

Documentarisation processes in Documents for Action

Manuel Zacklad

Université de Technologie de Troyes
Tech-CICO Laboratory (UTT / CNRS)

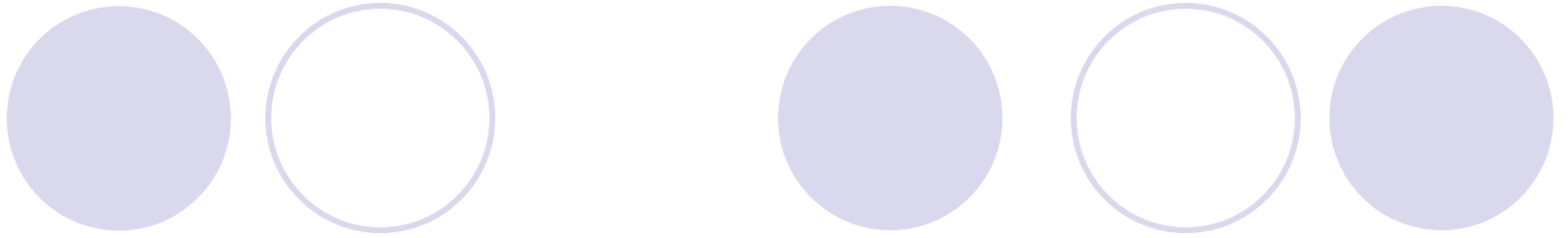
New document-related collective practices

- The development of the Web either in the internet or intranet context has resulted in a multiplicity of document-related collective practices
- From making documents available to a team in “small digital libraries” to the use of annotations while collectively writing or using tools such as Wiki and Blog.

Documents for cooperation



- We focus on situations where documents are used to mediate the coordination of a distributed group committed to working towards a common goal.
- The document can be viewed as a set of fragments contributed by various authors whose final content remains indeterminate.
- Its fast dissemination makes it a useful tool for conveying information, assisting decision-making and probing situations.



A pragmatic communicational definition of documents and semiotic product concepts

Approach



- Need for a broader approach than that usually adopted in studies on documentary topics in line with that of Briet 1951, Brown & Duguid 1995, Buckland 1997, Pédauque 2003, for example
- We continue on these pragmatic theoretical lines by investigating **symbolic communicational transactions** using a framework that we previously designed to account for cooperative activities

