n'est pas une balise HTML, c'est une instruction pour indiquer au navigateur en quelle version de HTML la page est écrite (ici HTML5).

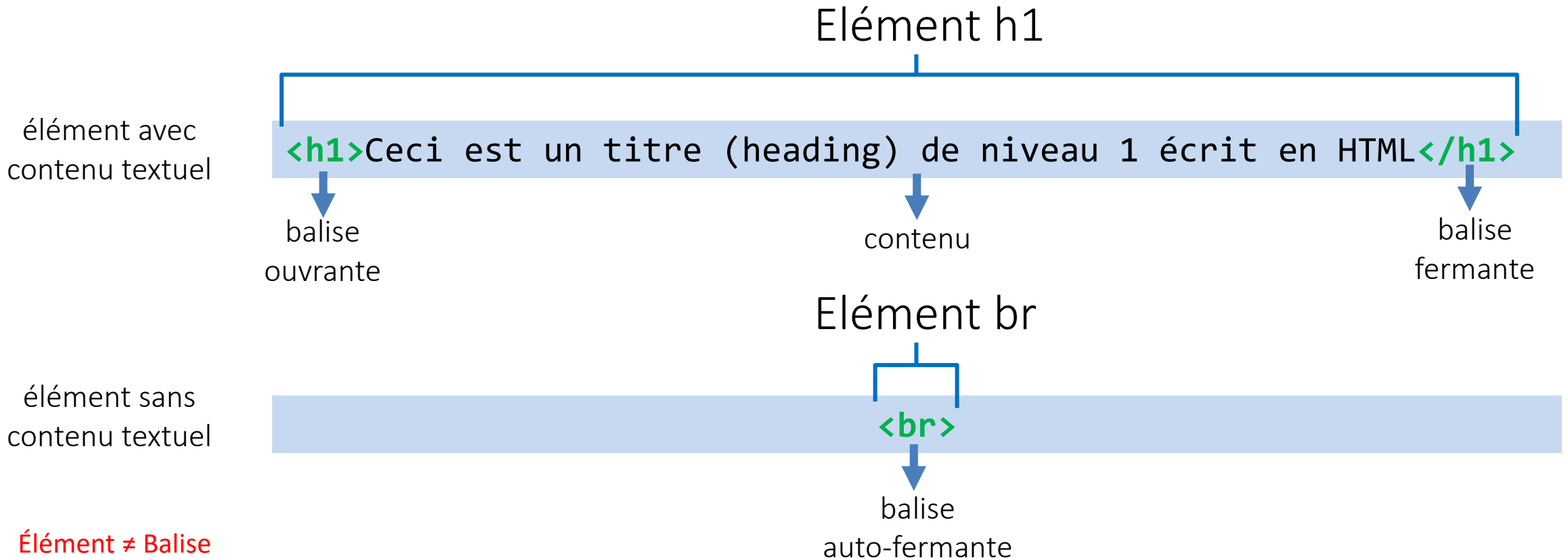


Dans de nombreux éditeurs de code, cette structure peut être générée automatiquement.

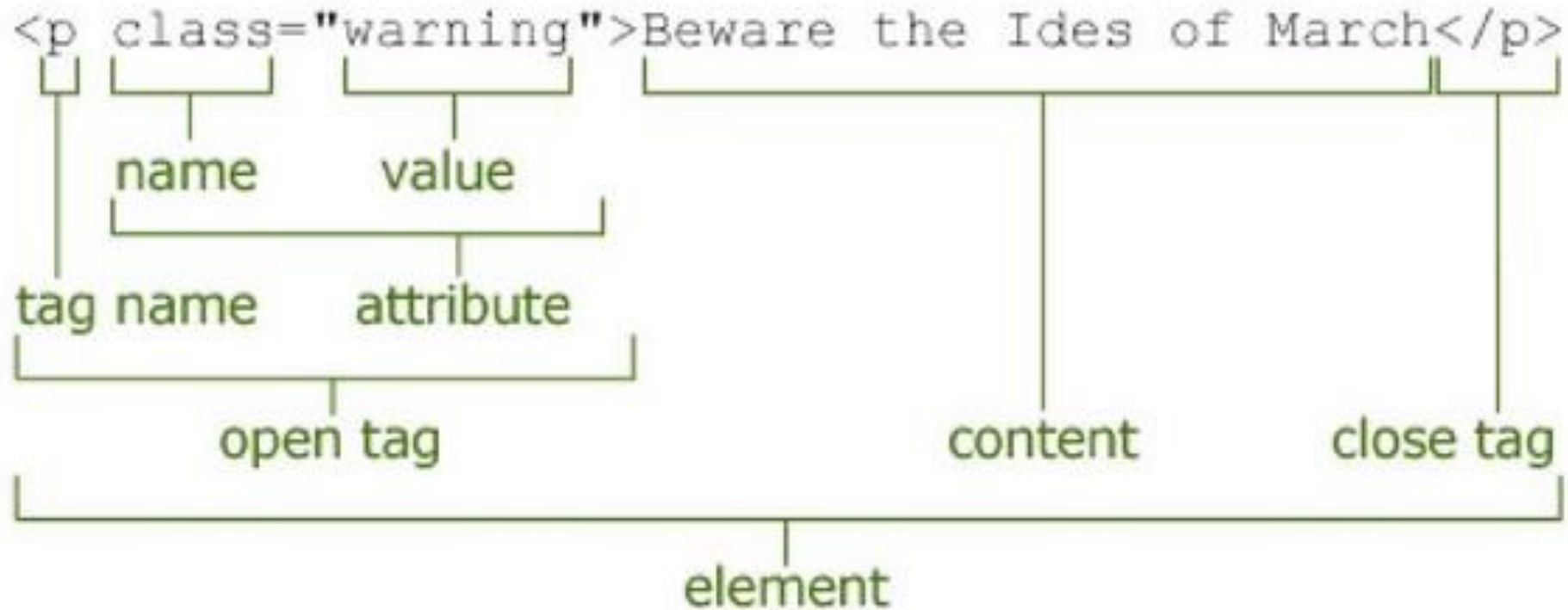
Démo avec VSCode

Éléments HTML

- Les balises permettent de structurer le contenu d'un document HTML en ce que l'on appelle des **éléments HTML**
- Un élément HTML peut être soit constitué d'une paire de balises (ouvrante et fermante) et d'un contenu, soit d'une balise unique (balise auto-fermante).



Element HTML



Éléments HTML



- Les éléments HTML peuvent être imbriqués
 - le contenu d'un élément HTML peut contenir un ou plusieurs autres éléments HTML

```
7 </head>
8 <body>
9 <h1>Le HTML5 c'est facile</h1>
10 <hr/>
11 <p>Dans le texte source les
12     espaces
13     et les retours à la ligne
14
15     n'ont pas d'importance
16 </p>
17 <p>On peut mettre :</p>
18 <ul>
19     <li>des liens hypertextes pour naviguer :
20         en cliquant <a href="https://www.w3schools.com/html/default.asp">ici</a>
21         vous accédez au tutorial W3Schools sur HTML5.
22     </li>
23     <li>
24         et aussi des <b>images</b> pour faire joli.<br>
25         
26     </li>
27 </ul>
28
29 <p>et en plus des vidéos !!</p>
30 <video width="320" height="240" controls>
31 <source src="video/chenilles.mp4" type="video/mp4">
```

Élément **h1**

Élément **ul**

Élément **a** imbriqué dans
l'élément **li** imbriqué dans
l'élément **ul**

Élément **li** imbriqué
dans l'élément **ul**

Éléments HTML

- Règles d'imbrication

- La balise fermante d'un élément imbriqué doit nécessairement apparaître avant la balise fermante de l'élément englobant



```
<p> toto <b> tata </b>titi </p>
```

```
<p>  
  toto  
  <b> tata </b>  
  titi  
</p>
```



```
<p> toto <b> tata </p>titi </b>
```

```
<p>  
  toto  
  <b> tata  
</p>  
  titi  
  </b>
```

- selon la nature de l'élément englobant, seul certains types de balises sont autorisées pour les éléments imbriqués



```
<ul>  
  <li>  
    <p> tata </p>  
  </li>  
</ul>
```

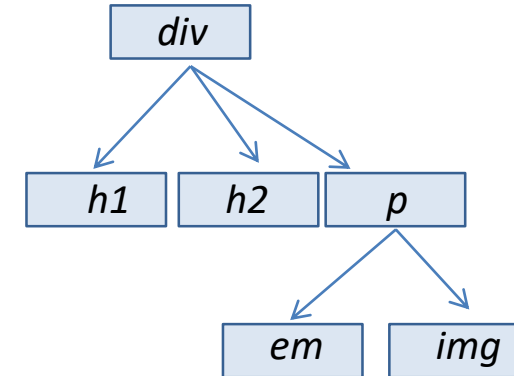
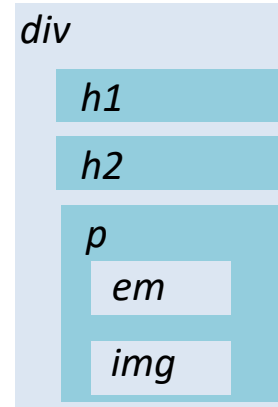


```
<ul>  
  <p>  
    <li> tata </li>  
  </p>  
</ul>
```

Éléments HTML

- l'imbrication des éléments HTML forme une structure hiérarchique (arbre)

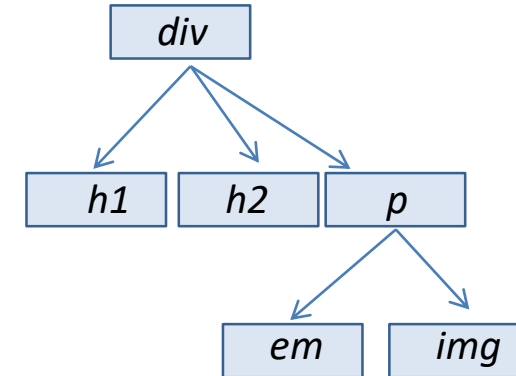
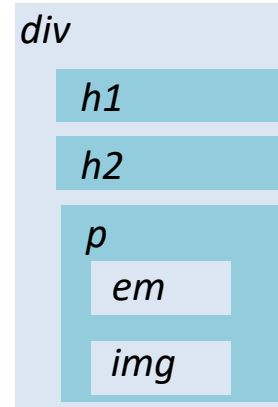
```
<div>
  <h1>titre</h1>
  <h2>un sous titre</h2>
  <p>un paragraphe avec
    <em>des mots importants</em>
    pour terminer avec une image
    .
  </p>
</div>
```



Éléments HTML

- l'imbrication des éléments HTML forme une structure hiérarchique (arbre)

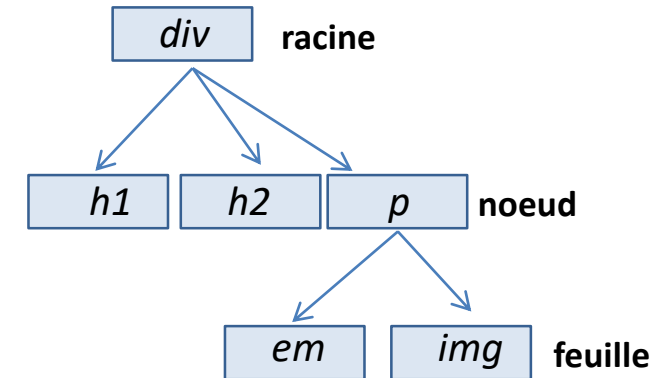
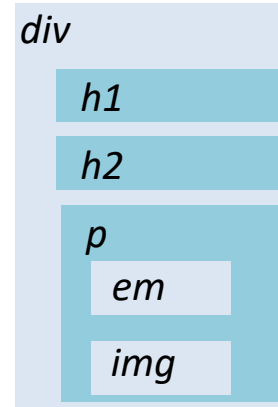
```
<div>
  <h1>titre</h1>
  <h2>un sous titre</h2>
  <p>un paragraphe avec
    <em>des mots importants</em>
    pour terminer avec une image
    .
  </p>
</div>
```



