Théorie de l'ingénierie des documents numériques

Ingénierie Documentaire
http://doc.crzt.fr

Stéphane Crozat

Contributions : Bruno Bachimont

14 septembre 2016

Table des matières

I - **Introduction à l'ingénierie documentaire** 4

1. Concept de document 4

2. Concept d'ingénierie documentaire 6

II - **L'essence du numérique** 7

1. Théorie du support et de la connaissance inscrite 7

2. La raison computationnelle 9

3. Ça a été manipulé ! 10

III - **Le document numérique entre signe et calcul** 11

1. Le paradigme logique 11

2. Le paradigme sémiotique 12

3. Les limites du paradigme logique 13

4. Les limites de l'approche sémiotique 14

5. Le paradigme documentaire 15

IV - **Le principe du balisage documentaire** 17

1. Balises et poignées de calcul 17

2. La structuration logique 17

3. Exemple de structuration logique 18

V - **Bibliographie** 20

VI - **Compléments** 21

VII - **Exercices** 22

1. Questions théoriques 23

1.1. Exercice : 23
1.2. Exercice : 24
1.3. Exercice : 24
1.4. Exercice : Ingénierie des connaissances et des contenus (commentaire de texte) 24
1.5. Exercice : Document numérique, ingénierie documentaire et archivage. 25

Abréviations 26

Bibliographie 27

Introduction à l'ingénierie documentaire

I

1. Concept de document

Fondamental

Un document est une *inscription* de *contenus* sur un *support* pérenne, établie dans un *contexte de production* et pour un *contexte de réception*.

Définition : Contenu

Un contenu est une forme d'expression pourvue d'une valeur culturelle associée à un *véhicule matériel* , il exprime une signification et suscite une réception et une interprétation.

Définition : Inscription

Une inscription est un contenu fixé sur un *support matériel*, tel qu'il lui apporte une permanence dans le temps.

Exemple : Exemple de contenu

Une définition donnée oralement est un contenu.

Une transmission hertzienne d'un flux audiovisuel est un contenu.

Exemple : Exemple d'inscription

Une définition écrite est une inscription.

Un enregistrement magnétique d'une transmission hertzienne d'un flux audiovisuel est une inscription.

Exemple : Exemple de document

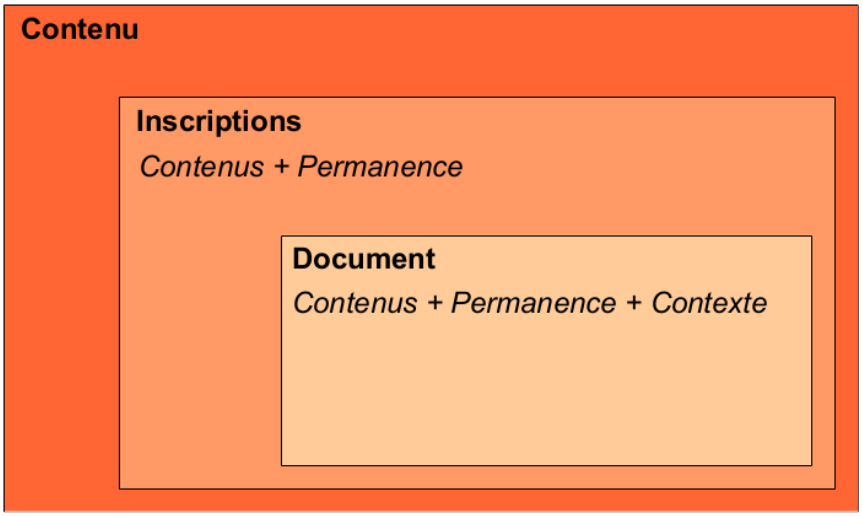
Un glossaire de définitions publié sur un livret, par un chercheur pour sa communauté est un document.

Une émission de télévision, produite par une chaîne pour ses téléspectateurs, enregistrée sur un support magnétique est un document.

Reformulations

Contenu	Forme d'expression
Inscription	Contenu + Permanence
Document	Inscription + Contexte

Document = Contenu + Permanence + Inscription



Notion de document

Fondamental : Caractéristiques du document

Un document est pourvu de trois propriétés fondamentales :

- *La fermeture* : Un document est délimité dans le temps et dans l'espace, il a un début et une fin, l'on sait de quoi il est constitué (parties), il est scénarisé (sa forme prescrit une ou plusieurs linéarités explicites, qui se s'imposent ou se proposent au lecteur).
- *L'intentionnalité* : Un document est intentionnel, il est destiné à un usage, il naît d'une intention auctoriale et éditoriale, il est produit dans un contexte d'écriture et publié pour un contexte de lecture.
- *La publication* : Un document est publié, il est rendu public sous une forme matérielle donnée et figée, son contexte de production et de diffusion est identifiable et objectivable.