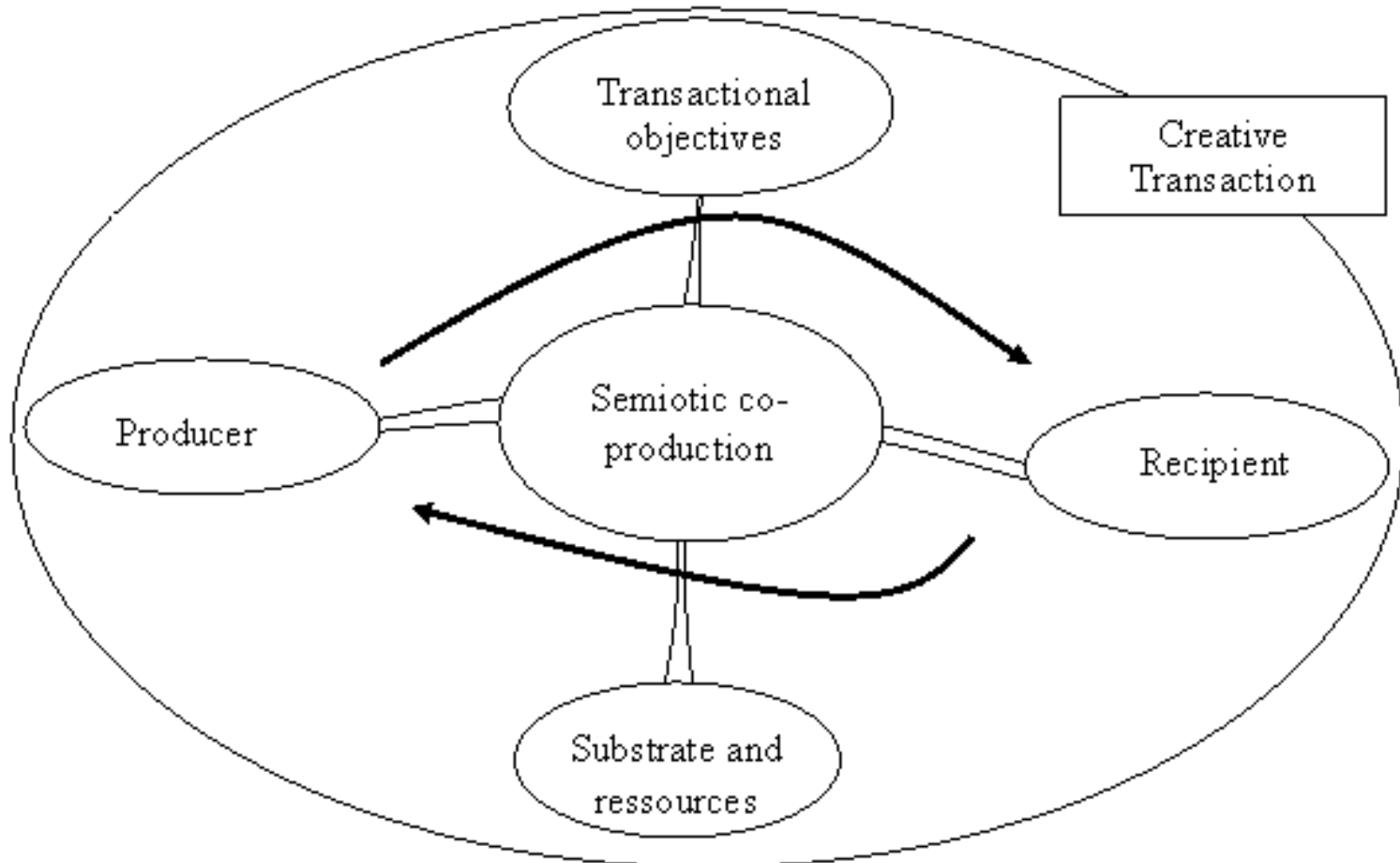
Hyperwriting, hyperfragmenting

- The fact that the public production phase of the document continues during large time scale is as important to us as the fact of including tangible objects in the definition of documents (Briet 1951).
- There is an hyper-writing phase, when digitized “Documents for Action” are drawn up out of many *evolving semiotic fragments* by the very collective activities they were intended to mediate.

