

## Lecture active et annotation de documents

Yannick Prié  
LIRIS – Université Lyon 1  
M2R Info / 2005-2006

## Des annotations partout !



Numéros de chaînes derrière une télécommande



Post-it et étagères



Marques sur une prothèse en construction



Images illustrant un texte

etc.

## Une approche radicale (Kirsh)

- "Annotations are ubiquitous in everyday life ... [they] might be considered as guides in everyday life, indicating proper use of objects or space. Rather than integral properties of objects, they are communicative devices. ... Perhaps it is easier to delineate what is not an annotation?"
- **Literary text annotations:** Written in books, articles, memos, etc. These annotations add emphasis, questions, or comments to the original text with color, words, symbols, etc.
- **Notes:** Written in response to lectures, practices, discussions, readings, and other events, these connect events with comments about them. For example, a choreographer comments on the flow of a dance in her notes; these prod memory later when redesigning a scene.

<http://adrenaline.ucsd.edu/onr/index.htm>

3

## Une approche radicale (Kirsh)

- **Measurements:** Used in construction, clothing alteration, prosthetic design, and other areas, indicate places of alteration or comment on the process through scribbled numbers or calculations.
- **Tags:** Including graffiti tags, construction flags, football field lines, and other markers, these denote a division or claim on space. Also can include road lines.
- **Signs:** Directional signage, describing the location of the post office or where to park a car, comments on the use of space.
- **Directions:** "Push," "Use with caution" and other directions comment on the use of an object. Even in their most basic form as affordances, without language, directions indicate how to interact with an object in its intended form.
- **Comments:** Similar to directions, comments like "fragile" and "poisonous" communicate what proper use of an object is. Rather than instructing, provide characteristics that are meant to guide use.
- **Labels:** Labels on bottles, books, boxes ; identify an object.
- **Handles:** Guidelines for use of an object – an implicit "carry me this way."

<http://adrenaline.ucsd.edu/onr/index.htm>

4

## Pour nous

- Notion de lecture active
  - lire en écrivant - marquant
- Panorama sur l'annotation comme marque dans un espace de connaissance documentaire, largement textuel
  - hors informatique : les grandes catégories
  - en informatique : applications
- Sources principales
  - C. Marschall
  - J. Virbel & co. (Toulouse)
- Remarque :
  - Annotations de documents web : cours suivant

5

[ Bellenger, 1978 ]

## Plusieurs types de lectures

- lecture jouissance
    - goûter le texte en tant que tel
  - lecture appropriation
    - comprendre en profondeur la teneur du texte
  - lecture exploration
    - visant la recherche d'une information supposée ou sue présente dans le texte
  - lecture assimilation
    - visant une connaissance du texte en tant que tel, par exemple pour faire un compte-rendu
  - lecture d'inspiration
    - en vue de rebond vers d'autres lectures ou des réalisations externes au texte
- lecture professionnelle
- Hochon et Lévard (1994)

6

## Lecture active

- Constat
  - Des systèmes hypertextes orientés lecture, pas vraiment lecture – écriture (on connaît le lecteur/navigateur, mais pas vraiment le lecteur/annotateur).
  - Ceci est en contradiction avec tous les grands principes de l' [hypertexte historique] .
- Nécessité
  - Etudier les grandes catégories d'annotations
  - Penser et théoriser la pratique d'annotation comme utilisation du document en vue de produire d'autres documents
  - Se concentrer sur la lecture comme *lecture active* et écriture, pour que l'annotation par des pratiques individuelles devienne un moyen d'étendre les hypertextes pour le bien de tous.
  - Ce faisant, on est en plein dans les réflexions (nombreuses) sur ce qu'est la lecture électronique, ce qu'est écrire dans les systèmes hypertextes, ce qu'est la création de connaissances, etc.