Éléments HTML

- l'imbrication des éléments HTML forme une structure hiérarchique (arbre)

```
<div>
  <h1>titre</h1>
  <h2>un sous titre</h2>
  <p>un paragraphe avec
    <em>des mots importants</em>
    pour terminer avec une image
    .
  </p>
</div>
```

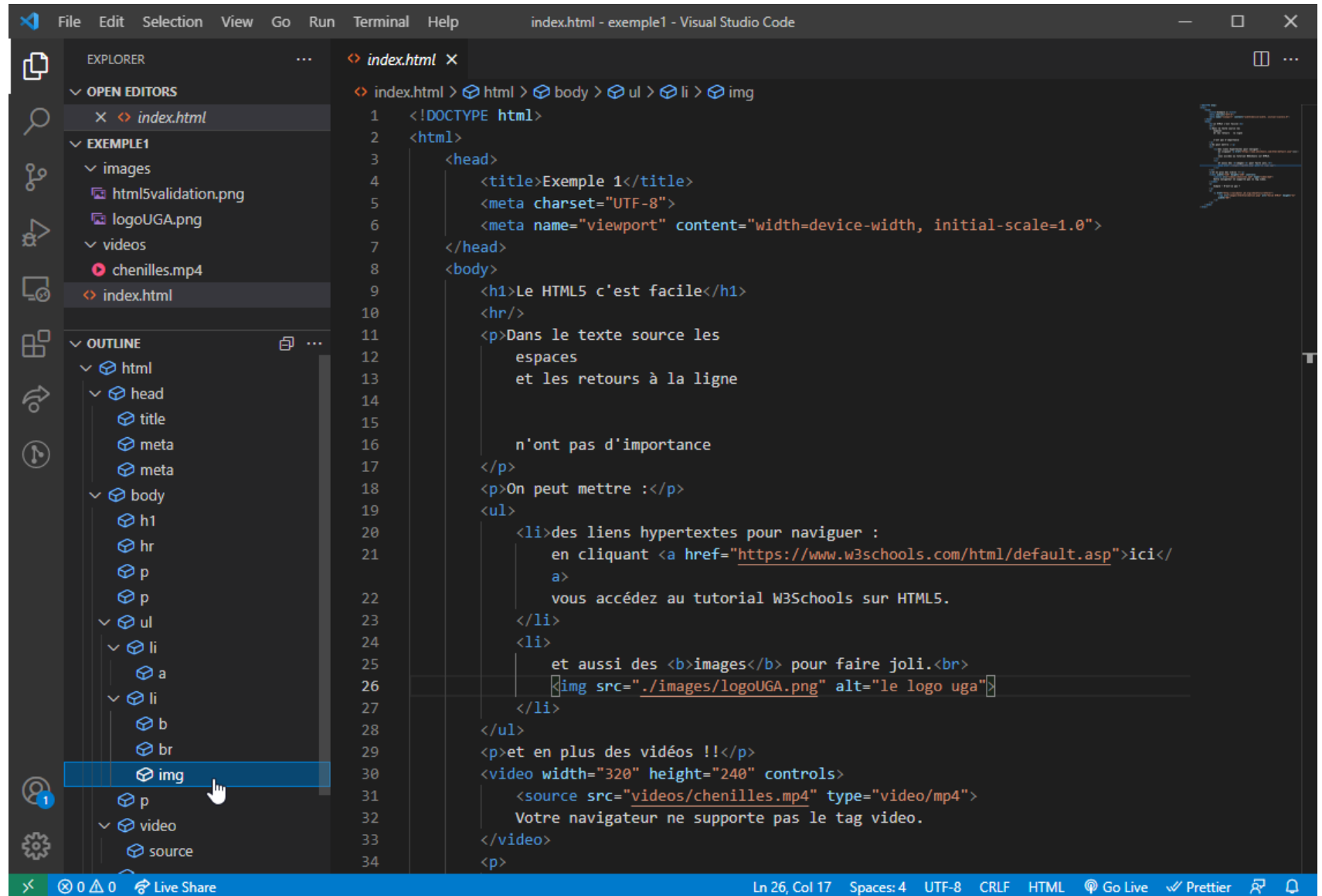


p est *père* de **em** et de **img**
p est *parent* de **em** et **img**
em et **img** sont les *enfants* de **p**
em est un *fils* de **p**
img est *frère* de **em**
div est un *ancêtre* de **em**
em est un *descendant* de **div**



Éléments HTML

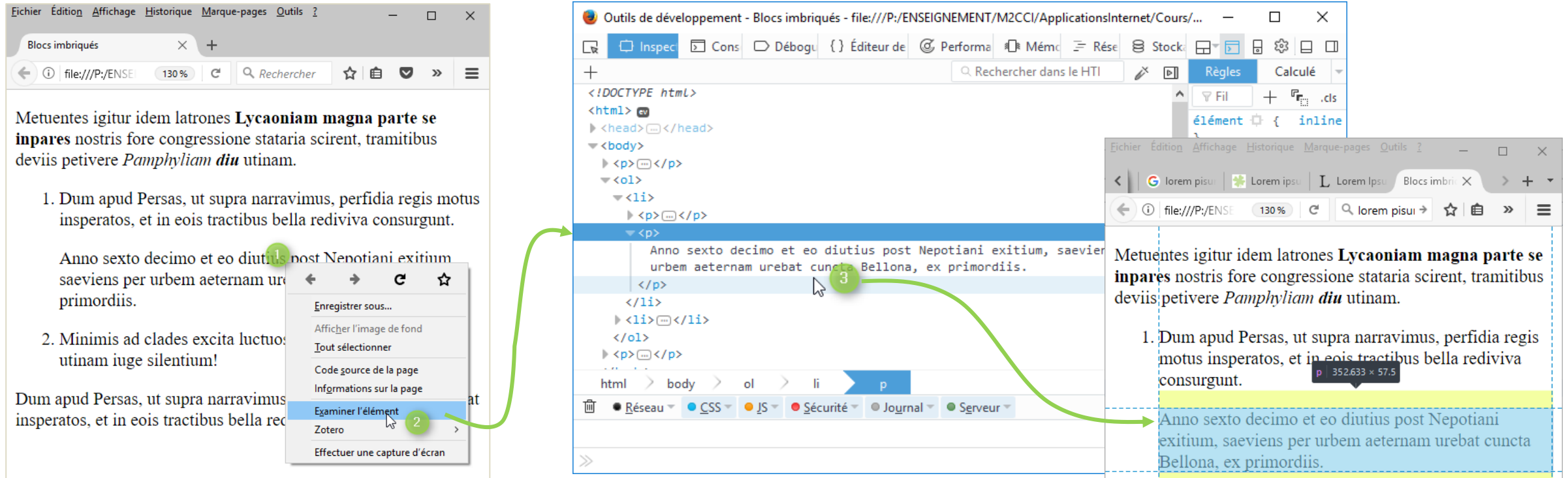
- Très souvent les éditeurs de code permettent de naviguer rapidement dans cette structure hiérarchique
 - exemple : Visual Studio Code



```
index.html - exemple1 - Visual Studio Code
EXPLORER
  OPEN EDITORS
    index.html
  EXEMPLE1
    images
      html5validation.png
      logoUGA.png
    videos
      chenilles.mp4
    index.html
  OUTLINE
    html
      head
        title
        meta
        meta
      body
        h1
        hr
        p
        p
        ul
          li
            a
          li
            b
            br
            img
        p
        video
          source
index.html
1 <!DOCTYPE html>
2 <html>
3   <head>
4     <title>Exemple 1</title>
5     <meta charset="UTF-8">
6     <meta name="viewport" content="width=device-width, initial-scale=1.0">
7   </head>
8   <body>
9     <h1>Le HTML5 c'est facile</h1>
10    <hr/>
11    <p>Dans le texte source les
12      espaces
13      et les retours à la ligne
14
15      n'ont pas d'importance
16    </p>
17    <p>On peut mettre :</p>
18    <ul>
19      <li>des liens hypertextes pour naviguer :
20        en cliquant <a href="https://www.w3schools.com/html/default.asp">ici</a>
21        vous accédez au tutorial W3Schools sur HTML5.
22      </li>
23      <li>
24        et aussi des <b>images</b> pour faire joli.<br>
25        
26      </li>
27    </ul>
28    <p>et en plus des vidéos !!</p>
29    <video width="320" height="240" controls>
30      <source src="videos/chenilles.mp4" type="video/mp4">
31      Votre navigateur ne supporte pas le tag video.
32    </video>
33    <p>
34
```

Éléments HTML

- de même les outils de développement intégrés aux navigateurs permettent d'*inspecter* les éléments



Cliquez avec le bouton droit* de la souris sur un élément (ou une zone vide) et choisissez "Inspecter" ou "Examiner l'élément" pour voir la structure du code HTML et matérialiser l'élément sélectionné dans la fenêtre du navigateur .

* ou bien CTRL + clic bouton droit

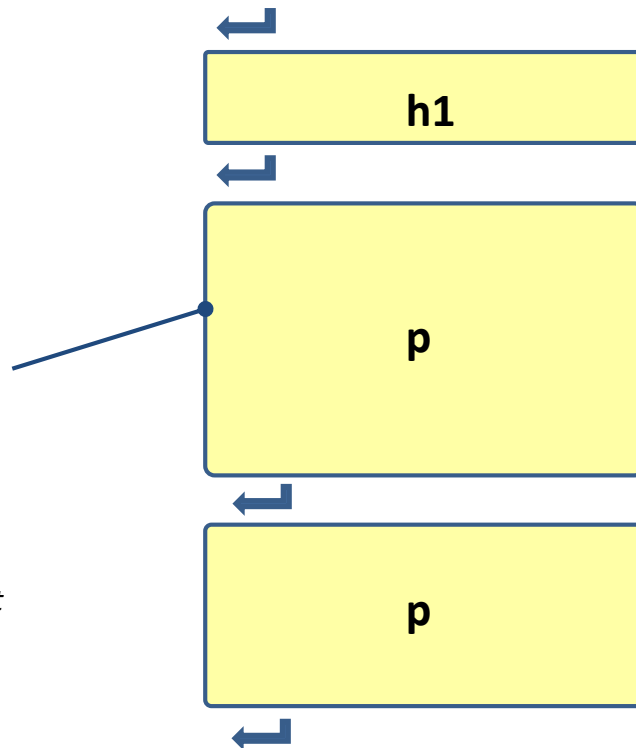
Éléments de bloc et éléments en ligne

- Deux types éléments
 - **bloc** (*block*) et **en ligne** (*inline*)
 - Dicte leur comportement (affichage, positionnement, ...)

Élément de bloc : par défaut affichage vertical

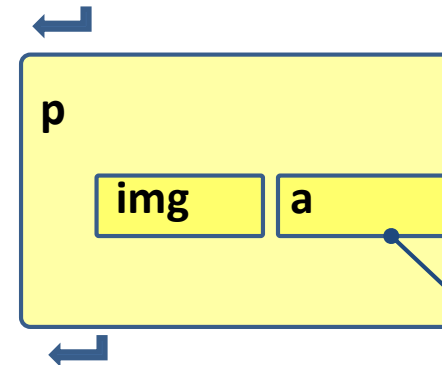
commence toujours sur une nouvelle ligne et prend la largeur totale disponible (s'étend sur la gauche et la droite aussi loin que possible).*

** Plus précisément un élément bloc prend la largeur de son conteneur*



Élément en ligne : par défaut affichage au fil du texte

ne démarre pas sur une nouvelle ligne et ne prend que toute la largeur nécessaire.



