

STANDARDS ET NORMES DE REPRÉSENTATION DES DONNÉES

<http://www.larbiquezouli.com>

Présenté par Dr Larbi GUEZOULI

STANDARDS ET NORMES DE REPRÉSENTATION DES DONNÉES

- Normes et Standards [1 Cours]
- Caractéristiques pour les documents numériques [1 Cours]
- Les niveaux de standardisation/normalisation d'un document numérique [1 Cours]
- Caractéristiques des processus de normalisation/standardisation pour le document numérique [1 Cours]
- Exemple de l'évolution des pratiques professionnelles [1 Cours]

2

NORMES ET STANDARDS

"La norme est un document établi par un consensus et approuvé par un organisme reconnu, qui, fournit, pour des usages communs et répétés, des règles, des lignes directrices ou des caractéristiques, pour des activités ou leurs résultats, garantissant un niveau d'ordre optimal dans un contexte donné" (Eric Sutter)

3

NORMES ET STANDARDS

1. Normes VS Standards

Une **norme** est un ensemble de **règles** de conformité ou de fonctionnement **légalisé** par un organisme de normalisation, comme l'ISO.

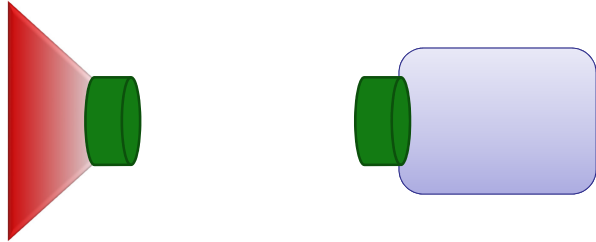
Un **standard** est un ensemble de **recommandations** ou de préférences préconisées par un groupe d'utilisateurs spécifiques et avisés.

Le défi des normes et standards est de définir un **vocabulaire**, des **structures** de données ainsi que des descriptions de **modèles** et de **scénarios** communs, afin de favoriser le travail **collaboratif** et d'établir une référence commune.

4

NORMES ET STANDARDS

1. Normes VS Standards (suite)



5

NORMES ET STANDARDS

2. Organismes de normalisation

Il existe plusieurs organismes de normalisation:

- International Organization for Standardization - ISO
- Commission Electrotechnique Internationale - IEC
- International Telecommunication Union - ITU
- National Information Standards Organization - NISO
- World Wide Web Consortium - W3C

6

NORMES ET STANDARDS

• Organisation internationale de normalisation - ISO

C'est une organisation internationale non gouvernementale.

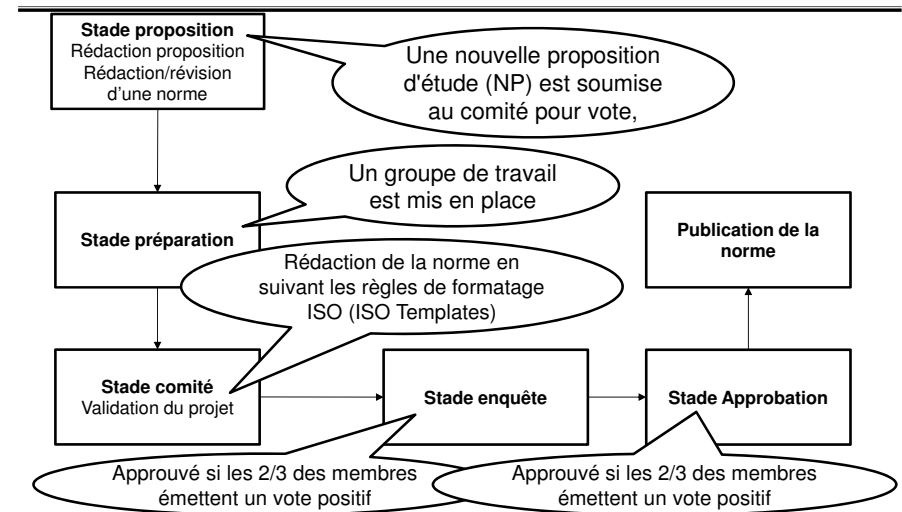
Elle débute en 1946 lorsque les délégués de 25 pays se sont réunis à l'Institute of Civil Engineers à Londres.