Le document est pourvu de deux facettes :

- C'est un *objet technique* considéré pour ses propriétés physique qui relève d'une manipulation technique
- C'est un *objet culturel* considéré pour sa signification qui relève d'une interprétation

Remarque : Document "au sens large"

Notons que les définitions précédentes laissent de côté la notion de document comme objet dont l'intentionnalité est reconstruite *a posteriori*, au sens où un objet archéologique est un document pour l'archéologue par exemple.

2. Concept d'ingénierie documentaire

Fondamental : Ingénierie documentaire

L'objet de l'ingénierie documentaire est la conception de systèmes techniques permettant et optimisant l'articulation de la manipulation technique et de l'interprétation culturelle des documents.

Définition : Ingénierie des documents numériques

L'ingénierie des documents numériques est le sous-ensemble de l'ingénierie documentaire, dont l'objet est la construction de systèmes informatiques et qui ne considère que les documents dont le support est numérique.

Fondamental : Problématique de l'ingénierie des documents numériques

L'essence computationnelle du numérique fait que les documents qu'il supporte *« ont été manipulés »*, ce qui interroge théoriquement les propriétés fondamentales du document (fermeture, intentionnalité, publication).

La question de la manipulation informatique est donc au cœur de l'ingénierie des documents numériques, qui s'attache d'une part à optimiser ces potentialités manipulatoires et d'autre part à protéger en pratique les propriétés fondamentales du document mises à mal par le numérique.

L'essence du numérique

II

1. Théorie du support et de la connaissance inscrite

La théorie du support, est une théorie philosophique de la connaissance, issue notamment de la phénoménologie de Husserl *, développée notamment à l'UTC à travers des philosophes comme Bernard Stielger (1994a *, 1994b *) et Bruno Bachimont (1996 *, 2004 *).

Le concept général de la théorie du support est que toute connaissance ne peut procéder que d'une inscription sur un support matériel.

Cette idée est développée en particulier autour du cas du support numérique.

Définition : Connaissance

Une connaissance est la capacité d'exercer une action pour atteindre un but.

- Une connaissance pratique renvoie à une activité dans le monde matériel (notion savoir-faire), elle est corrélé un objet pratique (ce sur quoi porte l'action pratique).
- Une connaissance théorique renvoie à la possibilité de produire ou reformuler des énoncés dans un code de communication (notion de savoir), elle est corrélée à un objet matériel, mais pour ce qu'il représente et non pour ce qu'il est (par exemple de l'encre sur du papier).

La connaissance pratique porte sur une modification physique du monde, la connaissance théorique porte sur une modification de notre représentation du monde.

(Bachimont, 2004 *, pp65,67)

Définition : La connaissance inscrite

Toute connaissance repose sur une *inscription*, dont elle est l'interprétation :

- la connaissance est l'interprétation de l'inscription
- l'inscription est la matérialisation de la connaissance

Ainsi un livre, une cassette vidéo ou un logiciel véhiculent des connaissances, sans les modéliser au sens de la logique formelle ou de l'IA *, mais en leur offrant un support de *mémorisation* et de *manifestation*.

La transmission de connaissance résulte de son inscription intentionnelle sur un support par un auteur pour un lecteur.

Complément : Le support comme prothèse

Le support technique devient alors une *prothèse* cognitive, un processus d'extériorisation, qui permet l'anticipation et la constitution de modes nouveaux de représentation : *« par pro-thèse nous entendrons toujours à la fois : posé devant, ou spatialisation (é-loignement); et posé d'avance, déjà là (passé) et anticipation (prévision), c'est à dire temporalisation. (Stiegler, 1994 *) ».*

Le support est donc un moyen de spatialiser l'information, pour la rejouer dans le temps.

Remarque

Le support est à prendre au sens le plus large, pour soutenir une telle thèse, il faut en effet élargir la notion d'inscription à tout ancrage matériel de la connaissance, et en particulier au corps :

- Une inscription est corporelle quand le support technique est le corps
- Une inscription est externe quand le support est externe au corps

(Bachimont, 2007 * , p255)

Fondamental : Le supplément

La signification de l'inscription est conditionnée par les *propriétés matérielles* du support :

- Le support impose un *supplément* à l'inscription car il ajoute de l'intelligibilité.
- Selon le type de support, le supplément d'intelligibilité sera différent.
- On aura donc une *rationalité spécifique par type de support.*

Complément : Les thèses de la théorie du support

La théorie du support s'articule autour de la thèse centrale suivante :

Les propriétés du substrat physique d'inscription, et du format physique de l'inscription, conditionnent l'intelligibilité de l'inscription.