Éléments de bloc et éléments en ligne

- exemples

The collage illustrates the difference between block and inline elements in HTML. It features a code editor on the left showing HTML code for three paragraphs. The first paragraph is a simple block element. The second paragraph contains a **Lycaoniam magna parte campestrum cum se inpare** inline element. The third paragraph is another simple block element. On the right, a browser window shows the rendered output. The first paragraph is a full-width block. The second paragraph shows the inline element as a highlighted green box that does not start on a new line and does not take up the full width. The third paragraph is another full-width block. Below the browser window, a callout box explains that block elements start on a new line and take up the full width. Another callout box explains that inline elements do not start on a new line and only take up the width they need.

```
<body>  
<p>  
Post quorum necem nihilo lenius ferociens Gallus ut leo cadaveribus  
pastus multa huius modi scrutabatur. quae singula narrare non refert,  
me professione modum, quod evitandum est, excedamus.  
</p>  
<p>  
Metuentes igitur idem latrones <strong>Lycaoniam magna parte campestrum cum  
se inpare</strong> nostris fore congressione stataria documentis frequentibus  
scirent, tramitibus deviis petivere <em>Pamphyliam diu</em> quidem intactam  
sed timore populationum et caedium, milite per omnia diffuso  
propinqua, magnis undique praesidiis communitam.  
</p>  
<p>  
Dum apud Persas, ut supra narravimus, perfidia regis motus agitat  
insperatos, et in eo tractibus bella rediviva consurgunt,  
anno sexto decimo et eo diutius post Nepotiani exitium,  
saeviens per urbem aeternam urebat cuncta Bellona, ex primordiis  
minimis ad clades ne forte paria quandoque temptentur, plus exemplis generalibus nocitura  
quam delictis.  
</p>  
</body>
```

Fichier Édition Affichage Historique Marque-pages Outils ?_

Block

file:///P:/ENSEIGN... 110% Rechercheur

Post quorum necem nihilo lenius ferociens Gallus ut leo cadaveribus pastus multa huius modi scrutabatur. quae singula narrare non refert, me professione modum, quod evitandum est, excedamus.

Metuentes igitur idem latrones Lycaoniam magna parte campestrum cum se inpare nostris fore congressione stataria documentis frequentibus scirent, tramitibus deviis petivere Pamphyliam diu quidem intactam sed timore populationum et caedium, milite per omnia diffuso propinqua, magnis undique praesidiis communitam.

Dum apud Persas, ut supra narravimus, perfidia regis motus agitat insperatos, et in eo tractibus bella rediviva consurgunt, anno sexto decimo et eo diutius post Nepotiani exitium, saeviens per urbem aeternam urebat cuncta Bellona, ex primordiis minimis ad clades ne forte paria quandoque temptentur, plus exemplis generalibus nocitura quam delictis.

Fichier Édition Affichage Historique Marque-pages Outils ?_

Block

Europe-Belarus - Biélorussie | Liste des pays par IDH — W... | 110% Rechercheur

Post quorum necem nihilo lenius ferociens Gallus ut leo cadaveribus pastus multa huius modi scrutabatur. quae singula narrare non refert, me professione modum, quod evitandum est, excedamus.

Metuentes igitur idem latrones Lycaoniam magna parte campestrum cum se inpare nostris fore congressione stataria documentis frequentibus scirent, tramitibus deviis petivere Pamphyliam diu quidem intactam sed timore populationum et caedium, milite per omnia diffuso propinqua, magnis undique praesidiis communitam.

Dum apud Persas, ut supra narravimus, perfidia regis motus agitat insperatos, et in eo tractibus bella rediviva consurgunt, anno sexto decimo et eo diutius post Nepotiani exitium, saeviens per urbem aeternam urebat cuncta Bellona, ex primordiis minimis ad clades ne forte paria quandoque temptentur, plus exemplis generalibus nocitura quam delictis.

Bloc: commence toujours sur une nouvelle ligne et prend la largeur totale disponible.

En ligne: ne démarre pas sur une nouvelle ligne et ne prend que la largeur nécessaire.

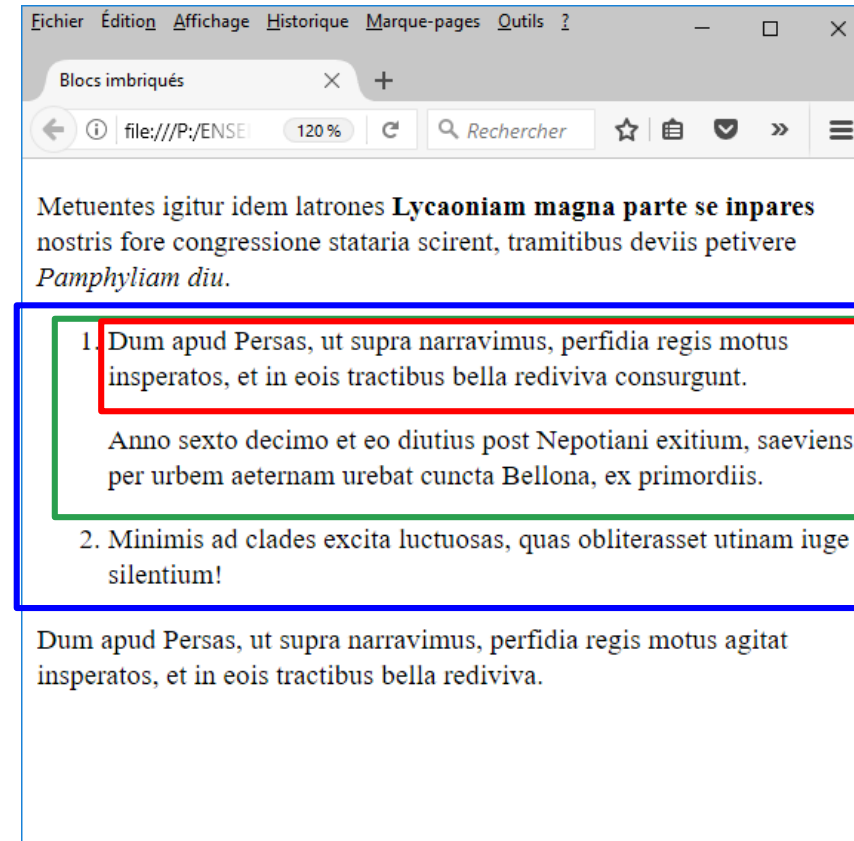
Éléments de bloc et éléments en ligne

- Éléments de bloc
 - Exemples
 - `<p>` paragraphe
 - `<h1>` `<h2>` ... `<h6>` titre (*header level 1, level 2 ...*)
 - `` liste ordonnée (*ordered list*)
 - `` liste non ordonnée (*unordered list*)
 - `` élément d'une liste (*list item*)
 - `<table>` tableau
 - `<tr>` ligne d'un tableau (*table row*)
 - ...
 - `<div>` bloc générique (pas de présentation ni marge) (*division*)
 - Peuvent contenir un (ou plusieurs) éléments de blocs et/ou en ligne.
 - Exception pour `<h1>` ... `<h6>` et `<p>` limités au contenu en ligne.

Éléments de bloc et éléments en ligne

- Exemple d'imbrication d'éléments de niveau bloc

```
<meta name="viewport" content="width=device-width, initial-scale=1.0">
</head>
<body>
<p>
Metuentes igitur idem latrones <strong>Lycaoniam magna parte
se inpares</strong> nostris fore congressione stataria
scirent, tramitibus deviis petivere <em>Pamphylia diu</em>.
</p>
<ol>
<li>
<p>
Dum apud Persas, ut supra narravimus, perfidia regis motus
insperatos, et in eo tractibus bella rediviva consurgunt
</p>
<p>
Anno sexto decimo et eo diutius post Nepotiani exitium,
saeviens per urbem aeternam urebat cuncta Bellona,
ex primordiis.
</p>
<li>
<p>
Minimis ad clades excita luctuosas,
quas obliterasset utinam iuge silentium!
</p>
</li>
</ol>
<p>
Dum apud Persas, ut supra narravimus, perfidia regis motus agit
insperatos, et in eo tractibus bella rediviva.
</p>
```



Éléments de bloc et éléments en ligne

– Éléments en ligne

- Exemples

- `<a>` lien hypertexte
- `` inclus une image
- `` indique que le texte a une importance particulière (par défaut affichage en gras)
- `` (pour emphase) marque un texte sur lequel on veut insister (par défaut affichage en italique)
- ...
- `` conteneur générique servant à regrouper d'autres éléments au fil du texte (équival. `<div>`)

- Ne peuvent contenir que des éléments en ligne (pas d'éléments de bloc).

```
<em>Pamphylia <strong>diu</strong></em> utinam.
```

*Pamphylia **diu** utinam.*

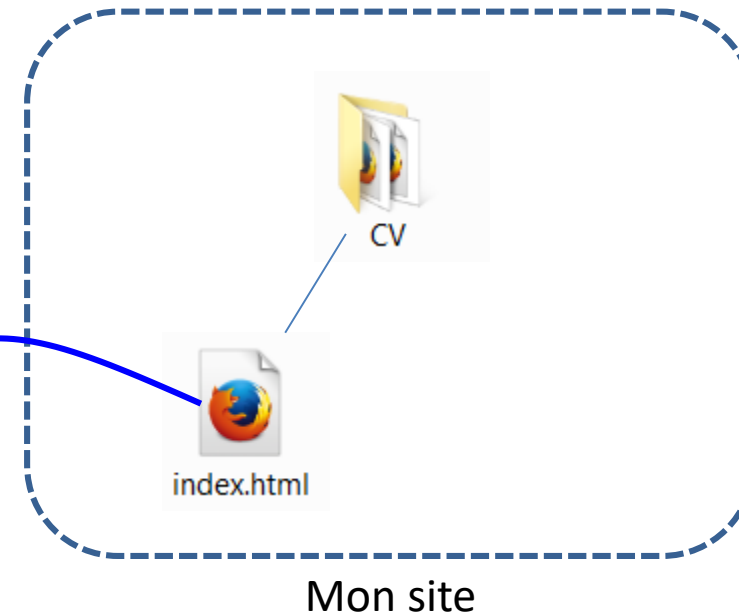
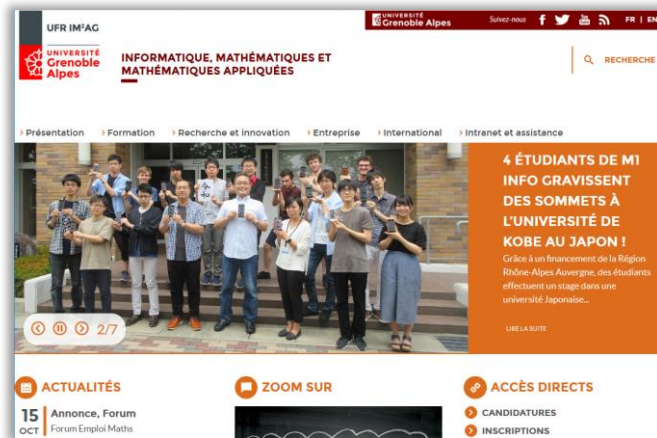
```
<em>Pamphylia <p>diu</p></em> utinam.
```

HTML : Liens hypertextes

- balise `<a>` , lien spécifié avec l'attribut `href`
<https://developer.mozilla.org/fr/docs/Web/HTML/Element/a>
- 3 types de liens
 - Vers document/ressource externe
 - Ex: lien vers une page HTML d'un autre site
 - Vers un document/ressource interne
 - Ex: lien vers une image provenant du même site que celui de la page HTML où se trouve le lien
 - A l'intérieur d'un même document (ancres)
 - Ex: lien vers un portion donnée d'une page HTML

HTML : Liens hypertextes

- Liens vers des ressources externes
 - L'url de la forme `http(s)://nomDeDomaine/...`
 - Ex: `Le site de l'ufr im2ag`