Elle compte aujourd'hui 162 pays membres, dont l'Algérie avec son organisme national IANor.



7

NORMES ET STANDARDS



NORMES ET STANDARDS

• Commission Electrotechnique Internationale - IEC

La Commission électrotechnique internationale ou International Electrotechnical Commission (IEC), est l'organisation internationale de normalisation chargée des domaines de l'électricité, de l'électronique, de la compatibilité, l'électromagnétique et de la nanotechnologie.

La IEC a été créée en 1906 et possède [69 pays membres](#).



9

NORMES ET STANDARDS

• International Telecommunication Union - ITU

L'union internationale des télécommunications (ITU) est une organisation fondée à Paris en 1865 sur les partenariats public-privé.

Elle compte aujourd'hui [193 pays membres](#).

Elle a son siège à Genève en Suisse.



10

NORMES ET STANDARDS

• World Wide Web Consortium - W3C

C'est un organisme de standardisation, fondé en octobre 1994.

Il est chargé de soutenir la compatibilité des technologies du Web telles que: HTML5, HTML, XHTML, XML, RDF, CSS, PNG, SOAP...

En 2013, il regroupait [383 entreprises partenaires](#).



11

NORMES ET STANDARDS

• National Information Standards Organization - NISO

Le NISO est un organisme dédié aux normes dans le domaine de l'édition, des bibliothèques et de l'accès à l'information.

Il est fondé en 1939 sous la forme d'un comité de l'ANSI. Il devient une structure indépendante en 1982 et prend le nom de NISO en 1984.



12

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

1. Identifiants des documents numériques

Les identifiants les plus utilisés pour les documents numériques, livres, journaux : ISBN, ISSN, DOI, OAI...

Ils permettent d'identifier chaque document numérique de **manière unique**.

13

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISBN

International Standard Book Number (ISBN) ou Numéro international normalisé du livre (NINL)

C'est un numéro international qui permet d'identifier, de manière unique, chaque édition de chaque livre publié en format numérique ou papier.

Il est utilisé pour les ouvrages et non pas pour les journaux ...

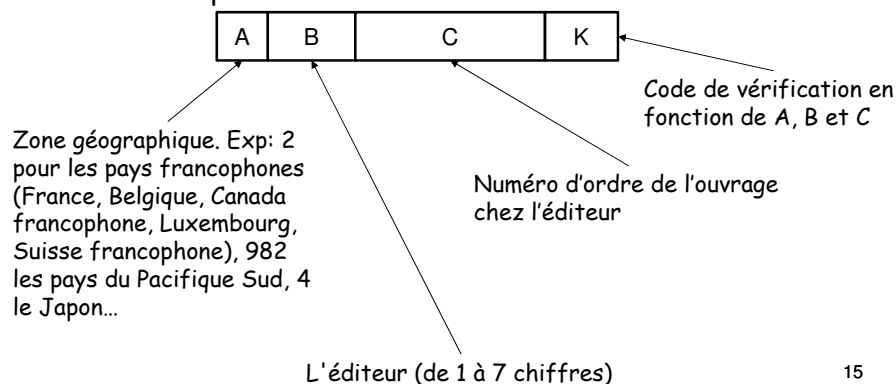
On peut trouver plusieurs ISBN pour un même livre s'il est diffusé en différents formats: papier, numérique (PDF)...