Elle comprend en outre les thèses suivantes :

1. une connaissance est la capacité d'effectuer une action dans un but donné.
2. un objet technique prescrit par sa structure matérielle des actions. L'objet technique est l'inscription matérielle d'une connaissance.
3. toute connaissance procède d'une genèse technique. Seule la répétition, prescrite par les objets techniques, de l'action permet d'engendrer la connaissance comme capacité à exercer une action possible.
4. la connaissance, engendrée par la technique, prescrit une transformation dans le monde des choses (l'objet technique est alors un instrument) ou une explicitation dans le monde des représentations (l'objet technique est alors une inscription sémiotique).

5. une pensée est une reformulation effectuée par la conscience sur le support corporel qu'est le corps propre. Penser, c'est s'écrire. Toute pensée, comprise comme reformulation a pour cible de réécriture le corps propre, et comme origine, le corps propre ou une inscription externe quelconque.
6. la conscience est un pur dynamisme intentionnel, source des ré-écritures considérées comme des interprétations et non comme un mécanisme.

(Bachimont, 2004 *, p77)

2. La raison computationnelle

Définition : La raison graphique

Goody nous montre que l'invention de l'écriture a modifié les schèmes de représentation de la connaissance, jusque là orale, donnant naissance à une *« raison graphique »* (Goody, 79) *. Les documents papiers ont permis la représentation spatiale de l'information, en lui donnant une permanence dans le temps.

Grâce à ces possibilités nouvelles d'inscription, de nouvelles connaissances ont pu naître de *l'émergence de représentations qui ne peuvent être formulées oralement.*

Exemple : Le tableau

Goody expose l'exemple du tableau ou de la liste qui permettent de mettre en relief des relations qui ne pourraient émerger par la description orale du tableau ou de la liste.

Soit par exemple la lecture orale suivante :

Lire le contenu du tableau ci-après à voix haute.

[cf. Exemple d'énoncé orale d'un tableau]

Soit la représentation tabulaire de la même information :

Année	Chiffre d'affaire	Bénéfice
2004	123.315	5.154
2005	115.247	7.156
2006	114.265	8.245
2007	112.250	8.300

La spatialisation permet de faire émerger des connaissances nouvelles, par exemple le fait que le fait que la baisse du chiffre d'affaire est corrélée à une hausse des bénéfices. Le changement de support a eu une influence sur la connaissance elle-même.

Exemple : Les mathématiques

Un autre exemple est celui des mathématiques, qui n'existeraient pas sans écriture.

Fondamental : Changer de support c'est changer de raisonnement

Le passage de l'oral à l'écrit n'est donc pas seulement un changement de support, c'est une révolution cognitive.

Définition : La raison computationnelle

De même que l'écrit a permis le passage du temporel au spatial par projection de la parole, le support numérique apporte de nouvelles formes de représentation des informations, basées sur le calcul. Bachimont parle de l'émergence d'une *« raison computationnelle »* (Bachimont, 2004 *).

En effet l'ordinateur ne traite que des séquences binaires qui, par le calcul, deviennent des signes sur un support tel que l'écran. C'est cette propriété du support numérique qui est fondamentale en tant qu'elle propose de nouvelles modalités d'inscription. Et ces nouvelles modalités induisent également la constitution de modes de représentation nouveaux, comme les tableaux pour la raison graphique en leur temps (Bachimont cite par exemple la couche, le réseau, etc.).

Fondamental : Enjeu

L'enjeu est alors de comprendre *comment* inscrire les informations sur ce support spécifique qu'est le numérique afin de repérer les structures d'inscription nouvelles pertinentes.

3. Ça a été manipulé !

Bachimont (2007 *, p33-34) propose de caractériser le *« noème »* du numérique (en référence à Roland Barthes à propos de la photographie : *« Ça a été »*), c'est à dire *ce qu'il faut comprendre et penser à propos du numérique*, comme : *« Ça a été manipulé »*.

C'est à dire que tout contenu numérique résulte toujours d'une construction dynamique via un calcul. Lorsque je frappe sur mon clavier un calcul transforme mon action en codage binaire et en stockage d'information dans la mémoire de l'ordinateur. Lorsque je regarde mon écran, je vois le résultat d'un calcul effectué sur le codage binaire à partir de la mémoire.

Le document numérique entre signe et calcul

III

1. Le paradigme logique

Le positivisme logique

Le positivisme logique est un courant de pensée fondé par le Cercle de Vienne (un groupe de philosophes et logiciens animé par Moritz Schlick entre 1929 et 1936).

Il pose que toute connaissance est formelle et donc se rapporte à une expression logique (au sens de la logique mathématique) (Soulez, 1986) * .

La calculabilité

La machine de Turing, inventé en 1936 par Allan Turing (à partir des travaux de Hilbert et Gödel), est un modèle abstrait d'une machine universelle permettant d'automatiser tout calcul symbolique (alors que l'ordinateur n'existe pas encore).