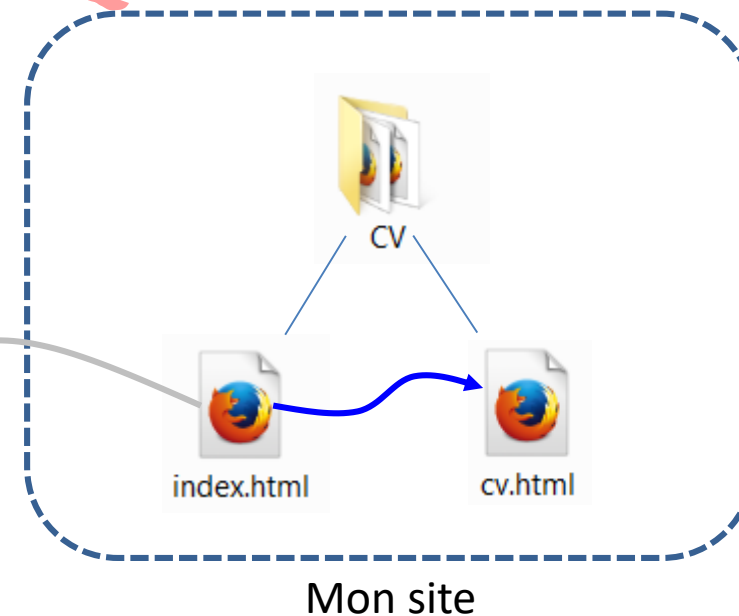
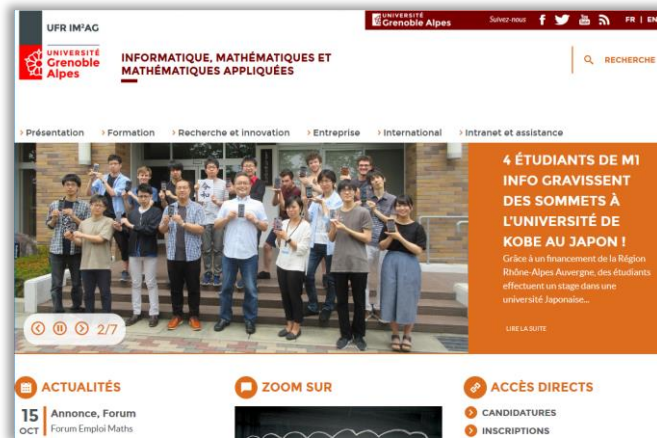
HTML : Liens hypertextes

- Liens vers des ressources internes

- On utilise des chemins relatifs
 - Les chemins absolus sont à proscrire

- Ex: `Mon CV`

~~`Mon CV`~~



HTML : Liens hypertextes

- Liens à l'intérieur d'un document (ancres)

- lien vers un élément précis d'une page,
- élément identifié par un attribut `id` (*identifier*)

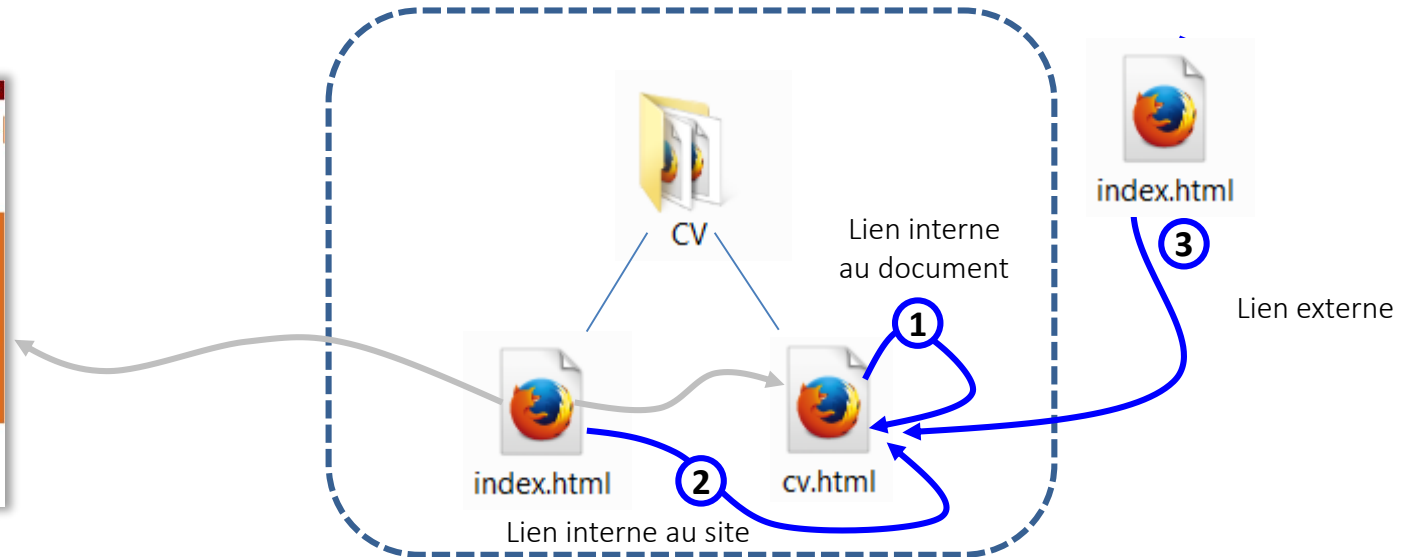
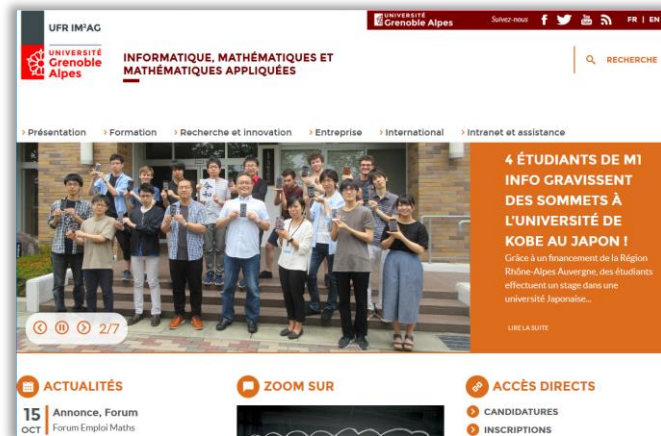
```
<section id="loisirs">mes loisirs</section>
```

- Utilisation d'un fragment (#) dans l'URL

```
<a href="#loisirs">Mes loisirs</a> ①
```

```
<a href="./cv.html#loisirs">Mes loisirs</a> ②
```

```
<a href="http://www.monserveur.org/CV/cv.html#loisirs">Loisirs de X</a> ③
```



HTML : Images

- balise `` , image spécifié avec l'attribut **src**
<https://developer.mozilla.org/fr/docs/Web/HTML/Element/img>
- attributs
 - **src** : URL de l'image (obligatoire)
 - un fichier* situé sur le même machine que la page
``
 - un fichier* situé sur une machine distante
``
 - **alt** : texte alternatif, à afficher si l' image ne peut pas l'être (recommandé)
 - **title** : bulle d'aide
 - **width, height** : taille de l'image en pixels

*l'image n'est pas nécessairement un fichier, l'url peut désigner un programme qui génère l'image à la volée

HTML : Images

- Navigateurs supportent différents formats d'image
- images bitmaps :
 - 3 principaux formats (cours M1105 (Web et documents numériques) DUT Informatique UGA – Sylvie Pesty):
 - JPEG (Joint Photographic Experts Group) : bien adaptée aux photos (images comportant beaucoup de couleurs), compression détériore la qualité (plus ou moins selon taux de compression)
 - PNG (Portable Network Graphics) : plus récent, libre de droit, compression sans détérioration de qualité
 - PNG 8 bits : stocke jusque 256 couleurs, les images peuvent être rendues transparentes
 - PNG 24 bits : 16 M de couleurs, les images peuvent être rendues transparentes sur 256 niveaux (permet un rendu plus lisse)
 - GIF (Graphics Interchange Format): ancien format très utilisé, limité à 256 couleurs, concurrencé par le PNG8 qui compresse mieux, mais peut être animé
- images vectorielles :
 - format SVG (Scalable Vector Graphics)
 - basé sur XML et développé par W3C

(Images Bitmap)

- images définis sous forme de tableau bidimensionnel
- la valeur de chaque case du tableau définit la couleur du pixel

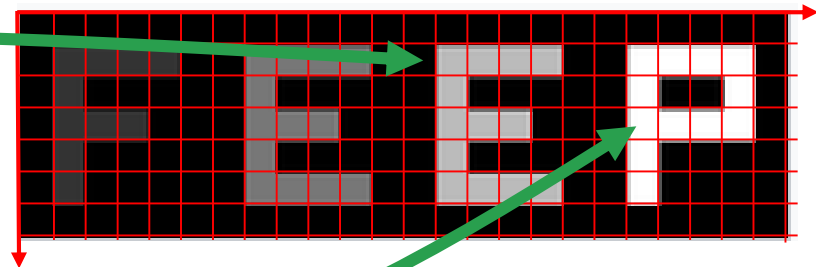
Exemple : Une image en niveaux de gris 8 bits (codage PGM - portable graymap - ASCII) – image non compressée

https://en.wikipedia.org/wiki/Netpbm#PGM_example

fichier image : `exemple.pgm`

```
P2
# Shows the word "FEEP" (example from Netpbm man page on PGM)
24 7
15
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 3 3 3 3 0 0 7 7 7 7 0 11 11 11 11 0 0 15 15 15 15 0
0 3 0 0 0 0 0 7 0 0 0 0 0 11 0 0 0 0 0 15 0 0 15 0
0 3 3 3 0 0 0 7 7 7 0 0 0 11 11 11 0 0 0 15 15 15 0
0 3 0 0 0 0 0 7 0 0 0 0 0 11 0 0 0 0 0 15 0 0 0 0
0 3 0 0 0 0 0 7 7 7 7 0 0 11 11 11 11 0 0 15 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

l'image correspondante
obtenue en l'ouvrant avec un logiciel
de traitement d'images (par ex. GIMP)



chaque caractère définit la couleur (ici niveau de gris, 0 noir, 15 : blanc) du pixel correspondant

HTML : images

- Comparaison image vectorielle / image bitmap

fichier HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta http-equiv="X-UA-Compatible" content="IE=edge">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Document</title>
</head>
<body>
  
  
</body>
</html>
```

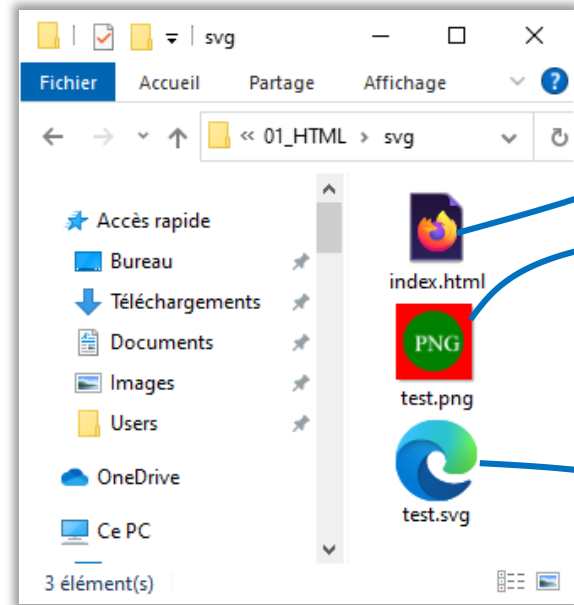
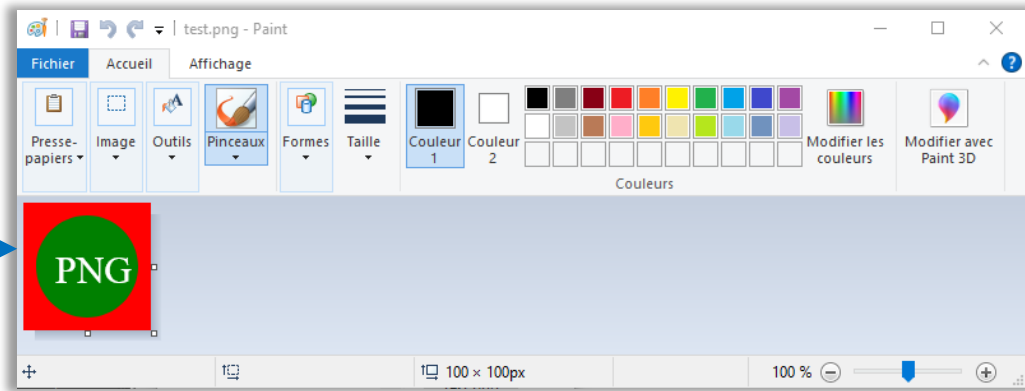