14

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISBN (suite)

Jusqu'à l'année 2006, le code ISBN était sur 10 chiffres divisés en 4 parties A-B-C-K



15

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISBN (suite)

Exp:

ISBN 2-271-05256-4

C'est un livre publié en France (A=2),

par l'éditeur CNRS (B=271)

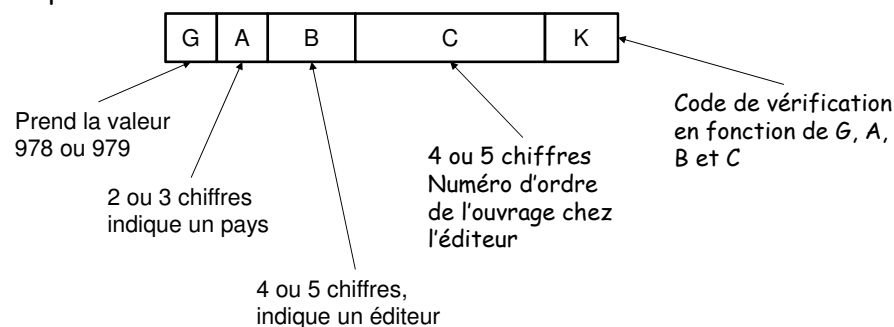
dont le n° d'ordre de ce livre chez cet éditeur est 5256.

16

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISBN (suite)

A partir de 2007, le code ISBN contient 13 chiffres, inspiré du code barre (norme EAN-13). Il est divisés en 5 parties G-A-B-C-K



17

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISBN (suite)

G = 978 indique que le code à le même format que ISBN10

Exp: ISBN 2-271-05256-4

devient ISBN 978-2-271-05256-8.

G = 979 utilisé par les nouveaux éditeurs ou par les anciens qui ont épuisés leur plage de codage ISBN10.

A=10 est le nouveau code pour les pays francophones.

L'Algérie en 2018 n'a pas encore de code sur 13 chiffres.

18

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISBN (suite)

Affectation des identifiants

L'Agence francophone pour la numérotation internationale du livre (AFNIL) attribue les codes ISBN à tous les éditeurs français, belges francophones et des pays d'Afrique francophones.

Le reste des pays reçoivent leurs codes par le système d'attribution mondial, organisé par l'International ISBN Agency, qui est localisé à Londres.

19

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISSN

L'International Standard Serial Number (ISSN) identifie une publication en série (journal, périodique ou collection).

Il est lié au titre. Si la revue change de titre, un nouveau numéro doit être attribué.

Une revue diffusée en plusieurs formats doit avoir plusieurs ISSN (autant de codes que de formats).

20

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISSN (suite)

Le code est formé de 2 parties de 4 chiffres chacune, séparées par un tiret.

Exp: ISSN 0994-7647.

Le dernier chiffre est un chiffre de contrôle, il prend la valeur X pour 10.

Les codes ISSN sont indépendants du pays d'origine.

21

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

ISSN (suite)

Affectation des identifiants

Le ISSN International Centre, situé à Paris, permet d'affecter les codes.

Lors de la demande, des informations (nom de la revue, pays, années de publication, éditeur, URL ou email...) sur la publication doivent être enregistrées.

22

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

DOI

Digital Object Identifier, est un mécanisme d'identification de ressources numériques documentaires ou autres.

C'est un identifiant qui ne dépend pas de l'emplacement de la publication. C'est aussi un mécanisme pour retrouver la ressource en ligne (son URL) à partir de son identifiant DOI.

Certaines métadonnées sont enregistrées lors de l'affectation du DOI.

Le serveur suivant permet de retrouver l'emplacement d'un document à partir de son DOI:

<http://dx.doi.org/>

23

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

DOI (suite)

Le code DOI est de la forme : doi:préfixe/suffixe.

Le préfixe est attribué à une organisation (bureau d'enregistrement).

Le suffixe est l'identifiant de la ressource numérique dans cette organisation.

Exp:

doi: 10.5802/aif.2754 identifie un article par le suffixe "aif.2754" dans l'organisation CEDRAM qui a le préfixe 10.5802.

24

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

OAI

Open Archives Initiative, aide à l'interrogation d'archives de documents numériques tel que HAL (<https://hal.archives-ouvertes.fr/>).