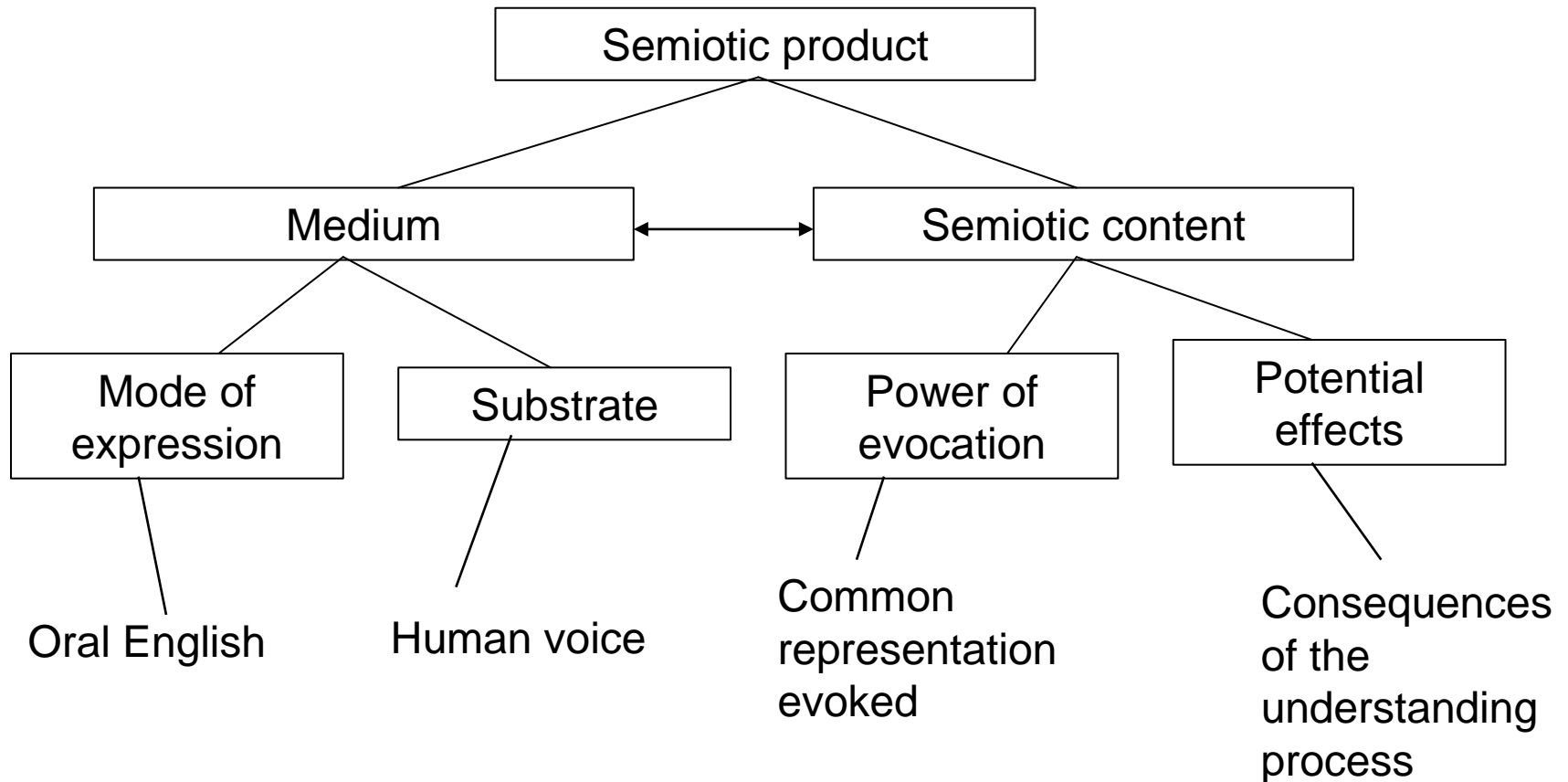
Symbolic communicational transactions

- To analyze what a "document" consists of it is necessary to take into account the long-term asynchronous production process, between the producers and beneficiaries bound together by common interests.
- In terms of the transactional theory of action, a document is the object of transactions between partners who commit themselves personally to participating and to providing the knowledge required to obtain an at least partly jointly produced "work"

Semiotic Transaction



Components of semiotic co-production



Distribution of transactions in a heterogeneous spatio-socio-temporal framework

- The producers and recipients of communicational transactions do not all belong to the same spatio-temporal framework.
- The semiotic production must have a sufficiently long-lasting form, so that they can be initiated, interrupted, updated and repeated in all the possible configurations involving the presence or absence of the producers and the recipients.

Eight strategies for coordinating distributed communicational transactions

- (1) *standardizing the transaction situation,*
- (2) *formalizing the mode of expression,*
- (3) *mnemotechnic ritualisation*
- (4) *encouraging abstraction,*
- (5) *substitutive mediation,*
- (6) *documentarisation,*
- (7) *increased recourse to techno-informational equipment,*
- (8) *substitutive coordination.*

Substitutive mediation



- *Substitutive mediation* focuses on the *material substrate* of medium on which the semiotic production process is based.
- Direct perception of the semiotic production process by the subject's own body is replaced here by an indirect mode of perception involving:
 - the *transcription onto a writing substrate*
 - or the *automatic encoding of one of the physical sources* conveying the signals forming the semiotic product onto a different substrate.

Ephemeral & perennial substrate

- *Ephemeral substrates* are characterized by the fact that the stimuli they convey can have transient effects on the recipients' sensory organs
 - the air, which is the ephemeral substrate conveying speech
- *Perennial substrate*, on the contrary, keep the form they were initially given, which makes it possible for the recipients to repeat the effects elicited by this form on their sensory organs
 - re-reading a few lines or re-viewing part of a film on personal digital equipment

Three substitutive mediation techniques

- 1) **Transcription** involves writing, and it is based on the use of a graphic code which has to be mastered by both the writers and the readers.
- 2) **Automatic encoding of a physical source** serves to convert a physical form (sound or light) in order to transfer it more easily onto an other substrate.
- 3) **Recording** makes it possible to save the physical signals encoded onto a perennial substrate