image bitmap 100 x 100 pixels

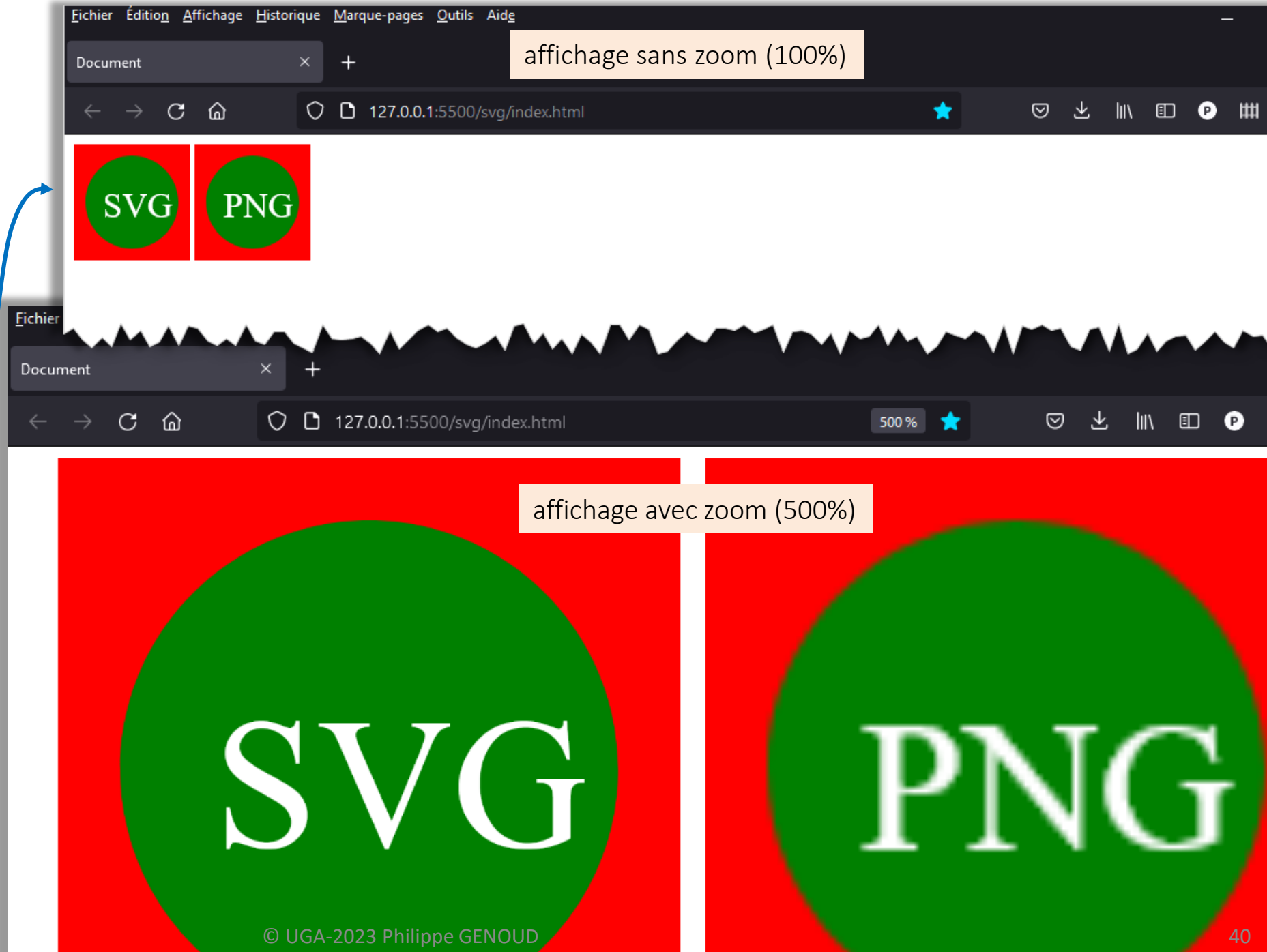
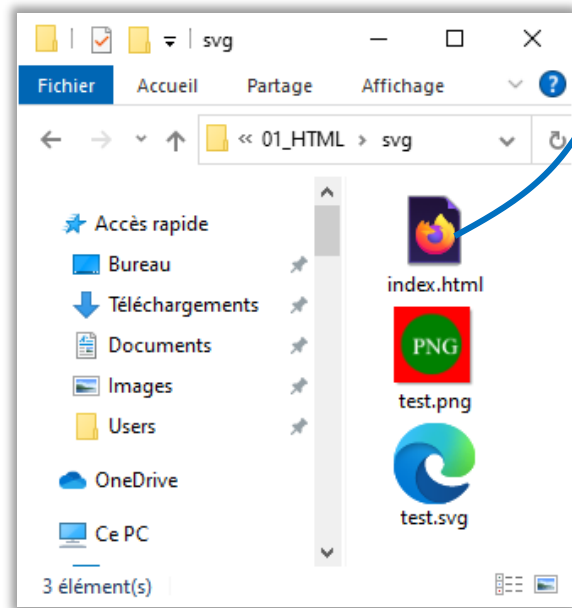


fichier SVG

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<svg version="1.1"
  baseProfile="full"
  width="100" height="100"
  xmlns="http://www.w3.org/2000/svg">
  <rect width="100%" height="100%" fill="red" />
  <circle cx="50" cy="50" r="40" fill="green" />
  <text x="55" y="62.5" font-size="30" text-anchor="middle" fill="white">SVG</text>
</svg>
```

HTML : images

- Comparaison image vectorielle / image bitmap

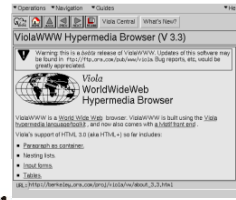


HTML - Historique

<https://www.w3.org/TR/html5/introduction.html#introduction-history>

1989

HTML 1.0



HTML début du HTML au CERN
Navigateur Mosaic

1991

HTML 2.0

HTML 3



Guerre des navigateurs - Balises propriétaires – cauchemar des développeurs

1995

HTML 3.2



Première normalisation de HTML
Syntaxe peu rigoureuse – rendu différent selon navigateurs

1998

HTML 4.0

CSS



Mise en ordre – séparation en 2 langages de la structure (HTML) et de la présentation (CSS)

1999

HTML 4.01

Quelques ajustements

2000

XHTML 1.0

Mise en conformité de HTML 4.01 avec XML (eXtensible Markup Language)

2004



Création du WHATWG (Web Hypertext Application Technology Working Group)
collaboration non officielle des différents développeurs de navigateurs web

2012

HTML 5



Working draft : Vidéo, Canvas, nouveau modèle pour éléments, APIs javascript

2014

W3C Recommendation (28/10/2014) <https://www.w3.org/TR/html5/>

2016

HTML 5.1

2017

HTML 5.2

W3C Recommendation (14/12/2017) <https://www.w3.org/TR/html52/>

May 2019 W3C cède la
gestion des standards HTML
et DOM au WATHWG

<https://html.spec.whatwg.org/>

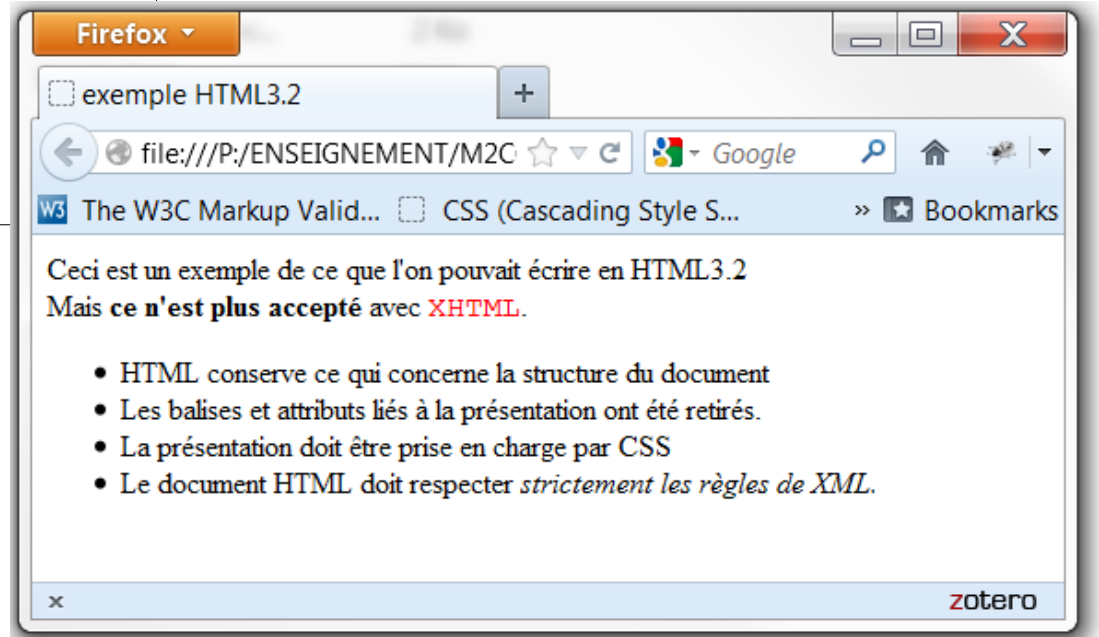
Exemple HTML 3.2

```
<html>
  <head>
    <title>exemple HTML3.2</title>
  </head>
  <body onload="alert(document.compatMode);">
    Ceci est un exemple de ce que l'on pouvait écrire en HTML3.2<BR>
    Mais <b>ce n'est plus accepté</b> avec
    <font face="Courier New, Courier, mono" color=red>XHTML</font>.
    <ul>
      <li>HTML conserve ce qui concerne la structure du document</li>
      <li>Les balises et attributs liés à la présentation ont été retirés.
      <li>La présentation doit être prise en charge par CSS
      <li>Le document HTML doit respecter <i>strictement</i>
        les règles de XML.
    </ul>
  </body>
</html>
```

Des balises de structure `` ``

Des balises de présentation ``

Une certaine liberté dans l'utilisation des balises (ex `` pas fermée)



XHTML 1.0

- Transposition en syntaxe XML de HTML 4.
 - Syntaxe plus rigoureuse qu'en HTML 4.
 - Toute balise ouverte doit être fermée
 - Balises en minuscule
 - Obligation de mettre des valeurs pour les attributs
 - Valeur d'attribut doivent être entre " " ou ' '
 - Suppression des balises et attributs de présentation
 - `<i>` `` `` `<center>` `<h1 align="center">`
 - On ne garde que la structure du document
 - `<h1>` `<h2>` ... `<h6>` `<p>`... `<div>`

XML (Extensible Markup Language) un métalangage informatique de balisage générique. Sa syntaxe est dite « extensible » car elle permet de définir différents langages avec chacun leur vocabulaire et leur grammaire, comme XHTML, XSLT, RSS, SVG...

Facilite l'échange automatisé de contenus complexes entre systèmes hétérogènes (interopérabilité)