L'identifiant a la forme :

oai: "nom d'un fournisseur de données, d'une archive": "identifiant de la ressource dans l'archive"

Exp: oai:hal.archives-ouvertes.fr:hal-00819683

25

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

2. Caractéristiques de fichier numérique

Un fichier numérique est caractérisé par un ensemble de paramètres techniques relatives à l'encodage de textes, d'images, de son, de vidéo, et autres types de données d'une ressource.

Ces caractéristiques sont formatés selon la norme MARC 21 (**M**ACHINE-**R**eadable **C**ataloguing).

26

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

2. Caractéristiques de fichier numérique (suite)

#a - Type de fichier;

#b - Format d'encodage: schéma, norme, etc. utilisée pour l'encodage du contenu numérique;

#c - Taille du fichier: en nombre d'octets;

#d - Résolution: en pixels;

#e - Encodage régional: Un code d'identification d'une région;

#f - Débit binaire codé: La vitesse à laquelle la lecture de fichiers audio, vidéo ou autres en mode continu est conçue;

27

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

2. Caractéristiques de fichier numérique (suite)

#0 - Numéro normalisé ou de contrôle d'une notice d'autorité

#2 - Source

#3 - Documents précisés

#6 - Liaison

#8 - Numéro de liaison de zone et de séquence

Exp: #a fichier audio #b MP3 #f 32 kbps #2 rda

#a fichier vidéo #b DVD-vidéo #e région 4 #2 rda

28

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

3. Caractérisation par contenu

Documents textuels

Un document textuel est caractérisé par extraction des mots clés à partir de son contenu.

Étapes à suivre:

- Élimination des mots vides
- Remplacer les termes restants par leurs origines
- Garder chaque terme avec ses positions dans le document, ses voisins, ...

29

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

3. Caractérisation par contenu (suite)

Images

Une image peut être caractérisée de différentes manières:

- Couleurs;
- Texture;
- Points d'intérêts;
- Contours;
- ...

30

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

3. Caractérisation par contenu (suite)

Documents audio

Une séquence audio peut être caractérisée de différentes manières aussi:

- Filtrage de fréquences (passe haut, passe bas, ...);
- Résumé textuel automatique;
- ...

31

CARACTÉRISTIQUES POUR LES DOCUMENTS NUMÉRIQUES

3. Caractérisation par contenu (suite)

Documents vidéo

Une séquence vidéo peut être caractérisée de différentes manières aussi:

- Caractérisation des images clés;
- Résumé textuel automatique;
- ...

32

Nous allons voir 3 niveaux:

- Codage des caractères
- Formats de fichiers
- Métadonnées pour la description du contenu

1. Codage des caractères

Les caractères dans un fichier texte peuvent être codés sous différents format:

- ASCII
- ISO-8859 (souvent appelée Latin-1)
- UNICODE
- UTF-8, UTF-16, UTF32 (rarement utilisé)

1. Codage des caractères (suite)

ASCII: American Standard Code for Information Interchange. Il est basé sur l'alphabet anglais.

Il est codé sur 8 bits.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	Codes de contrôles ou codes inutilisés															
1																
2		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	

1. Codage des caractères (suite)

ISO-8859-1: Souvent appelée Latin-1.

Il existe plusieurs tables:

ISO-8859-2

...

ISO-8859-16

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
8	Codes de contrôles ou codes inutilisés															
9																
A		ı	ç	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ö	ñ	ò	ó	ô	õ	÷	ø	ù	ú	û	ü	ý	þ	ÿ	

1. Codage des caractères (suite)

UNICODE:

Proposé par le consortium UNICODE issu de l'association de constructeurs d'ordinateurs.

Face aux limites du codage sur 8 bits, ils propose d'utiliser plusieurs octets pour coder un caractère.

Avec le standard UNICODE on utilise 16 bits pour coder un caractère.

37

1. Codage des caractères (suite)

UTF-8:

Unicode transformation format. C'est un format de longueur variable d'octets.