Elle montre que toute formalisation logique est calculable par une machine.

Complément : Exemple de machine de Turing

https://interstices.info/jcms/nn_72391/comment-fonctionne-une-machine-de-turing

L'intelligence artificielle

L'IA * est fondée en août 1956, au Dartmouth College, lors d'un séminaire organisé par Mac Carthy et auquel sont présents des figures telles que Simon, Newell, Minsky ou Shannon.

Elle se construit sur un concept de modélisation (représentations logiques des problèmes à traiter) et d'effectivité (les représentations sont calculables).

Le mouvement de l'IA (dite forte) a pour hypothèse qu'un ordinateur peut penser puisque la connaissance est formalisable et que le formalisé est calculable par une machine de Turing.

Complément : Exemples de réalisations célèbres de l'IA

- *Les langages de représentation des connaissances*

LISP en 1958 par John Mac Carthy
- *Les systèmes experts*
  DENDRAL en 1965 par Edward Feigenbaum
- *Le traitement automatique de la langue naturelle*
  ELIZA en 1966 par Joseph Weizanbaum
- *La robotique*
  SHRDLU en 1970 par Terry Winograd

Fondamental : Hypothèse du paradigme logique

Le paradigme logique considère les inscriptions numériques en tant que formalisations de la connaissance.

L'ordinateur calcule sur la connaissance formalisée.

Exemple : Formalisation logique de connaissances

```
1 Est-un (Voiture, Véhicule)
2 Possede (Voiture, Roue, 4)
3 Est-un (Ferrari, Voiture)
```

Un ordinateur pourra calculer (déduire ?) qu'une Ferrari est un véhicule et qu'il possède quatre roues.

2. Le paradigme sémiotique

Numérisation

La numérisation d'un contenu correspond à son codage (donc un calcul) de tel façon que l'on soit capable par un décodage (un autre calcul) de le restituer *tel qu'il a été codé*.

Un tel traitement correspond à une *numérisation de la forme*, car le contenu est inscrit mais non formalisé (au sens de l'IA *).

Orthothétie

Stiegler parle d'orthothétie pour exprimer la correspondance exacte entre information codée et information décodée (Stiegler, 1994) *.

Le numérique par essence n'est pas orthothétique, le calcul l'empêche, mais elle peut être restituée par un dispositif qui la simule (principe du WYSIWYG *).

Exemple : Exemple de systèmes numériques sémiotiques

- Scanner
- Traitement de texte
- etc.

Fondamental : Hypothèse du paradigme sémiotique

Le paradigme logique considère les inscriptions numériques en tant que simple numérisation de formes sémiotiques.

L'humain interprète ce qu'il perçoit.

Exemple : Présentation sémiotique de contenu

Une *voiture* est un véhicule à quatre roues.

- *Exemple* : Une Ferrari est une voiture

Un humain pourra déduire qu'une Ferrari est un véhicule et qu'il possède quatre roues. Un ordinateur ne peut que présenter cette information, il n'accède pas au contenu.

3. Les limites du paradigme logique

L'aporie de l'IA

L'hypothèse du positivisme logique qui fonde l'IA, à savoir que la connaissance est formelle et n'est que formelle, se révèle impossible à tenir. Il n'est pas possible de formaliser logiquement l'ensemble des connaissances humaines et l'intelligence humaine ne peut être réduite à une manipulation de symboles sans signification.

L'IA comme projet de création de machines pensantes est remis en cause (Winograd et Flores, 1989) *.

Des machines qui pensent aux machines qui donnent à penser

Les années 1980 voient alors la naissance d'une IA modeste (ou *« faible »*), dont l'objet n'est plus la construction de machines qui pensent, mais de *« machines qui donnent à penser »* (Bachimont, 1996 *).

Exemple : Traduction automatique

Parmi les programmes les plus ambitieux et les plus stratégiques de l'IA, celui de la traduction automatique, largement financé par la défense américaine en période de guerre froide, avait pour terrain d'application la traduction anglais-russe. Une illustration mythique d'échec de ce programme était le test de traduction de la phrase : *« The spirit is strong, but the flesh is weak »* (*« l'esprit est fort mais la chair est faible »*) qui donna après traduction automatique en russe, puis de nouveau en anglais : *« The vodka is good, but the meat is rotten »* (*« la vodka est forte, mais la viande est avariée »*).

Au delà de son caractère amusant cet exemple illustre très bien l'aporie d'une informatique symbolique qui bute sur des concepts de base tels que le contexte, la polysémie, etc.

Complément : Théorème d'incomplétude de Gödel

Gödel montre en 1931 que les mathématiques ne peuvent être réduite à la logique. Un corollaire direct est qu'il existe des connaissances qui ne sont pas représentables par la logique.