```
<?xml version="1.0"?>
<questionnaire>
  <question>
    Qui était le premier
    empereur romain ?
  </question>
  <réponse>
    Auguste
  </réponse>
  <!-- Note : tu auras besoin
  de plus de questions.-->
</questionnaire>
```

XML



XSD

XSD (XML Schema Description)
Permet de définir (et valider) la structure
d'un document XML

XSLT (eXtensible Stylesheet Language Transformations)
permet notamment de transformer un document XML
vers un autre format



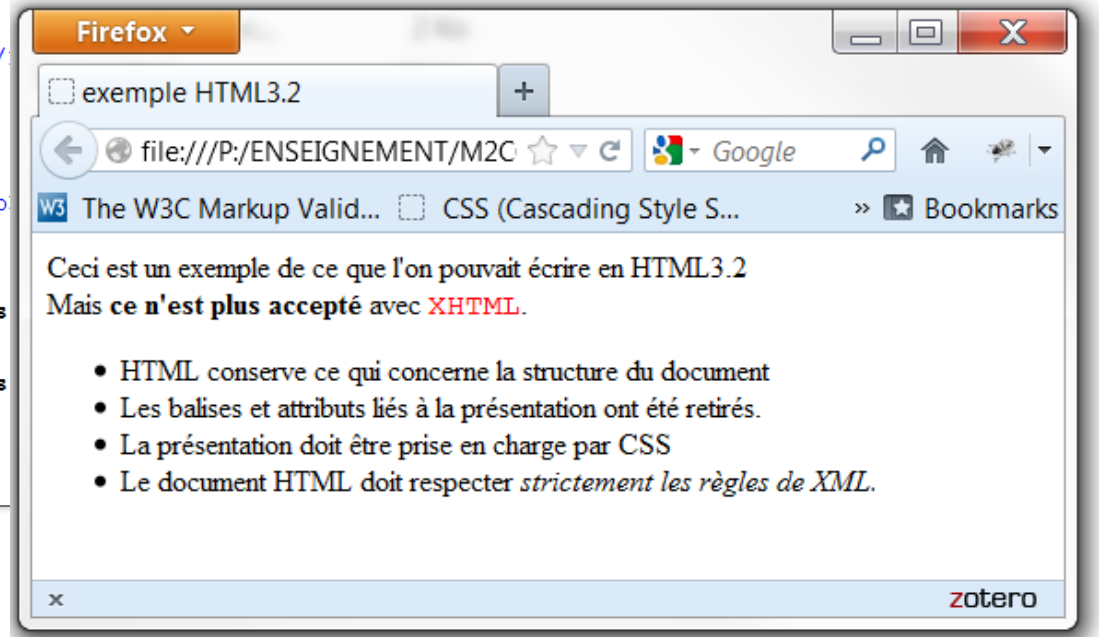
XSLT

Exemple XHTML 1.0

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr">
  <head>
    <title>exemple XHTML</title>
    <meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
    <style type="text.css">
      .codeRouge {
        color:red;
        font-family:"Courier New, Courier, mono";
      }
    </style>
  </head>
  <body onload="alert(document.compatMode);">
    <p>Ceci est une traduction en XHTML strict de l'exemple précédent<br/>
    Tout <strong>ce qui n'est plus accepté</strong> avec
    <span class="codeRouge">XHTML</span> a été corrigé et
    le document a été validé avec le validateur du W3C
    (<a href="http://validator.w3.org/">http://validator.w3.org/</a>).</p>
    <ul>
      <li>HTML conserve ce qui concerne la structure du document</li>
      <li>Les balises et attributs liés à la présentation ont été retirés</li>
      <li>La présentation doit être prise en charge par CSS</li>
      <li>Le document HTML doit respecter <em>strictement</em> les règles</li>
    </ul>
  </body>
</html>
```

DOCTYPE permet de spécifier au navigateur la version de l'HTML ou de l'XHTML utilisée

style de présentation CSS (Cascading Style Sheet)



XHTML Doctype

- DOCTYPE permet de spécifier au navigateur la version de l'HTML ou de l'XHTML utilisée
- XHTML propose 3 type de documents :

- mode strict
 - respect total de la sacrosainte règle du W3C : séparation contenu/présentation

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
```

- mode transitionnel
 - mode hybride qui accepte des balises jugées obsolètes (*deprecated*)

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

- mode frameset
 - permet de concevoir des pages comportant des cadres (*frames*).

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
```

- document XHTML pour être correct doit être:
 - bien formé (respectueux des règles de syntaxe XML)
 - valide (conforme aux spécifications du doctype)

Pour en savoir plus sur les doctypes :
http://www.w3schools.com/tags/tag_doctype.asp

HTML₅



- Dernière évolution des standards définissant HTML.
- regroupe deux concepts différents :
 - nouvelle version du langage HTML, avec de nouveaux éléments, attributs et comportements
 - nouvelles balises sémantiques pour la structure **<header>**, **<footer>**, **<article>**, **<section>**
 - nouveaux type d'inputs pour les formulaires **date**, **number**, **email**, ...
 - ...
 - un ensemble plus large de technologies qui permettent des sites web plus variés et puissants, et des applications web
 - support pour Audio/Video
 - APIs javascript
 - canvas (dessin)
 - géolocalisation
 - Stockage
 - connectivité
 - ...

<https://developer.mozilla.org/fr/docs/Web/Guide/HTML/HTML5>

HTML₅ : balises sémantiques

- Les balises sémantiques permettent de définir clairement la signification des éléments à la fois pour le navigateur (browser) et le développeur. Elles faciliteront la mise en page à l'aide des CSS.

<header></header>

introduction du document entier (en-tête de page) ou d'une section, d'un article,...

<footer></footer>

conclusion du document entier (pied de page), ou d'une section, d'un article,...

<nav></nav>

division pour les liens de navigation (le menu)

<section> </section>

division générique regroupant un même sujet

<article> </article>

division de contenu indépendant (ex. article de blog)

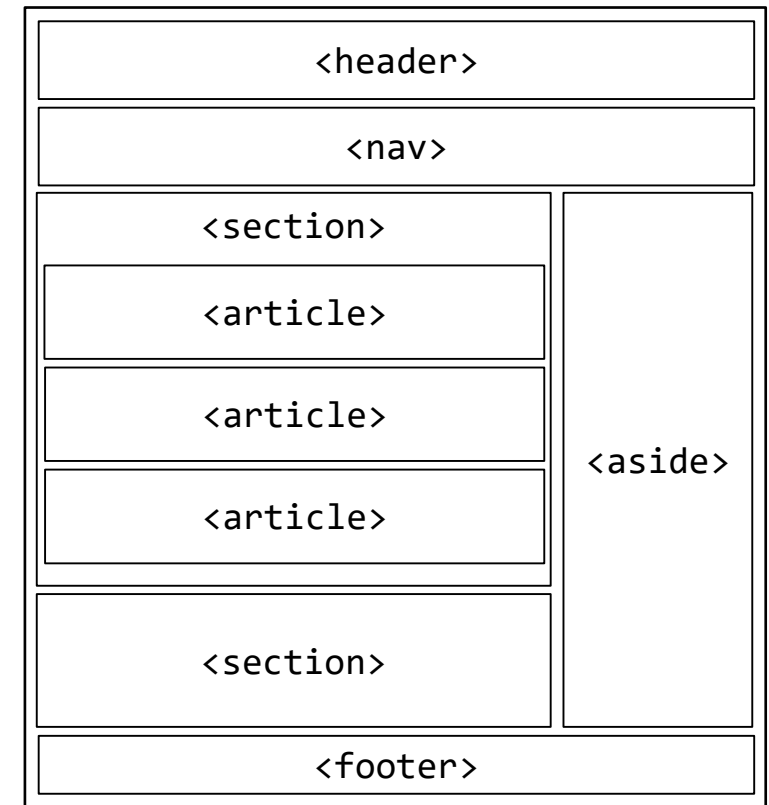
<aside></aside>

aparté en français, représente une partie d'un document dont le contenu n'a qu'un rapport indirect avec le contenu principal du document<main></main>

<details></details>

élément de divulgation de détails

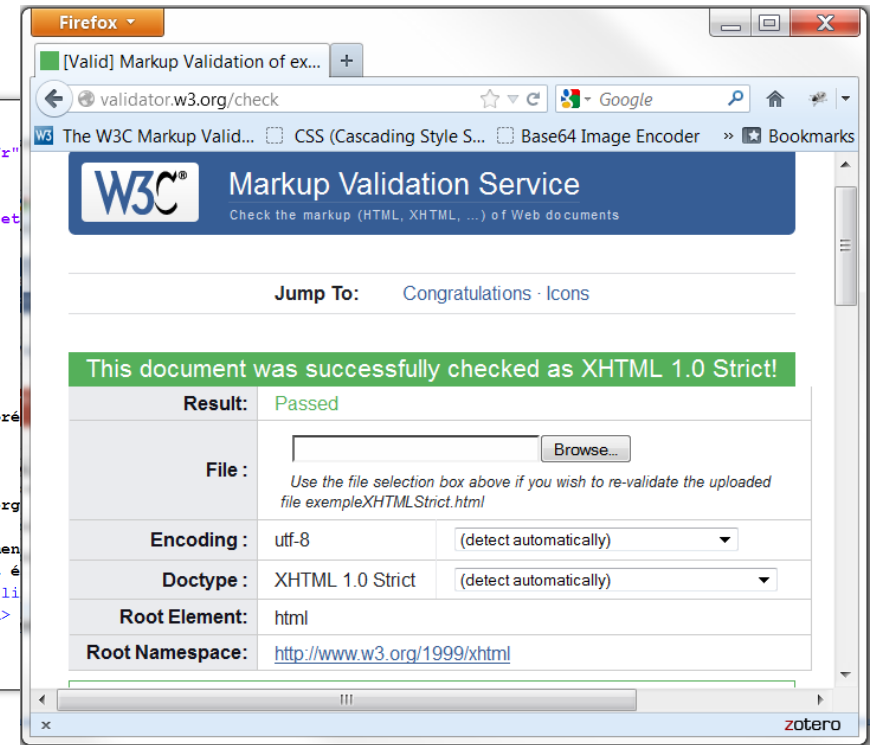
...



Validation de documents HTML

- permet de tester la correction d'un document HTML
 - assurance d'un affichage cohérent entre les principaux navigateurs
 - accessibilité des pages , y compris aux internautes souffrant d'un handicap
 - démonstration de vos capacités professionnelles (code de qualité respectant les normes)
- Valideur du W3C
 - <http://validator.w3.org>
 - <http://validator.w3.org/nu>

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="fr" lang="fr"
<head>
  <title>exemple XHTML</title>
  <meta http-equiv="Content-Type" content="text/html; charset=
  <style type="text.css">
    .codeRouge {
      color:red;
      font-family:"Courier New, Courier, mono";
    }
  </style>
</head>
<body onload="alert(document.compatMode);">
  <p>Ceci est une traduction en XHTML strict de l'exemple pré
  Tout <strong>ce qui n'est plus accepté</strong> avec
  <span class="codeRouge">XHTML</span> a été corrigé et
  le document a été validé avec le valideur du W3C
  (<a href="http://validator.w3.org/">http://validator.w3.org
  <ul>
    <li>HTML conserve ce qui concerne la structure du document
    <li>Les balises et attributs liés à la présentation ont été
    <li>La présentation doit être prise en charge par CSS</li>
    <li>Le document HTML doit respecter <em>strictement</em>
  </ul>
</body>
</html>
```



Pour conclure

- HTML5 est la dernière version du langage à balises HTML (HyperText Markup Language) servant à décrire des documents web. Les éléments du langage HTML5 servent principalement à :
 - structurer le texte : divisions, titres, paragraphes, tableaux, listes...
 - inclure des objets externes : images, sons, vidéos
 - réaliser des liens hypertextes
- Quelques ressources
 - le standard HTML (les spécifications de WHATWG)
 - <https://html.spec.whatwg.org/multipage/>
 - Les ressources de Mozilla Developers Network
 - <https://developer.mozilla.org/fr/docs/Web/HTML>
 - Les tutoriaux de w3schools :
 - <https://www.w3schools.com/html/default.asp>

Philippe GENOUD (LIG-Steamer)
Philippe.Genoud@imag.fr

HTML et encodage des caractères

dernière modification : 28/09/2023 10:19



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

HTML Encodage des caractères

- du texte HTML peut être distribué sur le réseau en utilisant différents encodage des caractères.
- balise `<meta>` dans l'en tête indique au navigateur quel jeu de caractères (et encodage) est utilisé dans le document