Dans UTF-8, on remarque que :

- un octet de bit de poids fort 0 code un caractère ASCII sur un octet;
- un octet de bits de poids fort 11 est le premier octet d'un caractère codé sur 2 octets (nbOctets = nb 1 de poids fort);
- un octet de bits de poids fort 10 est à l'intérieur d'un caractère codé sur plusieurs octets.

38

1. Codage des caractères (suite)

UTF-8:

Représentation binaire	Signification
0xxxxxx	1 octet codant 1 à 7 bits
110xxxx 10xxxxx	2 octets codant 8 à 11 bits
1110xxxx 10xxxxx 10xxxxx	3 octets codant 12 à 16 bits
11110xxx 10xxxxx 10xxxxx 10xxxxx	4 octets codant 17 à 21 bits

Caractère	Commentaire	Hex	Numérique	Binaire	Binaire UTF-8
A		x41	65	1000001	01000001
Ù		xDA	0218	11011010	11000011 10011010
€	(Monnaie) signe Euro	x20AC	8364	10000010101100	11100010 10000010 10101100
♩	(Musique) Clé de sol	x1D11E	119070	11101000100011110	11110000 10011101 10000100 10011110

1. Codage des caractères (suite)

UTF-16:

C'est un format de longueur variable de mots (2 octets).

Quand l'octet de poids le plus fort est au début on parle de **big endian**.

Et l'inverse est appelé **little endian**.

40

1. Codage des caractères (suite)

UTF-16:

UTF-16 en BE (big endian)

Bits	00000000	000uuuuu	xxxxxyy	yyyyyyyy
UTF-16 BE (2 octets)	si	uuuuu = 00000	xxxxxyy	yyyyyyyy
UTF-16 BE (4 octets)	110110ww	wwxxxxxx	110111yy	yyyyyyyy
	avec www = uuuuu - 1 si uuuuu > 00000			

UTF-16 en LE (little endian)

Bits	00000000	000uuuuu	xxxxxyy	yyyyyyyy
UTF-16 LE (2 octets)	si	uuuuu = 00000	yyyyyyyy	xxxxxyy
UTF-16 LE (4 octets)	wwxxxxxx	110110ww	yyyyyyyy	110111yy
	avec www = uuuuu - 1 si uuuuu > 00000			

41

2. Formats des fichiers

- MSWord
- Format RTF (Rich Text Format)
- PostScript («.ps»)
- PDF (Portable Document Format)
- Formats de documents semi-structurés:
 - SGML : Standard Generalized Mark-Up
 - XML : Extensible Mark-up Language
 - HTML : Hypertext Mark-up Language
 - ...

42

2. Formats des fichiers (suite)

Formats de fichiers multimédias:

GIF : Graphics Interchange Format
 JPEG : Joint Photographics Expert Group
 TIFF : Tagged-Image File Format
 PNG : Portable Network Graphics
 QuickTime Movies
 AVI : Audio-Video Interleaved
 MPEG : Moving Picture Expert Group
 RealVideo
 WAVE
 RealAudio

43

3. Métadonnées pour la description du contenu

La norme Dublin Core

Contient 15 éléments de données descriptifs:

Le titre	Le créateur
L'éditeur	Le sujet
La description	La source
La langue	La relation
La couverture	La date
Le type	Le format
L'identificateur	Le collaborateur
Les droits.	

44

3. Métadonnées pour la description du contenu (Suite)

RDF : Resource Description Framework

C'est un modèle standard qui est élaboration par le W3C.
XML est la syntaxe d'encodage.

Comme le XML est très utilisé, nous allons le voir dans
ce qui suit plus en détail.

XML: eXtensible Markup Language

C'est un est un langage informatique de balisage
générique.

Les balises permettent de structurer de manière
hiérarchisée et organisée les données d'un document.

Une balise porte un nom qui est entouré de chevrons <>.

En XML, on distingue 2 types de balises : les balises par
paires et les balises uniques.