Documentarisation

- **Documentarisation** is an extended version of substitutive mediation, which occurs when the substrates used are perennial.
- It consists of endowing the substrates with *specific attributes making it possible:*
 - *(i) to manage them along with other substrates,*
 - *(ii) to handle them physically, which is a prerequisite to be able to browse semantically among the semiotic content, and lastly,*
 - *(iii) to guide not only the recipients, but also the producers themselves, around the substrate by providing one or several maps of the semiotic contents.*



Processes and means of documentarisation

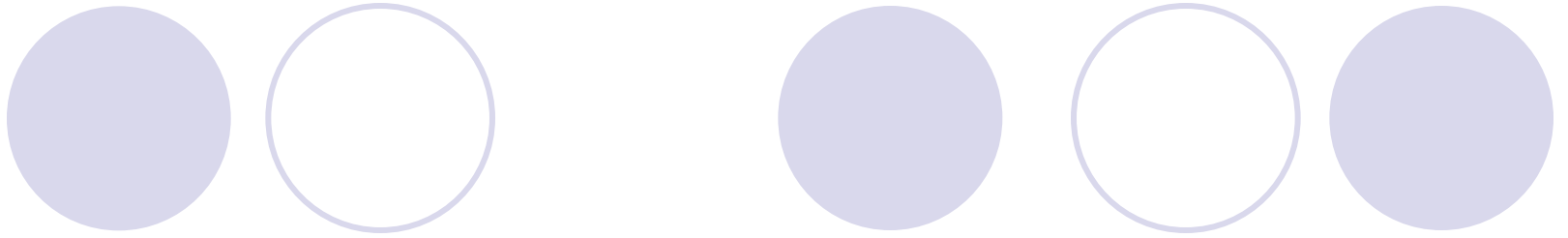
Substitutive mediation vs documentarization



- Transcription and recording procedures (that is substitutive mediation) are prerequisites for successful documentarisation processes, but they do not suffice:
 - Sort term personal note-taking or an occasional recording intended to facilitate some semiotic activity do not constitute document if the substrate is not worked out
- If the *documentary investment* has not been made, it will not be possible to go back to re-use these expedients.

Document

- In line with our definition of documentarization document will be taken to mean *a semiotic product transcribed or recorded on a perennial substrate, which is endowed with specific attributes intended to facilitate the practices associated with its subsequent uses in the framework of distributed communicational transactions.*
- These attributes make it possible for the document to move through time and space among the communities of interpretation, with a view to prolonging and extending the communicational transactions initiated by its producers.



Implicit and explicit semantic links between fragments of a document

Two problems in managing documents

- External management of the documents:
to store and search them among others in *libraries, archives, filing cabinets, administrative, technical and documentary databases of various kinds*
- Internal management of the documents:
document seen as forming a set of micro-semiotic products which must be organized and searched



Internal management: fragments

- The use of a perennial substrate make it possible to inscribe or record an enormous number of signs, corresponding to micro-communicational transactions, on the same physical substrate.
- A transcribed or recorded micro-transaction is a fragment.
- There is a need to provide readers with means of semantic orientation within a document by linking together the fragments.

Implicit links

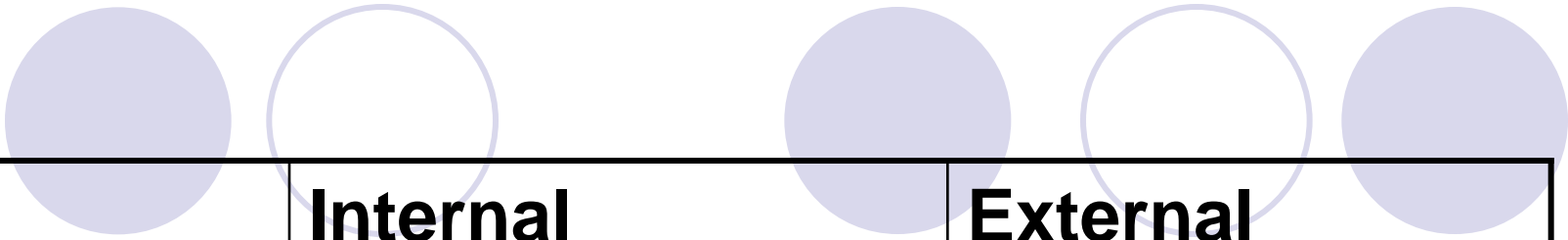


- Implicit links serve to connect together fragments of semiotic productions using various narrative strategies of temporal, structural or functional kind, for instance.
- They make use of all the implicit relationships authorized in the sharing of a common representational ground, which is developed and enlarged as the transaction progresses.
- This linkage is carried out in the way which seems the most natural, while apparently giving the more direct access to the semiotic content.