```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
    <title>LAT-1</title>
  </head>
  <body>
    Vivement l'été
  </body>
</html>
```



```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="UTF-8">
    <title>UTF-8</title>
  </head>
  <body>
    Vivement l'été
  </body>
</html>
```



Utiliser UTF-8 valeur recommandée par les standards du web

placer la spécification du **charset** au début de la partie **head**.

HTML Encodage des caractères

- jeu de caractères (charset) ≠ encodage
- jeu de caractères:
 - liste de caractères identifiés avec un numéro unique (point de code)
- encodage de caractères
 - algorithme qui traduit un point de code en binaire

Bits		Column													
b ₇	b ₆	b ₅	b ₄	b ₃	b ₂	b ₁	b ₀	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	NUL	DLE	SP	0	@	P	\	p
0	0	0	0	1	1	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	0	1	0	0	0	0	STX	DC2	"	2	B	R	b	r
0	0	0	1	1	0	0	0	ETX	DC3	#	3	C	S	c	s
0	0	1	0	0	0	0	0	EOT	DC4	\$	4	D	T	d	t
0	0	1	0	1	0	0	0	ENQ	NAK	%	5	E	U	e	u
0	0	1	1	0	0	0	0	ACK	SYN	&	6	F	V	f	v
0	0	1	1	1	0	0	0	BEL	ETB	'	7	G	W	g	w
1	0	0	0	0	0	0	0	BS	CAN	(8	H	X	h	x
1	0	0	0	1	0	0	0	HT	EM)	9	I	Y	i	y
1	0	0	1	0	0	0	0	LF	SUB	*	:	J	Z	j	z
1	0	0	1	1	0	0	0	VT	ESC	+	;	K	[k	{
1	0	1	0	0	0	0	0	FF	FS	,	<	L	\	l	
1	0	1	0	1	0	0	0	CR	GS	-	=	M]	m	}
1	0	1	1	0	0	0	0	SO	RS	.	>	N	^	n	~
1	0	1	1	1	0	0	0	SI	US	/	?	O	_	o	DEL

ASCII chart from a pre-1972 printer manual
<https://en.wikipedia.org/wiki/ASCII>

exemple : ASCII
 (American Standard code for Information Interchange)

point code

66^{ème} caractère du code ASCII

→ 65

⚠ on compte à partir de 0

A

encodage

4 1
 $2^2 2^1 2^0$ $2^3 2^2 2^1 2^0$
 100 0001

Hexadécimal
 16 nombres représentés par 0 ... 9 A B C E F

Binaire

6 B
 k 110 1101

$65 = 2^6 + 2^0$

HTML Encodage des caractères

- Code ASCII insuffisant pour représenter les caractères de langues autres que l'anglais
- Définition d'autres jeux de caractères pour supporter différentes langues

ISO 8859-1 (Latin1)

Jeu de caractères ASCII										iso-8859-1											
+	0	1	2	3	4	5	6	7	8	9	+	0	1	2	3	4	5	6	7	8	9
30			!	"	#	\$	%	&	'		160		ı	ç	£	¥	ı	š	ˆ	ø	
40	()	*	+	,	-	.	/	0	1	170	ª	«	¬	-	®	-	°	±	²	³
50	2	3	4	5	6	7	8	9	:	;	180	˜	µ	¶	·	¸	¹	º	»	¼	½
60	<	=	>	?	@	A	B	C	D	E	190	¸	¿	À	Á	Â	Ã	Ä	Å	Æ	Ç
70	F	G	H	I	J	K	L	M	N	O	200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
80	P	Q	R	S	T	U	V	W	X	Y	210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
90	Z	[\]	^	_	`	a	b	c	220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
100	d	e	f	g	h	i	j	k	l	m	230	æ	ç	è	é	ê	ë	ì	í	î	ï
110	n	o	p	q	r	s	t	u	v	w	240	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù
120	x	y	z	{		}	~				250	ú	û	ü	ý	þ	ÿ				

ISO 8859-5 (Cyrillique)

Jeu de caractères ASCII										iso-8859-5											
+	0	1	2	3	4	5	6	7	8	9	+	0	1	2	3	4	5	6	7	8	9
30			!	"	#	\$	%	&	'		160		Ё	Ђ	Ѓ	Е	Ѕ	І	Ї	Ј	Љ
40	()	*	+	,	-	.	/	0	1	170	Њ	Ћ	Ќ	-	Ў	Ц	А	Б	В	Г
50	2	3	4	5	6	7	8	9	:	;	180	Д	Е	Ж	З	И	Й	К	Л	М	Н
60	<	=	>	?	@	A	B	C	D	E	190	О	П	Р	С	Т	У	Ф	Х	Ц	Ч
70	F	G	H	I	J	K	L	M	N	O	200	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	а	б
80	P	Q	R	S	T	U	V	W	X	Y	210	в	г	д	е	ж	з	и	й	к	л
90	Z	[\]	^	_	`	a	b	c	220	м	н	о	п	р	с	т	у	ф	х
100	d	e	f	g	h	i	j	k	l	m	230	ч	ц	ш	щ	ъ	ы	ь	э	ю	я
110	n	o	p	q	r	s	t	u	v	w	240	Ѕ	ё	ђ	ѓ	е	ѕ	і	ї	ј	љ
120	x	y	z	{		}	~				250	њ	ќ	ќ	ѕ	ў					

⚠ Attention
contrairement
à la table du code
ASCII slide
précédent, ces tables
donnent les points de
code et non pas
l'encodage

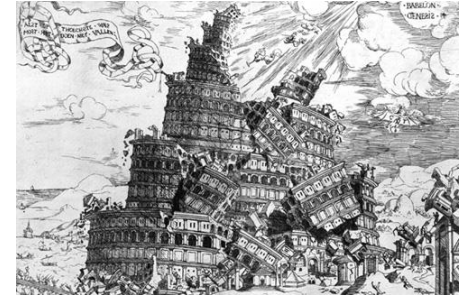
a → 97

https://gilles-hunault.leria-info.univ-angers.fr/selfhtml_fr/internationalisation/jeux_caracteres.htm

ANSI (Windows-1252) jeu de caractères de Windows. Identique à ISO-8859-1 (certains caractères de contrôle plus utilisés ont été remplacés par des symboles utiles € , œ ...).

HTML Encodage des caractères

- problèmes des jeux de caractères ISO
 - Prolifération des jeux de caractères : ISO-8859-2 (Latin-2) (langues d'Europe centrale : croate, serbe, ...), ISO-8859-3 (Latin-3) (galicien, le maltais, l'espéranto...)
 - Un même point de code et donc un même encodage peuvent correspondre à des caractères différents
 - Impossibilité de supporter les langages non alphabétiques (ex. idéogrammes chinois)



Cornelis Anthonisz (1547)

ISO 8859-1 (Latin1)

Jeu de caractères ASCII											iso-8859-1										
+	0	1	2	3	4	5	6	7	8	9	+	0	1	2	3	4	5	6	7	8	9
30			!	"	#	\$	%	&	'		160		ı	ı	ı	ı	ı	ı	ı	ı	ı
40	()	*	+	,	-	.	/	0	1	170	ª	«	¬	-	»	-	°	±	²	³
50	2	3	4	5	6	7	8	9	:	;	180	ˆ	µ	¶	·	,	1	º	»	¼	½
60	<	=	>	?	@	A	B	C	D	E	190	¸	˘	ˆ	ˆ	ˆ	ˆ	ˆ	ˆ	ˆ	ˆ
70	F	G	H	I	J	K	L	M	N	O	200	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ
80	P	Q	R	S	T	U	V	W	X	Y	210	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û
90	Z	[\]	^	_	`	a	b	c	220	Ü	Ý	Þ	ß	à	á	â	ã	ä	å
100	d	e	f	g	h	i	j	k	l	m	230	æ	ç	è	é	ê	ë	ì	í	î	ï
110	n	o	p	q	r	s	t	u	v	w	240	ä	ñ	ò	ó	ô	õ	ö	÷	ø	ù
120	x	y	z	{		}	~				250	ú	û	ü	ý	þ	ÿ				

ISO 8859-5 (Cyrillique)

Jeu de caractères ASCII											iso-8859-5										
+	0	1	2	3	4	5	6	7	8	9	+	0	1	2	3	4	5	6	7	8	9
30			!	"	#	\$	%	&	'		160		Ë	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ
40	()	*	+	,	-	.	/	0	1	170	Њ	Ћ	Ќ	-	Ў	Џ	А	Б	В	Г
50	2	3	4	5	6	7	8	9	:	;	180	Д	Е	Ж	З	И	Й	К	Л	М	Н
60	<	=	>	?	@	A	B	C	D	E	190	О	П	Р	С	Т	У	Ф	Х	Ц	Ч
70	F	G	H	I	J	K	L	M	N	O	200	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я	а	б
80	P	Q	R	S	T	U	V	W	X	Y	210	в	г	д	е	ж	з	и	й	к	л
90	Z	[\]	^	_	`	a	b	c	220	м	н	о	п	р	с	т	у	ф	х
100	d	e	f	g	h	i	j	k	l	m	230	ч	ц	ш	щ	ъ	ы	ь	э	ю	я
110	n	o	p	q	r	s	t	u	v	w	240	Њ	ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ
120	x	y	z	{		}	~				250	Њ	ћ	ќ	ѕ	џ	џ				

A : 65 → 2⁶ + 1 → 0100 0001
 4 1
 hexadécimal

1110 1001 → 2⁷ + 2⁶ + 2⁵ + 2³ + 2⁰ → 233
 E 9
 hexadécimal

→ é ISO 8859-1
 → щ ISO 8859-5

HTML Encodage des caractères

- caractères Unicode jeu de caractère universel

Unicode est un standard informatique qui permet des échanges de textes dans différentes langues, à un niveau mondial. Il est développé par le Consortium Unicode, qui vise au codage de texte écrit en donnant à tout caractère de n'importe quel système d'écriture un nom et un identifiant numérique, et ce de manière unifiée, quelle que soit la plate-forme informatique ou le logiciel utilisés.



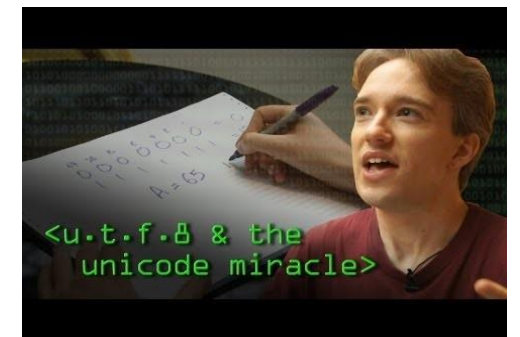
<https://home.unicode.org/about-unicode/>

v15 (sept 2022)
149 186 caractères

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0000	NUL	SOH	STX	ETX	END	ACK	DEL	BS	HT	LF	VT	FF	CR	SO	CI	
0010	DL	DC1	DC2	DC3	DC4	DC5	DC6	DC7	DC8	DC9	DC10	DC11	DC12	DC13	DC14	DC15
0020		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
0030	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0040	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0050	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
00E0	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
00F0	ð	ñ	ò	ó	ô	õ	÷	ø	ù	ú	û	ü	ý	þ	ÿ	
0440	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я

<https://unicode-table.com/>

	point de code	
	décimal	hexadécimal
A	65	0041
é	233	00E9
Щ	1097	0449
😎	128 526	1F60E



<https://www.youtube.com/watch?v=MijmeoH9LT4>

HTML Encodage des caractères

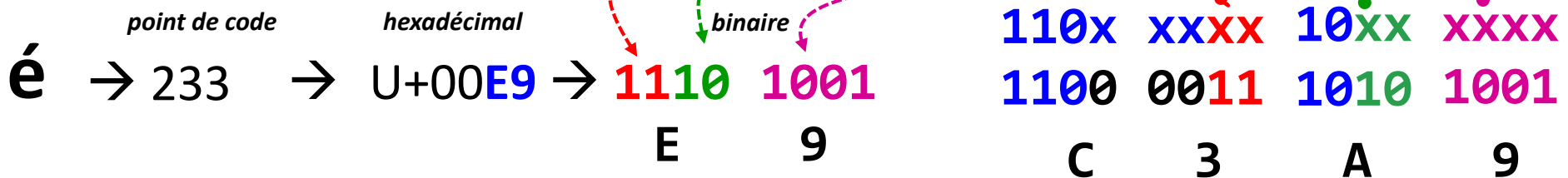
- caractères Unicode :
 - v 15.0 (sept. 2022) : 149 186 caractères (https://unicode.org/faq/basic_q.html)
 - un octet ne suffit plus
- encodages UTF (*Unicode Transformation Format*)
 - codage de taille variable
 - UTF-8 : 1 , 2 , 3 ou 4 octets (préconisé pour le web)
 - UTF-16 : 2 ou 4 octets (codage des chaînes en Java et dans l'autres langages de programmation)
 - codage de taille fixe
 - UTF-32: 4 octets

https://en.wikipedia.org/wiki/Comparison_of_Unicode_encodings

HTML Encodage des caractères

- encodages UTF-8
 - codage de taille variable
 - UTF-8 : 1 , 2 , 3 ou 4 octets

Number of bytes	Bits for code point	First code point	Last code point	Byte 1	Byte 2	Byte 3	Byte 4
1	7	<small>0</small> U+0000	<small>127</small> U+007F	0xxxxxxx			
2	11	<small>128</small> U+0080	<small>2047</small> U+07FF	110xxxxx	10xxxxxx		
3	16	<small>2048</small> U+0800	<small>65535</small> U+FFFF	1110xxxx	10xxxxxx	10xxxxxx	
4	21	<small>65536</small> U+10000	<small>1 114 111</small> U+10FFFF	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx



HTML Encodage des caractères