7

## Dimensions de l'annotation

(forme de l'annotation)

- Annotation formelle vs informelle
  - des méta-données avec un schéma pour l'interopérabilité ... aux notes de marges.
  - possibilité de passage de l'informel vers le formel (ex : smiley)
- Annotation explicite vs tacite
  - lié à l'intelligibilité de l'annotation à son auto-suffisance
  - du « No! » au soulignement,
  - on ne va que rarement au-delà de la stricte nécessité pour la tâche en cours (pourquoi expliciter si ce n'est pas nécessaire ?) : une annotation peut être interprétable par son auteur (un certain temps), par autrui éventuellement

8

## Dimensions de l'annotation

(fonctions de l'annotation)

- Annotation comme écriture vs annotation comme lecture
  - « communier » avec le document, entre
    - à un bout, le lecteur de M. de Certeau, qui ne fait que se promener... et oublier
    - à l'autre, l'« idéal postmoderne de l'hypertexte polyvalent » : le texte comme un médium participatif dans lequel l'annotateur est écrivain.
  - l'annotation est un pont entre l'écriture et la lecture.
  - fonctions [Desmontil / Jacquin 2002] :
    - information, illustration, extension du document
    - forum
    - opérationnalisation
    - aide au processus rédactionnel
    - support de lecture

9

## Dimensions de l'annotation

(communication auteur / lecteur de l'annotation)

- Annotation hyperextensive vs extensive vs intensive (niveau de lecture du document)
  - annotations orientées liens (le texte dans son entour hypertexte, lecture superficielle)
  - notes de lecture rapide (typographie : mise en gras ou en italique par exemple, surlignage)
  - annotations orientées texte uniquement (pour une lecture approfondie)
- Annotations permanentes vs éphémère
  - note de bas de page vs note de rédaction

10

## Dimensions de l'annotation

(communication auteur / lecteur de l'annotation)

- Annotations publiées vs privées
  - privé pur (evt. diffusé)
    - note pour soi (evt. passées anonymement)
  - public – publié
    - one to many
      - note de bas de page
      - édition savante
    - many to many
      - forum
- Annotation globale vs institutionnelle vs groupe de travail vs personnelle
  - tout est possible

11

## L'« école toulousaine »

- Formes / moyens de l'annotation
- Objectifs de l'annotation

12

## Formes – moyens

- mise en valeur
  - surlignage (trait, couleur)
  - mise en forme (modification de la présentation du texte)
- liens
  - liens à partir du texte
    - référence (d'un élément de texte vers un autre, même page, pages différentes, document différents, liens simples ou multiples)
    - dictionnaire et annexes de compulsion (ajout placé en dehors du texte)
  - liens vers le texte
    - marque page
    - index et annexes d'entrées
  - liens à double sens
    - référence double
    - sommaire et annexes doubles

13

## Formes – moyens

- annotations et ajouts de modification
  - annotations d'ajout
    - ajout d'objet annotatif (quelconque)
    - ajout de post-it
  - annotations de modification
    - correction (jusqu'à mise à jour)
    - restructuration (suppression, déplacement, ... de passage)
- annotations type forums (annotation d'annotation)

14

## Lieux de l'annotation

- Dans la page (dans : souligné, autour : note de bas de page, sur : barré, à côté du texte)
- Entre les pages (signet, intercalaire)
- Hors-document (cahier de note)

15

## Objectifs de l'annotation

(taxonomisation de Veron 97)

- classifier
  - annotation = trace de restructuration du document par le lecteur
  - **hiérarchiser** (donner un jugement qualitatif suivant importance, la représentativité)
  - **architecturer** (donner des catégories logico-linguistiques)
    - (extension en fonction des objets)
    - document
    - division
    - liste
    - paragraphe
    - unité de texte
  - **contextualiser** (créer un passage hors duquel l'annotation peut ne plus avoir le sens voulu : ex. Souligner un mot, et puis donner le passage)

16

## Objectifs de l'annotation

(taxonomisation de Veron 97)

- Compléter
  - **Reformuler** (modifier le contenu du texte)
    - (extension dépend du rapport taille reformulation / taille initiale)
    - développer
    - reformuler
    - résumer
  - **commenter**
  - **documenter** (attacher un objet quelconque)
- **Corréler** (deux passages)
- **Programmer** (de future actions, par exemple : traduire, relire)

17

## L'« école toulousaine »

- Formes / moyens de l'annotation
- Objectifs de l'annotation
- Liens entre objectifs et moyens
  - ambiguïté de la relation, multivocité
  - mais certaines relations sont fortement improbables
    - difficile de reformuler en soulignant
- Remarque
  - l'annotation est rarement un but en soi