Notons que Turing a également démontré que certains problèmes sont indécidables par une machine de Turing (sans pour autant que cela remette en cause sa vision positiviste).

Complément : La chambre chinoise de Searle

L'ordinateur ne pense pas car il n'accède par au sens, ainsi que l'illustre Searle avec la métaphore de la "chambre chinoise" : Un opérateur qui recevrait des idéogrammes chinois et disposerait de procédures de traitement adapté à ces signes pourrait exécuter des opérations correctes, par application stricte des procédures (algorithmes). Pour autant cet opérateur ne comprendrait pas le chinois.

Ainsi, réaliser automatiquement des actions au sens d'une machine de Turing n'est pas penser.

Complément : Représentations locales versus représentation globale

La représentation formelle des connaissances a donné des résultats opérationnels tout à fait satisfaisant (par exemple certains systèmes experts), tant que cette représentation et que les raisonnements qui la concerne restent local à un domaine, en particulier technique (médecine par exemple).

C'est la formalisation globale du monde qui est aporétique.

4. Les limites de l'approche sémiotique

L'enregistrement orthothétique numérique est impossible

Parce que le contenu numérique est toujours calculé (*« il a été manipulé »*), ce que l'on lit n'est jamais ce qui a été écrit.

La simulation de l'enregistrement orthothétique est une falsification

L'approche consistant a simuler une orthothéticité masque le fonctionnement réel de la machine. Si les algorithmes sont suffisamment robustes cela n'aura pas de conséquence pratique, mais si au contraire une faille s'immisce cela entraîne l'incompréhension de l'utilisateur.

C'est une sous-utilisation de l'informatique

Si l'usage du support numérique reste limité au codage et décodage orthothétique de l'information, les potentialités du calcul, le propre du numérique, sont sous-utilisées.

Exemple : Photocopieuse numérique

Une approche exclusivement sémiotique équivaudrait au champ fonctionnel d'une photocopieuse, sans la certitude de la conformité de la copie à l'original fournie par la photocopieuse.

Complément : De la numérisation à l'informatisation

La question se pose donc de l'opérationnalisation, au sens du calcul, de connaissances en langue naturelle : Comment manipuler par des algorithmes des informations codées orthothétiquement ?

*« [...] le formalisme comme principe de modélisation n'est plus tenable. En revanche, le formalisme comme principe d'effectivité reste indispensable (Bachimont 1996) * »*

L'enjeu est donc d'*enrichir* le codage orthétique pour lui ajouter des propriétés qui seront exploitables par le calcul.

Bachimont parle d'informatisation, comme état supérieur de la numérisation, du point de vue de la manipulation calculatoire (Bachimont, 2004) *.

Complément : La désorientation

L'impossibilité de l'enregistrement orthothétique et le calcul comme principe de manipulation conduisent à une perte de *référence* (in-vérifiabilité théorique de ce que suis je en train de lire) avec comme conséquence potentielle la désorientation du lecteur. Une informatique documentaire se doit donc de prendre acte de cet état de fait pour l'intégrer dans sa réflexion et ses instrumentations.

5. Le paradigme documentaire

Coupler enregistrement sémiotique et modélisation logique

Le support numérique est un *outil sémiotique*, permettant de mémoriser et de restituer de l'information moyennant un *calcul*.

Une approche uniquement calculatoire butte sur les limites de la formalisation en tant que principe de représentation de la connaissance et une approche uniquement sémiotique butte sur la nature même du support qui est le calcul. D'où la nécessité de coupler les deux.

Une informatique documentaire consiste à associer une dimension spatio-temporelle, l'inscription sémiotique de contenus, et une dimension calculatoire, le modèle de manipulation de ces inscriptions. Le support numérique n'a plus vocation à modéliser la connaissance, il a vocation à la véhiculer, et l'objectif d'une informatique documentaire est de concevoir des modèles qui optimisent ce véhicule.

Fondamental : Hypothèse du paradigme documentaire

Le paradigme documentaire propose une description calculable d'inscriptions sémiotiques. Il s'agit d'un compromis entre la formalisation logique du sens et l'enregistrement sémiotique de la forme.

L'ordinateur calcule sur les descriptions formelles pour décider des modalités de présentation des inscriptions, et l'humain interprète le résultat présenté qu'il perçoit.

Exemple

```
1 <definition>
2     Une <notion>voiture</notion> est un véhicule à quatre roues.
3 </definition>
4 <exemple>
5     Une Ferrari est une voiture.
6 </exemple>
```

L'ordinateur peut répondre à une requête telle que "Je voudrais une définition de la notion voiture" en présentant l'information :

Une *voiture* est un véhicule à quatre roues.