XML: eXtensible Markup Language (suite)

Les balises par paires sont composées en réalité de 2
balises que l'on appelle ouvrantes et fermantes.
<balise>Contenu</balise>.

Une balise unique est en réalité une balise par paires qui
n'a pas de contenu. <balise />

XML: eXtensible Markup Language (suite)

Règles de nommage des balises :

- Les noms peuvent contenir des lettres, des chiffres
ou des caractères spéciaux.
- Les noms ne peuvent pas débuter par un nombre ou un
caractère de ponctuation.
- Les noms ne peuvent pas commencer par les lettres
XML.
- Les noms ne peuvent pas contenir d'espaces.
- Eviter les caractères - , ; . < et >

XML: eXtensible Markup Language (suite)

Les attributs:

Une balise peut contenir des attributs.

Un attribut peut se décrire comme une option ou une donnée cachée.

Exp: <prix devise="euro">25.3</prix>

Quelques règles pour les attributs :

- Les règles de nommage sont les mêmes que pour les balises.
- La valeur d'un attribut doit impérativement être délimitée par des guillemets, simples ou doubles.
- Les attributs d'une balise doivent être différents.

49

XML: eXtensible Markup Language (suite)

Les commentaires:

Un commentaire est un texte qui permet de donner une indication et ne sera pas traité par l'interpréteur comme les balises.

Exp:

<!-- Ceci est un commentaire ! -->

50

XML: eXtensible Markup Language (suite)

Structure d'un document XML:

Un document XML peut être découpé en 2 parties : le **prologue** et le **corps**.

Le prologue correspond à la première ligne du doc XML.

Exp: <?xml version = "1.0" encoding="UTF-8" standalone="yes" ?>

version: La version de XML;

encoding: Jeu de caractères utilisé dans le doc XML;

standalone: Le document XML est autonome ou un autre document lui est rattaché.

51

XML: eXtensible Markup Language (suite)

Structure d'un document XML: (suite)

Le corps contient l'ensemble des balises qui décrivent les données.

Toutes les balises du document XML doivent être incluse à l'intérieur d'une seule balise appelée **racine**.

Exp: <?xml version = "1.0" encoding="UTF-8" standalone="yes" ?>

```
<racine>
  <balise1>texte</balise1>
  <balise2>texte</balise2>
  <balise3/>
</racine>
```

52

XML: stands for eXtensible Markup Language (suite)

Document XML bien formé

- La version 1.1 du XML, le prologue est bien renseigné.
- Une seule balise racine.
- Le nom des balises et des attributs sont conformes aux règles de nommage.
- Les balises en paires sont correctement fermées.
- Les valeurs des attributs sont entre guillemets simples ou doubles.
- Les balises ne se chevauchent pas, il existe une arborescence dans le corps du document XML.

53

XML: stands for eXtensible Markup Language (suite)

DTD (Document Type Definition)

Un document XML est rarement utilisé seul mais souvent avec souvent avec d'autres technologies.

La technologie DTD est un ensemble de règles pour définir l'architecture du document XML, ainsi que la hiérarchie qui existe entre les balises.

Une règle est définie:

```
<!ELEMENT balise (contenu)>
```

balise: nom de la balise.

54

XML: stands for eXtensible Markup Language (suite)

DTD (Document Type Definition) (suite)

contenu: peut contenir:

- #PCDATA: veut dire que le contenu est une valeur simple
- Nom d'une autre balise qui doit être dans la balise de la règle
- EMPTY: veut dire que la balise est vide.
- ANY: veut dire les 3 cas précédents en même temps (c'est à éviter).