18

# Travaux pratiques

Catégoriser ces exemples d'annotation



# Retour sur la lecture active

- Remarque : la lecture scientifique professionnelle
  - parangon qui en vaut un autre, pensé pour soi-même...
  - largement fondée sur une pratique annotative
- Christian Jacob (La lecture assistée par ordinateur)
  - « l'idée essentielle me semble être que l'on déconstruit le texte comme réalité physique associée au livre, au profit d'un modèle topographique dynamique, d'un ensemble de lieux qui peuvent être reliés soit par un parcours linéaire bidirectionnel (forward/backward) soit sous une forme plus cartographique avec des figures arborescentes voire labyrinthiques, de manière à suivre toutes les connexions qui se forment au cours d'un travail d'interprétation ou au cours d'une problématique.
  - Ce double processus de balisage et d'établissement de liens correspond à une étape essentielle du travail intellectuel et permet une mobilisation immédiate, réitérable et visuelle de l'ensemble des intuitions que l'on peut avoir au cours de la lecture linéaire de l'œuvre. »

# Numérique : approche MAPS

[Virbel 94]

Poste de lecture active

- Marquage
  - structurer le texte de l'auteur
- Annotation
  - annotation
- Prospection
  - rechercher des éléments du texte, analyse plus ou moins automatique
- Structuration
  - structurer du texte provenant de différents documents en une seule entité

# Passage au numérique...

- Mimer ce qu'il se passe sur le papier
  - annotation en même temps qu'on lit pour aider la lecture
    - architecturer
    - résumer
    - préparer un traitement futur
- Automatiser des tâches à partir des annotations comme matériel numérique exploitable pour des outils sophistiqués
  - résumé : représentation concise
  - relecture assistée
  - recherche d'information
  - classification / comparaison de documents
  - conversations
  - personnalisation
  - ...

# Formes numériques possibles

- Mise en forme
- Postits
- Threads de discussion : conversation
- Liens
- Tours guidés
- Labels (index)
- Notes
- Marques d'approbation
- ... tous objets (ex. image)

# Questions liées au numérique

[Veron 97]

- Stockage
  - qui, où, comment
- Filtrage, sécurité et droits de lecture/écriture/modification des annotations
- Utilisabilité
  - liberté d'utilisation
- Uniformité
  - tout in situ, de la même forme : visualisation / annotation
  - extensibilité ?
- Robustesse et cohérence
- Acceptabilité par l'utilisateur
- Réseaux :
  - échelle et performance
  - protocoles, formats, plateformes

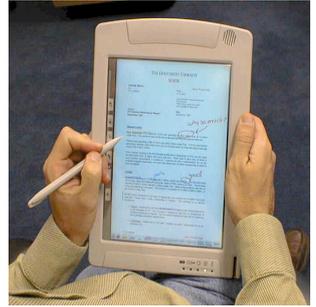
## Exemples : bureautique

- Mise en forme
  - Formats de caractères, paragraphes, ...
- Word
  - [ Annotations ]
  - [ Discussions ]
  - Notes, etc.
- Powerpoint
  - [ Annotation / commentaires ]
  - Briefing associate : dessiner des schémas et décrire les connaissances du document – cf. Web sémantique (<http://www.semanticweb.org/SWWS/program/full/paper54.pdf>)
- Excel
- Méta-données et propriétés de documents
- ...

25

## Exemples : e-book

- Tous les outils du type e-book proposent des fonctionnalités d'annotation (acrobat, etc.)
- X-libris
  - celui qui est allé le plus loin
  - tablette, annotation au stylet <http://www.fxpal.com/xlibris/index.htm>



26

## Xlibris : annotation libre

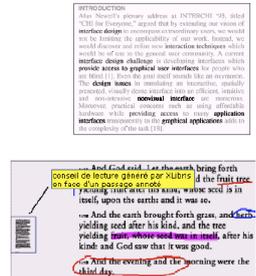
- Stylo : écriture
- Stabylo (surlignage)
- Couleurs