Il n'a aucune idée de ce qu'est une voiture, mais il permet à l'humain de le savoir en calculant une présentation adéquat du contenu qui véhicule cette connaissance.

Le principe du balisage documentaire

IV

1. Balises et poignées de calcul

L'ingénierie documentaire met à profit deux thèses complémentaires :

- le contenu est numérisé dans sa forme signifiante : il est manipulable par la machine mais indépendamment du sa signification qui lui reste inaccessible ;
- le contenu est enrichi par des balises qui sont connues syntaxiquement et sémantiquement par la machine ; elle sait quoi en faire.

Fondamental

Le principe du balisage consiste à enrichir un contenu numérisé (dans sa forme sémiotique), sans l'altérer, pour lui ajouter des poignées qui vont être manipulables par l'ordinateur (logiquement).

Remarque

Le contenu est donc interprétable par l'homme et la machine, chacun via ce qui lui est destiné :

- l'humain interprète le contenu signifiant numérisé ;
- la machine interprète les balises ;

Exemple : XML

XML est une illustration de ce principe, puisque l'on va coupler une information sémiotique (texte, image, etc.) destinée à l'interprétation humaine, avec un ensemble de balises qui permettent de décrire formellement une structure documentaire qui sera alors manipulable par le calcul.

2. La structuration logique

« Un document peut être décrit comme une collection d'objets comportant des objets de plus haut niveau composés d'objets plus primitifs. Les relations entre ces objets représentent les relations logiques entre les composants du

> document. Par exemple [...] un livre est divisé en chapitres, chaque chapitre en sections, sous-sections, paragraphes, etc. Une telle organisation documentaire est appelée représentation de document structuré. (traduit depuis la préface de Structured documents *)

Définition : Structuration logique

On appelle structuration logique d'un contenu une inscription explicitant la structure de ce contenu en fonction de son organisation et des attributs intrinsèques qui le caractérisent et non en fonction de propriétés de présentation sur un support.

Définition : Document abstrait

Un document décrit par sa structure logique est appelé document abstrait, on parle aussi de *document structuré*.

Définition : Structuration physique

On appelle structuration physique ou mise en forme d'un contenu une inscription décrivant la façon dont ce contenu doit être présenté sur un support donné.

Définition : Document formaté

Un document décrit par sa structure physique est appelé document formaté, c'est en général également ce dont on parle quand on parle simplement de document.

Remarque

Il est possible de calculer une ou plusieurs structurations physiques pour une même structuration logique. Il est possible de calculer d'autant plus de structurations physiques que la structuration logique est indépendante de ses supports de présentation.

Remarque

La structuration logique est associée au fond du contenu, tandis que la structuration physique est associée à sa forme sur un support.

Complément : Voir aussi

Langages XML orienté documents (cf. p.)

3. Exemple de structuration logique

Exemple : Un exercice structuré logiquement

Soit la structuration logique d'un exercice :

```
1 Exercice = {Enonce, Question, Indice, Solution}
```

```
2 avec
3 Enonce = Soit un triangle rectangle disposant d'un angle de 30 degrés.
4 Question = Donner la valeur des autres angles du triangle.
5 Indice = La somme des angles d'un triangle est égale à 180 degrés.
6 Solution = 90 et 60 degrés.
```

Il est possible à partir de cette représentation de calculer différentes présentations. Pour l'écran on peut générer une présentation HTML, en laissant la solution en hyperlien cliquable. Pour le papier on peut générer une présentation PDF, en affichant la solution sur une page séparée de l'énoncé. Pour un usage multimédia on pourra générer une présentation SMIL, avec affichage de l'énoncé, lecture de la question, et affichage de la solution après un temps de pause.

Notons que si l'on avait choisi une des représentations physiques, plutôt que la représentation logique, il n'aurait pas été possible de générer les autres représentations.

Exemple : Un exercice mis en forme

Soit la mise en forme en HTML du même exercice :

```
1 <HTML>
2 <BODY>
3 Soit un triangle rectangle disposant d'un angle de 30 degrés. </BR>
4 <B> Donner la valeur des autres angles du triangle. </B> </BR>
5 <A HREF="ex001i01.html"> Vous avez besoin d'aide ? </A> </HR>
6 <A HREF="ex001s01.html"> Vérifier votre réponse ! </A>
7 </BODY>
8 </HTML>
```

On voit que dans ce format la structure logique n'apparaît plus explicitement et qu'il n'est plus possible d'identifier l'énoncé, la question et la solution sans comprendre le contenu.