55

XML: stands for eXtensible Markup Language (suite)

DTD (Document Type Definition) (suite)

Exp:

```
<!ELEMENT personne (nom)>
```

```
<!ELEMENT nom #PCDATA>
```

```
<personne>
```

```
  <nom>John DOE</nom>
```

```
</personne>
```

```
<!ELEMENT personne (nom)>
```

```
<!ELEMENT nom EMPTY>
```

```
<personne>
```

```
  <nom ></nom>
```

```
</personne>
```

56

XML: stands for eXtensible Markup Language (suite)
DTD (Document Type Definition) (suite)

D'autres cas existent pour contenu:

- **Séquence** `<!ELEMENT balise (balise2, balise3, etc.)>`
- **Liste de choix** `<!ELEMENT balise (balise2| balise3| etc.)>`
- **Balise optionnelle** `<!ELEMENT balise (balise2, balise3?, etc.)>`
- **Balise répétée optionnelle** `<!ELEMENT balise (balise2, balise3*, etc.)>`
- **Balise répétée** `<!ELEMENT balise (balise2, balise3+, etc.)>`

57

XML: stands for eXtensible Markup Language (suite)
DTD (Document Type Definition) (suite)

Les attributs

`<!ATTLIST balise attribut type mode>`

balise: nom de la balise concernée

attribut: nom de l'attribut concerné

type: (val1 | val2 | ...) liste de valeurs possibles

CDATA texte non parsé

ID identifiant unique

58

XML: stands for eXtensible Markup Language (suite)
DTD (Document Type Definition) (suite)

mode: #REQUIRED attribut obligatoire

#IMPLIED attribut optionnel

"val" valeur par défaut

#FIXED "val" valeur constante

Exp:

`<!ATTLIST personne nom CDATA #REQUIRED>`

`<!ATTLIST objet devise CDATA #FIXED "Euro">`

`<objet devise="Dollar" ></objet>` (invalide)

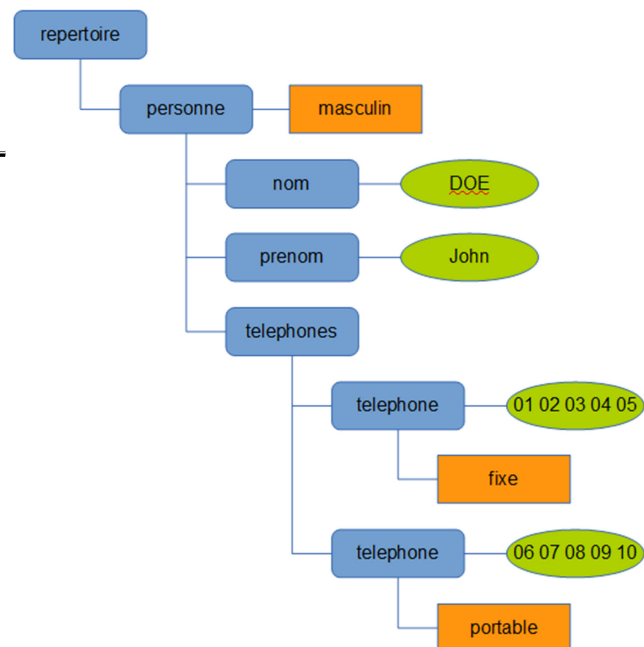
59

XML: stands for eXtensible Markup Language (suite)
DOM: Parseur XML

Document Object Model est une bibliothèque intégrée dans les différents langages de programmation (C++, JAVA, PHP, C#, ...).

Le parseur DOM permet de **structurer** le fichier XML sous forme d'un arbre pour pouvoir récupérer les informations et les manipuler dans le langage utilisé.

60



61

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

1. Processus d'élaboration d'une norme en France

Les Etapes	Le rôle de l'expert
Identification d'un besoin	Peut exprimer un besoin de norme
Evaluation d'opportunité	
Programmation collective	Participe à la commission de normalisation
Création d'une CN	
Elaboration d'un document de travail	Défend son point de vue et contribue à la rédaction du projet
Consultation formelle Consensus des experts	
Enquête probatoire	Emet des commentaires
Elaboration du projet final	Contribue à la rédaction de la version finale
Homologation et publication	