27

## Xlibris : visualisation

- Skimming (écrémage)
  - niveaux de gris
  - lecture allégée
  - annotations « anonymes »
- Mise en évidence des mots clés par TF/IDF
  - mise en gras
- Calcul
  - suggestions de lecture après recherche sur un moteur (Google), thumbnail
  - [ Lectures complémentaires ]



28

## Xlibris : visualisation



Parcours des annotations : représentation concise



Edition de documents : découpage des morceaux annotés, assemblage, lien avec Word

29

## Au fait, ça marche ?

- Systèmes généralement peu utilisés
- Xlibris
  - prototype de recherche
- Office – Word
  - utilisé pour les groupes de travail
  - commentaires de correction
  - discussion sur les documents
  - utilisation YP
  - lecture / écriture complexe
- Autres exemples
  - annotations sur le web : à suivre

30

## Annotation dynamique

- Actes annotatifs
  - lecture active : écriture
  - identification d'un ajustement à un conflit
  - analyse cf. actes de langage (cf. Searle)
  - considérer des suites d'actes, chercher à les qualifier, les reconnaître, etc.
- Annotation = (engagement annotatif, contexte annotatif)
  - Engagement : acte annotatif, moyen annotatif, but annotatif, raison annotative
  - Contexte : auteur, lecteur, support documentaire, campagne annotative, rôle annotateur, date annotation, session annotative
  - etc.
- Système de gestion de documents orienté lecture (développement se poursuit actuellement)
  - mémorisation des différentes activités de lecture, consultation, tri, etc. en lien avec des sessions de lecture, actes
  - en même temps, forme d'opportunisme, telle que chaque idée instantanée ou intuition soit immédiatement enregistrée

31

## Annotation, navigation, édition électronique : vers une géographie de la connaissance

- Lecture savante = lecture professionnelle
  - Ex : lecture approfondie d'une œuvre, lecture d'une œuvre dans le contexte d'une autre
  - Retour à une situation d'avant le bon à tirer ?
- Deux grandes classes d'interventions sur le texte
  - Hiérarchisation
  - Qualification
- Annotation permanente, inscription, mais mémoire finie
  - L'effet miette de pain du Petit Poucet
  - La machine est là pour aider, organiser l'appropriation des inscriptions, l'accès à la lecture passée

32

## Annotation, navigation, édition électronique : vers une géographie de la connaissance

- Approche d'une lecture active et des descriptions formalisées des structures (DFS) en trois niveaux.
  - Niveau du fonds
    - Ex. structure de la bibliothèque
  - Niveau de la grille de lecture
    - Découpage / annotation / grille
  - Niveau de l'utilisation
    - synthèse, navigation, etc.
    - « niveau des historiques de la lecture effective, constituée en objet hypertextuel qui est la grille de lecture dynamique. »
- D'où 4 grandes catégories d'opérations, qui sont les fonctionnalités de base de l'annotation électronique.
  - Hiérarchisation
  - Qualification (annotation / commentaire)
  - Navigation et recherche
  - Représentation (assistance à la navigation et à la recherche).
- Besoins de standard normatifs suivant les activités, mais qui permettent l'émergence de singularités.

33

## En conclusion

- L'annotation
  - peut être considérée à de nombreux niveaux
  - est rarement une fin en soi
  - ressort de l'écriture
  - s'intègre dans une activité productrice de sens, et souvent de documents
- Actuellement
  - Des systèmes qui imitent la pratique papier, l'enrichissent
  - Des tentatives de constituer une lecture véritablement active → hypertexte à l'ancienne

34

## En 2006, sur le Web

- Vers la bidirectionnalité des liens ?
  - blogs
  - extensions navigateurs
- La reconnaissance des documents en construction
  - documents pour l'action (Tech-CICO, Troyes)
- Vers un changement de modèle ?
  - du document d'auteur au document collectif

35

## Remerciements - emprunts

- E. Desmontil / C. Jacquin (IRIN – Nantes)
- L. Denoue (FXPal)

36

## Aux origines de l'hypertexte

- V. Bush, As we may think, The Atlantic Monthly, 1945.
- Principe général
  - Photos, facsimilés, ... : milliards de documents.
  - Traitements : calculs par des machines très rapides, simple sélection (pensée non créative).
  - Les progrès de l'informatique et du calcul ne pourront pas grand-chose pour la Science : l'intuition et la créativité ne sont pas dans la machine.