Fondamental

L'exemple montre que l'on peut calculer la mise en forme à partir de la structure logique, *mais non l'inverse.*

Bibliographie

V

Complément : Ouvrage de référence

Ingénierie des connaissances et des contenus (Bachimont) *

Chapitre 1, 8, 9, 10, 11

Complément : Roger T. Pédauque

1. Document : forme, signe et médium, les re-formulations du numérique *
2. Le texte en jeu : Permanence et transformations du document *
3. Document et modernités, troisième texte *

Complément : Autres textes de Bachimont

- Herméneutique matérielle et artéfacture *
- Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle
  *
- Instrumentation numérique des documents : pour une séparation fonds/forme *
- Réinterroger les structures documentaires : de la numérisation à l'informatisation *

Complément : Autre texte lié

- *Numérique et ingénierie documentaire*

Compléments

VI

- *Préservation culturelle et numérique* (B. Bachimont)
- *Numérique et ingénierie documentaire* (L. Gaillard)

Exercices

VII

1. Questions théoriques

1.1. Exercice :

[60 minutes]

Chaque question sera traitée sur une demi-page environ. Vous mobiliserez vos connaissances théoriques (concepts clés du cours) ainsi que pratiques (exemples concrets). Vous veillerez particulièrement à la précision de vos énoncés et à ne pas effectuer de contre-sens ou d'erreur majeure. La qualité générale de la rédaction (et de l'orthographe) sera prise en compte.

Question 1

Pourquoi peut-on dire que la séparation entre le fond et la forme est un principe théoriquement faux, bien que mobilisé dans les chaînes éditoriales XML ?

Indice :

- La connaissance n'existe qu'inscrite, donc il n'existe pas de fond indépendant d'une forme
- Les chaînes éditoriales ne font que simuler ce principe, en favoriser une forme qui serait plus proche du "fond" - en fait de l'intention auctoriale - et plus éloignée de la mise en forme.
- C'est un principe de rationalité, cela permet d'accéder à certaines propriétés intéressantes pour l'ingénierie du contenu (réutilisation, multi-supports, etc.)
- Exemple : en pratique il est difficile d'écrire tout indépendamment d'une forme pour un auteur par exemple, l'aller-retour entre forme d'écriture et forme de publication est souvent nécessaire.

Question 2

Expliquer le principe d'une chaîne éditoriale XML basée sur la séparation entre un format de stockage orienté métier (fond) et des formats de mise en forme orientés présentation (forme).

Question 3

Pourquoi doit-on se poser la question de savoir si un document numérique est encore un document ?

Question 4

En quoi le principe du balisage est-il un élément de réponse à la problématique de l'ingénierie des documents numériques ?

1.2. Exercice : [90 minutes]

Pour chaque question, vous produirez un argumentaire en deux ou trois pages maximum. Vous mobiliserez les concepts vus en cours, dans les ouvrages de référence, et dans vos propres lectures complémentaires. Vous veillerez à un usage précis des termes employés, et à la rigueur de l'argumentaire : Vous avez un espace d'expression limité, il convient de le "rentabiliser" au mieux en ne produisant que des énoncés dignes d'intérêt pour le lecteur - ou, en clair, évitez le remplissage et l'utilisation approximative de notions mal maîtrisées !

Question 1

Quelles différences peut-on relever entre un document et une base de données ? Peut-on les exploiter de la même manière ? Contiennent-ils le même type d'information ? On reviendra pour cela sur les types de contenus, les formats et les exploitations que l'on peut en faire.

Question 2

Est-ce qu'un auteur a pleinement sa place au sein des chaînes éditoriales XML ? Peut-on anticiper des difficultés à leur mise en œuvre et donc déduire des recommandations pratiques pour l'organisation du travail et l'accompagnement des usages ? Quels seraient les métiers proches, dans le monde classique de l'édition et de la presse par exemple, du rôle attendu de l'auteur dans la chaîne éditoriale ?

1.3. Exercice : [15 minutes]

En quoi peut-on rapprocher l'indexation de l'opération de balisage que l'on peut effectuer sur un document ? Ont-elles la même finalité ? Utilisent-elles les mêmes outils ? Peut-on dire que l'une est une partie de l'autre ? Un moyen pour l'autre ?

1.4. Exercice : Ingénierie des connaissances et des contenus (commentaire de texte) [30 min]

«

> Le document numérique [...] n'est pas réellement un document au sens où le document papier en est un. Le document numérique est à proprement parler un complexe documentaire composé de ressources enregistrées, d'un dispositif de reconstruction du contenu pour l'afficher dans une forme perceptible et intelligible, et finalement des vues reconstruites. Aucune de ces composantes ne peut prétendre à elle seule d'être le document, ni même un document, en supposant que les composantes soient elles-mêmes des documents intervenant dans la construction d'un autre. En effet, la ressource est persistante car elle est enregistrée sur un support pérenne. Mais elle n'est pas consultable telle quelle sans la médiation d'un dispositif technique de lecture. Autrement dit, sans ce dispositif, la ressource reste inintelligible et imperceptible. Mais la vue reconstruite n'est pas davantage un document : si

elle est bien intelligible, elle n'est en général pas persistante, car elle ne dure que le temps de la reconstruction (affichage à l'écran, restitution sonore). Enfin, le dispositif n'est pas davantage le document même s'il contient un paramétrage permettant de configurer la ressource pour être consultée.