```
<!DOCTYPE html>
<html>
  <head>
    <title>TEST</title>
  </head>
  <body>
    Vivement l'été
  </body>
</html>
```

UTF-8

```
$ od -t x1 testUTF8.html
```

```
00000000 3c 21 44 4f 43 54 59 50 45 20 68 74 6d 6c 3e 0d
00000200 0a 3c 68 74 6d 6c 3e 0d 0a 20 20 20 20 3c 68 65
00000400 61 64 3e 0d 0a 20 20 20 20 20 20 20 20 3c 74 69
00000600 74 6c 65 3e 54 45 53 54 3c 2f 74 69 74 6c 65 3e
00001000 0d 0a 20 20 20 20 3c 2f 68 65 61 64 3e 0d 0a 20
00001200 20 20 20 3c 62 6f 64 79 3e 0d 0a 20 20 20 20 20
00001400 20 20 20 56 69 76 65 6d 65 6e 74 20 6c 27 c3 a9
00001600 74 c3 a9 0d 0a 20 20 20 20 3c 2f 62 6f 64 79 3e
00002000 0d 0a 3c 2f 68 74 6d 6c 3e
00002111
```

Lat-1
ISO 8859-1

```
$ od -t x1 testLAT1.html
```

```
00000000 3c 21 44 4f 43 54 59 50 45 20 68 74 6d 6c 3e 0d
00000200 0a 3c 68 74 6d 6c 3e 0d 0a 20 20 20 20 3c 68 65
00000400 61 64 3e 0d 0a 20 20 20 20 20 20 20 20 3c 74 69
00000600 74 6c 65 3e 54 45 53 54 3c 2f 74 69 74 6c 65 3e
00001000 0d 0a 20 20 20 20 3c 2f 68 65 61 64 3e 0d 0a 20
00001200 20 20 20 3c 62 6f 64 79 3e 0d 0a 20 20 20 20 20
00001400 20 20 20 56 69 76 65 6d 65 6e 74 20 6c 27 e9 74
00001600 e9 0d 0a 20 20 20 20 3c 2f 62 6f 64 79 3e 0d 0a
00002000 3c 2f 68 74 6d 6c 3e
```

HTML Encodage des caractères

```
<!DOCTYPE html>
<html>
  <head>
    <title>TEST</title>
  </head>
  <body>
    Vivement l'été
  </body>
</html>
```

UTF-8

```
$ od -t x1 testUTF8.html
```

```
00000000 3c 21 44 4f 43 54 59 50 45 20 68 74 6d 6c 3e 0d
00000200 0a 3c 68 74 6d 6c 3e 0d 0a 20 20 20 20 3c 68 65
00000400 61 64 3e 0d 0a 20 20 20 20 20 20 20 20 3c 74 69
00000600 74 6c 65 3e 54 45 53 54 3c 2f 74 69 74 6c 65 3e
00001000 0d 0a 20 20 20 20 3c 2f 68 65 61 64 3e 0d 0a 20
00001200 20 20 20 3c 62 6f 64 79 3e 0d 0a 20 20 20 20 20
00001400 20 20 20 56 69 76 65 6d 65 6e 74 20 6c 27 c3 a9
00001600 74 c3 a9 0d 0a 20 20 20 20 3c 2f 62 6f 64 79 3e
00002000 0d 0a 3c 2f 68 74 6d 6c 3e
00002111
```

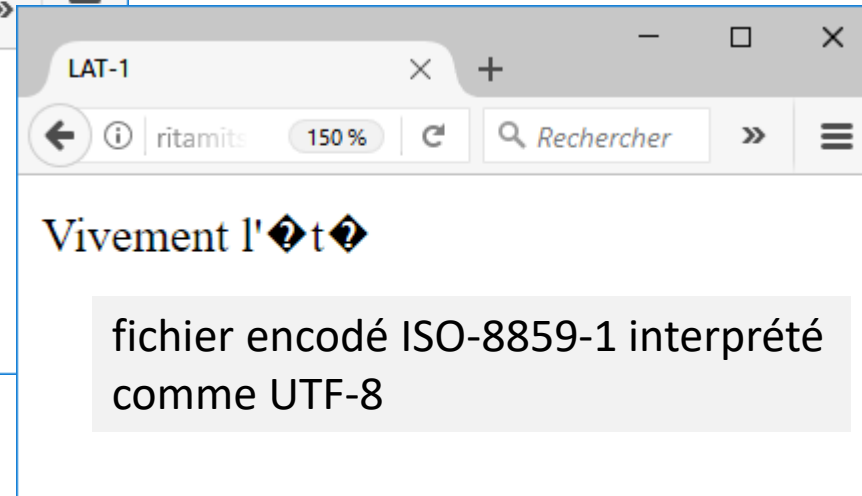
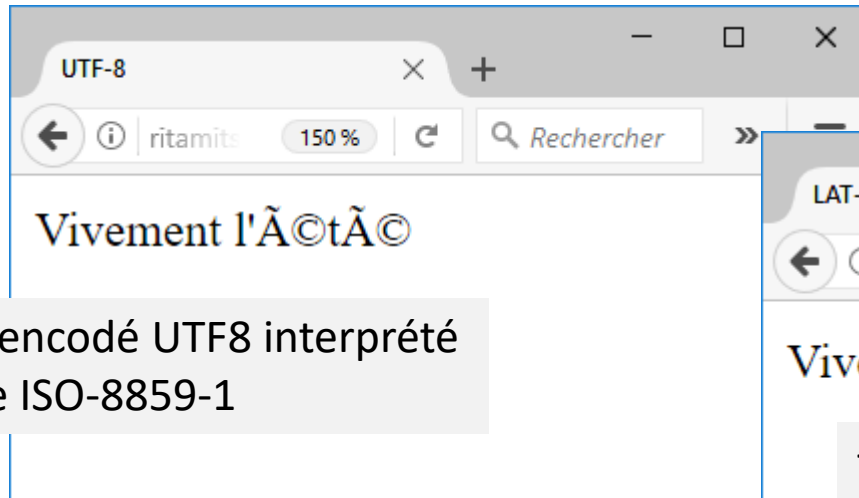
Lat-1
ISO 8859-1

```
$ od -t x1 testLAT1.html
```

```
00000000 3c 21 44 4f 43 54 59 50 45 20 68 74 6d 6c 3e 0d
00000200 0a 3c 68 74 6d 6c 3e 0d 0a 20 20 20 20 3c 68 65
00000400 61 64 3e 0d 0a 20 20 20 20 20 20 20 20 3c 74 69
00000600 74 6c 65 3e 54 45 53 54 3c 2f 74 69 74 6c 65 3e
00001000 0d 0a 20 20 20 20 3c 2f 68 65 61 64 3e 0d 0a 20
00001200 20 20 20 3c 62 6f 64 79 3e 0d 0a 20 20 20 20 20
00001400 20 20 20 56 69 76 65 6d 65 6e 74 20 6c 27 e9 74
00001600 e9 0d 0a 20 20 20 20 3c 2f 62 6f 64 79 3e 0d 0a
00002000 3c 2f 68 74 6d 6c 3e
```

HTML Encodage des caractères

- problèmes d'encodage
 - pas de déclaration d'encodage et l'encodage choisi par le navigateur n'est pas le bon
 - la déclaration d'encodage n'est pas la bonne



pour comprendre menu encodage dans firefox
<https://hsivonen.fi/encoding-menu>

https://www.w3schools.com/html/html_charset.asp

https://lig-membres.imag.fr/genoud/teaching/PL2AI/cours/exemples/01_HTML/encodage/



Le serveur lig-membres.imag.fr ajoute un header content-type avec charset='utf-8' donc le navigateur interprète toutes les pages comme encodées en UTF-8 quelle que soit la balise `<meta charset="...">`

HTML Encodage des caractères

```
<!DOCTYPE html>
<html>
  <head>
    <title>TEST</title>
  </head>
  <body>
    Vivement l'été
  </body>
</html>
```

UTF-8

```
ritamitsouko:genoud% od -t x1 test-UTF8.html
0000000 3c 21 44 4f 43 54 59 50 45 20 68 74 6d 6c 3e 0d
0000020 0a 3c 68 74 6d 6c 3e 0d 0a 20 20 20 20 3c 68 65
0000040 61 64 3e 0d 0a 20 20 20 20 20 20 20 20 3c 74 69
0000060 74 6c 65 3e 54 45 53 54 3c 2f 74 69 74 6c 65 3e
0000100 0d 0a 20 20 20 20 3c 2f 68 65 61 64 3e 0d 0a 20
0000120 20 20 20 3c 62 6f 64 79 3e 0d 0a 20 20 20 20 20
0000140 20 20 20 56 69 76 65 6d 65 6e 74 20 6c 27 c3 a9
0000160 74 c3 a9 0d 0a 20 20 20 20 3c 2f 62 6f 64 79 3e
0000200 0d 0a 3c 2f 68 74 6d 6c 3e 0d 0a
0000213
```

Lat-1
ISO 8859-1

```
ritamitsouko:genoud% od -t x1 testANSI.html
0000000 3c 21 44 4f 43 54 59 50 45 20 68 74 6d 6c 3e 0d
0000020 0a 3c 68 74 6d 6c 3e 0d 0a 20 20 20 20 3c 68 65
0000040 61 64 3e 0d 0a 20 20 20 20 20 20 20 20 3c 74 69
0000060 74 6c 65 3e 54 45 53 54 3c 2f 74 69 74 6c 65 3e
0000100 0d 0a 20 20 20 20 3c 2f 68 65 61 64 3e 0d 0a 20
0000120 20 20 20 3c 62 6f 64 79 3e 0d 0a 20 20 20 20 20
0000140 20 20 20 56 69 76 65 6d 65 6e 74 20 6c 27 e9 74
0000160 e9 0d 0a 20 20 20 20 3c 2f 62 6f 64 79 3e 0d 0a
0000200 3c 2f 68 74 6d 6c 3e 0d 0a
0000211
```