37

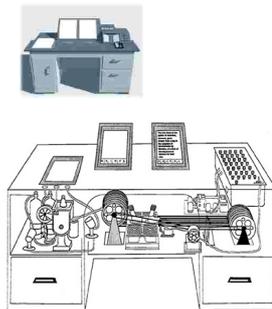
## Bush dans le texte

- ...Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to have rules as to which path will locate it, and the rules are cumbersome. **Having found one item, moreover, one has to emerge from the system and re-enter on a new path...**
- The human mind does not work that way. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts [...] the speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature.
- Man cannot hope fully to duplicate this mental process artificially, but he certainly ought to be able to learn from it. [...] The first idea, however, to be drawn from the analogy concerns selection. **Selection by association, rather than by indexing, may yet be mechanized.**

38

## Memex

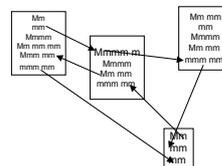
- Un bureau à microfilms : livres, communications, ...
- Lecture en utilisant l'indexation et la sélection : livre, page.
- Possibilité de voir plusieurs pages en même temps, de les rappeler, etc.
- Possibilité d'ajouter des pages (scanner)



39

## Memex

- Annotation dans les marges
- Construction de trails (chaines de liens)



Theodor H. Nelson, "As We Will Think," Proceedings of Online 72 Conference, England, 1972.



40

## Bush dans le texte

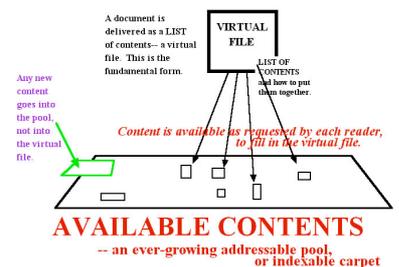
- All this is conventional [...] It affords an immediate step, however, to **associative indexing**, the basic idea of which is a provision whereby any item may be caused at will to select immediately and automatically another. This is the essential feature of the memex. **The process of tying two items together is the important thing.**
- When the user is building a **trail**, he names it, inserts the name in his code book, and taps it out on his keyboard [...] the items are permanently joined. [...]
- Thereafter, at any time, when one of these items is in view, the other can be instantly recalled merely by tapping a button below the corresponding code space. Moreover, when numerous items have been thus joined together to form a trail, they can be reviewed in turn, rapidly or slowly, by deflecting a lever like that used for turning the pages of a book. It is exactly as though the physical items had been gathered together to form a new book. It is more than this, for any item can be joined into numerous trails.

41

## Xanadu

- Modèle hypertexte depuis les années 60
- Modèle simple + gestion des droits

### AN AUTHOR-BASED, LITERARY AND CULTURAL DESIGN The Xanadu Document Model — built on the assumption of perpetual change and re-use



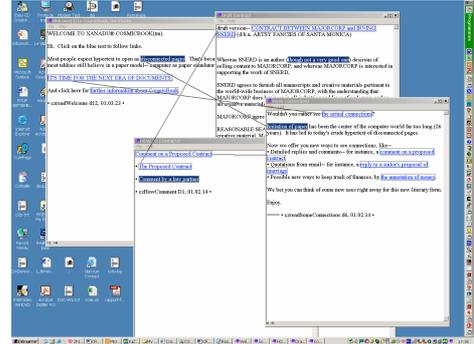
## Ted Nelson: rebel with a cause

Today's one-way hypertext-- the World Wide Web-- is far too shallow. The Xanadu project foresaw world-wide hypertext decades ago, and endeavored to create a much deeper system. The Web, however, took over with a very shallow structure. Our simple method, very different, allows--

- UNBREAKABLE LINKS.
- COPYRIGHT SIMPLIFICATION AND SOFTENING: by special permissions and methods, quotations of any size may be used by anyone and mixed together frictionlessly.
- ORIGIN CONNECTION: All quotations and excerpts stay connected to their original.
- TWO-WAY LINKS: anyone may publish connected comments to any page.
- SIDE-BY-SIDE INTERCOMPARISON OF CONNECTED DOCUMENTS: showing two-way links, differences between versions, origins of contexts. (For a simple working demonstration, see our new free [CosmicBook\(tm\)](#) reader.)
- DEEP VERSION MANAGEMENT: documents may be changed incrementally (with each version available); versions may branch; authors may easily see exact differences between versions.
- INCREMENTAL PUBLISHING: new changes may be continually made by authors without breaking links. How can this be? Very simple, but very different.