Dans ces conditions, le contenu affiché résulte autant de la ressource numérique que du paramétrage lié au format d'enregistrement et au dispositif de lecture. La paternité du contenu est en quelque sorte répartie entre la ressource et les conditions sous lesquelles la consulter au moyen du paramétrage adéquat du dispositif de lecture.

Mais cette distance introduite entre la vue reconstruite et la ressource enregistrée a pour conséquence que la reconstruction de la vue peut mobiliser différentes ressources, différents fragments d'origines diverses : il n'y a pas de relation intrinsèque entre la vue consultée et la ressource préservée, sinon la confiance que l'on a dans le décodeur. Autrement dit, le contenu consulté possède de lui-même une nature composite et la tendance technique propre au numérique ne peut que distendre encore ce lien entre la vue et la ressource pour rendre la vue encore plus anonyme, issue d'une reconstruction éparse et diversifiée.

(Bachimont B., Ingénierie des connaissances et des contenus, Lavoisier, 2007, pp223-224)

Question

Commentez ce texte.

Vous prendrez soin d'expliquer le point de vue de l'auteur en le remettant dans son contexte. Vous mettrez en perspective ses hypothèses en expliquant leurs conséquences pratiques ; vous vous appuierez à cet effet sur un ou plusieurs exemples concrets.

La rigueur de la rédaction et du vocabulaire utilisé sera pris en compte.

1.5. Exercice : Document numérique, ingénierie documentaire et archivage.

En quoi la préservation des documents numériques interroge-t-elle la nature des documents numériques ; et quel son rapport avec l'ingénierie documentaire ?

Mobilisez les contenus que vous avez étudiés dans le cadre de l'UV :

- Bachimont B. (2007). Ingénierie des connaissances et des contenus. Hermes.
- Bachimont B. (2012). Archivage et préservation numérique. Cours de NF29 du 19 décembre 2012, UTC.
- Crozat S. (2012). Chaînes éditoriales et rééditorialisation de contenus numériques. In "Le document numérique à l'heure du web de données", coord. L. Calderan. ADBS éditions.

Abréviations

IA : Intelligence Artificielle

WYSIWYG : What You See Is What You Get

Bibliographie

Jacques André, Richard Furuta, Vincent Quint, *Structured documents*, Cambridge University Press, 1989.

Bachimont Bruno, *Arts et sciences du numérique : ingénierie des connaissances et critique de la raison computationnelle*, Mémoire de HDR, Université de Technologie de Compiègne, 2004.

Bruno Bachimont, Stéphane Crozat, *Instrumentation numérique des documents : pour une séparation fonds/forme*, revue I3, vol 4, num 1, 2004. [http://archivesic.ccsd.cnrs.fr/sic_00001017.html]

Bruno Bachimont, *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*, Lavoisier, Hermès, 2007

Bachimont Bruno, *Herméneutique matérielle et artéfacture*, Thèse en épistémologie, 1996.

Stéphane Crozat, Bruno Bachimont, *Réinterroger les structures documentaires : de la numérisation à l'informatisation*, revue I3, vol 4, num 1, 2004. [http://archivesic.ccsd.cnrs.fr/sic_00001015.html]

Goody Jack, *La raison graphique : La domestication de la pensée sauvage*, Les éditions de minuit, 1979.

Husserl Edmund, *Idées directrices pour une phénoménologie*, Gallimard, 1950 (Ideen I, 1913).

Roger T. Pédauque, *Document : forme, signe et médium, les re-formulations du numérique*, premier texte, 2003. [http://archivesic.ccsd.cnrs.fr/sic_00000511]

Roger T. Pédauque, *Le texte en jeu : Permanence et transformations du document*, second texte, 2005. [http://archivesic.ccsd.cnrs.fr/sic_00001401]

Roger T. Pédauque, *Document et modernités*, troisième texte, 2006. [http://archivesic.ccsd.cnrs.fr/sic_00001741]

Soulez Antonia, *Manifeste du Cercle de Vienne et autres récits*, Presses Universitaires de France, 1985.

Stiegler Bernard, *La technique et le temps, Tome I : La faute d'Épiméthée*, Galilée, 1994.

Stiegler Bernard, *La technique et le temps, Tome II : La désorientation*, Galilée, 1994.

Winograd Terry, Flores Fernando, *L'intelligence artificielle en question*,Presses Universitaires de France, 1989.