18 mois

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

2. Processus d'élaboration d'une norme en Europe

Les Etapes au CEN	Le rôle de l'expert
Proposition de travaux statut, dates cibles	Exprime son avis
Inscription au programme de travail	Participe à la commission de normalisation
Elaboration du projet	Défend son point de vue et Contribue à la rédaction du projet
Enquête CEN (5 mois)	Emet une position assortie d'éventuels commentaires techniques
Vote formel (2 mois)	Exprime son avis
Adoption de la norme européenne	

3 ans

UAP (5 mois)

63

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

3. Processus d'élaboration d'une norme Internationale

Les Etapes à l'ISO	Le rôle de l'expert
Proposition de travaux (NP) statut, dates cibles	Exprime son avis
Inscription au programme de travail	Participe à la commission de normalisation
Elaboration du projet	Défend son point de vue et Contribue à la rédaction du projet
Enquête CD (3 mois)	Exprime son avis
Enquête DIS (5 mois)	Emet une position assortie d'éventuels commentaires techniques
Vote FDIS (2 mois)	Exprime son avis
Adoption Publication	

3 ans

64

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

4. Processus de normalisation complexe et transverse

Le **processus de normalisation** des articles scientifiques par exemple qui sont accessibles sur le net possède des **prérequis**.

Il a besoin:

- des normes du multimédia.
- des normes documentaires.
- des normes du commerce électronique.

65

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

4. Processus de normalisation complexe et transverse (suite)

Ces prérequis rend le processus de normalisation **complexe**.

On a besoin de :

- Se réorganiser
- Coopérer avec d'autres groupes de travail
- Établir un nouveau comité stratégique d'orientation

66

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

5. Analyse de contenu

L'analyse du contenu de documents numériques est caractérisée par:

- la création de résumés et sa normalisation
- la traduction du contenu en langages d'indexation et de recherche.

67

CARACTÉRISTIQUES DES PROCESSUS DE NORMALISATION/STANDARDISATION POUR LE DOCUMENT NUMÉRIQUE

5.1. Normalisation de résumé

Au niveau international, en **1976** est apparue une norme pour la rédaction des résumés dans des disciplines spécialisées.

Cette norme **attire l'attention** sur les éléments essentiels à faire figurer dans le résumé.

Un **type de résumé**, qui convient à la **recherche par ordinateur**, a été mis au point au Bureau International du Travail (BIT).

Les **descripteurs** affectés au document sont **incorporés** dans le texte du résumé et **marqués** par des barres.

68

CARACTÉRISTIQUES DES PROCESSUS DE
NORMALISATION/STANDARDISATION POUR
LE DOCUMENT NUMÉRIQUE

5.2. Traduction en langages d'indexation et de recherche

Dans le but de faciliter la recherche, le contenu doit être traduit en une **représentation d'indexation**.

Il s'agit d'établir un vocabulaire d'indexation avec une **syntaxe simple et conventionnelle**.

Des **règles** sont en cours d'élaboration sur le plan international, pour l'indexation des documents à l'aide de descripteurs.

69

CARACTÉRISTIQUES DES PROCESSUS DE
NORMALISATION/STANDARDISATION POUR
LE DOCUMENT NUMÉRIQUE

5.2. Traduction en langages d'indexation et de recherche (suite)

Broad System of Ordering (BSO) propose, dans son programme UNISIST, des termes d'indexation sous forme de thésaurus.

Un thésaurus est un répertoire structuré de termes (mots-clés) pour l'analyse de contenu et le classement de documents.

70

CARACTÉRISTIQUES DES PROCESSUS DE
NORMALISATION/STANDARDISATION POUR
LE DOCUMENT NUMÉRIQUE

5.2. Traduction en langages d'indexation et de recherche (suite)

Exp:

Thèse: Mohamed BAHROUN, « Construction de thesaurus pour la recherche interactive dans une base d images satellitaires », Maroc, 2011

Thèse: François Lagunas, « Construction de thésaurus de mots composés », Paris, 2004

...

71