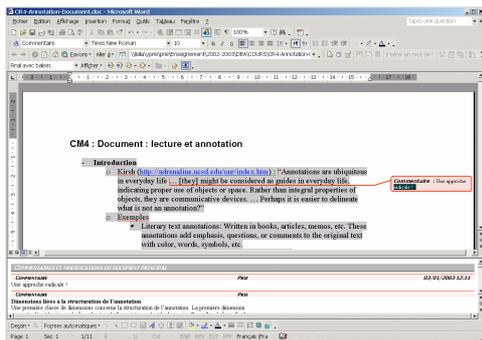
<http://xanadu.com/xuTheModel/>

## CosmicBook



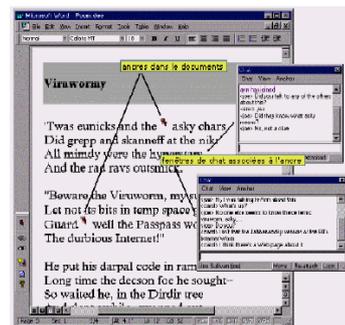
[Retour]

## Microsoft Word : annotations



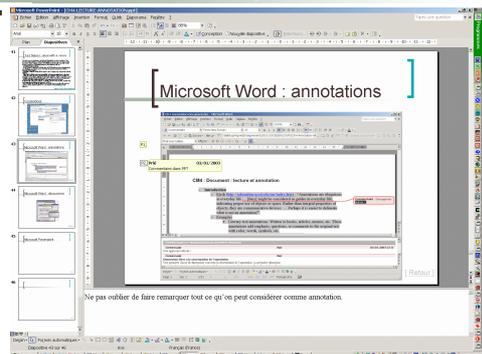
[Retour]

## Microsoft Word : discussions



[Retour]

## Microsoft Powerpoint



[Retour]

## Xlibris : conseil de lecture

**Further reading list for Beyond Paper: Supporting Active Reading with Free-Form Digital Ink Annotations**  
Xerox 10K 1994

**Digital Publishing**

The FaculTech family of digital publishers scans hard copy and converts it to digital documents, or accepts digital documents directly from networked personal computers or workstations. These friendly, electronic publishing packages also allow the manipulation of images in the creation of new documents.

**Personal Document Reader Research Roadmap - 1st Edition**

Digital copiers may be considered as a hardware alternative to pen computers. Ink on paper user interfaces (e.g., Protolite (Racal) and Xerox DAF (Xerox)) is analogous to digital ink, and a paper display may be updated by running it through a digital copier. The advantages of digital copiers over pen computers are two-fold: paper is better than any electronic display, and it is

**Dynomite - A Dynamically Organized Ink and Audio Notebook**

- support retrieval beyond the visual scanning offered by paper notebooks.

**DYNOMITE DESIGN**

Based on our notebook usage study, we designed and implemented the Dynomite system to augment and complement the Dynomite system to augment and

**A Comparison of Reading Paper and On-Line Documents**

**Paper.** Movement through paper documents was characterized by its speed and automaticity. For example, page turning in the paper condition was often anticipatory, with one hand often lifting a page even before it was read, minimizing disruption between reading the text at the end of one page and the beginning of another. That it was so

**ink,** analogous to data types in a programming language. Properties are associated with ink and use the entire page and also so that all the ink with a particular property can be retrieved and shown at once. This use of properties for retrieval is quite different than "color" and "thickness" ink

example, one subject tried to get to a particular location in the document using the drag function on the scroll bar, but the document contents did not move concurrently in response to the dragging. As a result, the subject had to release the page icon and then wait to find out if she was in the right place. This kind of scanning can make finding exact locations extremely time-consuming.

[Retour]