Explicit links



- The explicit system of orientation parallels and complements the implicit one by:
 - explicitly presenting the initial browsing plan in the form of sub-titles or Tables of Contents
 - or by suggesting other more unexpected browsing projects: indexes or tables of references which are organized differently from “Tables of Contents” (keywords, names of quoted authors).
- Explicit linkage may require ontological resources of a formal or semiotic kind making it possible for users to browse efficiently through the semiotic contents.



	Internal management	External management
Explicit	Chapter numbers, Table of contents, indexes	The numbering of the volumes of a collection
Implicit	Implicit logical access to content following the writing of the page according to a narrative strategy	A succession of paper sheets forming a pile



***Emergent communicational transactions in
fragmented documents: Documents for Action and
annotations***

Smooth and fragmented documents



- A document, in the form of a single macro-transactional project summed up by its title, will not achieve it until a possibly very large number of arguments have been put forward (micro-transactions)
- In the case of a technical or legal document, the length is often justified by the need to put forward detailed arguments in response to the potential objections liable to be raised by the future recipients.

Smooth documents



- The “smoother” or more unified a document seems to be, the more implicit the various arguments corresponding to micro-communicational transactions will be.
- Each literary approach corresponding to a genre favor a specific implicit type of semantic link between the fragments.

Fragmented documents: DofA

- In fragmented documents the various micro-communicational transactions have been explicitly linked up using :
 - numbering, sub-titles, or a set of specific attributes announcing the status of the fragments in the document and their relationships with the surrounding fragments.
- Strong cases of fragmentation can occur in a special class of documents coordinating distributed communicational transactions: “**documents for action**”.



Cooperative semiotic production processes: documents and files for action

Increasing use of documentarisation

- In an increasing number of cooperative professional activities, the intermediate or final product of the cooperative process is a semiotic product
- The use of documentarisation strategies prompted by the extended use of techno-informational equipment is becoming extremely popular:
 - for example, word processors which can be used to make annotations as well as being well integrated into electronic mail systems.

Example of DofA



- A few examples of DofA:
 - engineering design documents (mechanical descriptions, software programs),
 - contractual documents in business, which are gradually transformed from business proposals into definite, formal, contracts
 - patients' case-records in the field of health care,
 - quality insurance documents
 - management counseling assessments,
 - free open source software forum

Main characteristics of DofA

- Extended state of incompleteness: they go through a long process of completion during the active collective semiotic production phase, during which we call them *evolving DofA* (as opposed to *stabilised DofA*).
- Perenniality: due to the participants' commitment to the semiotic content of these documents.
- Fragmentation: at least during the evolving phase, they contain several fragments which are often only loosely semantically linked and which cannot be implicitly integrated into the document.

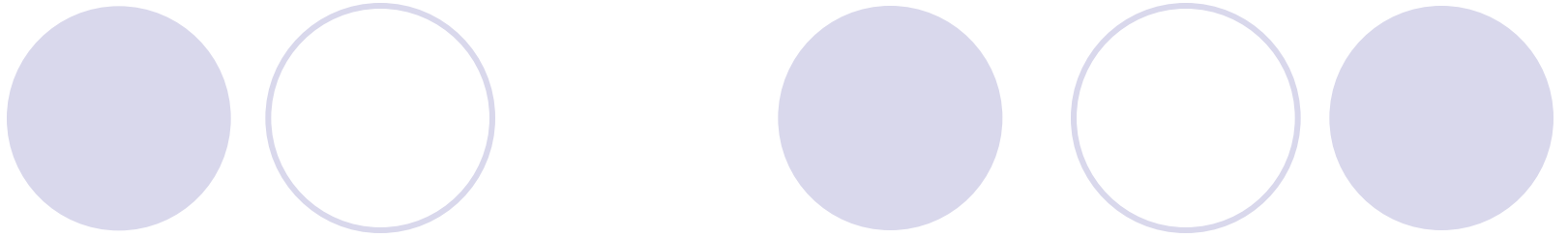
Main characteristics of DofA (cont.)

- Non trivial relationships between DofA fragments and their producers: the various fragments of DofA are often produced by different authors who have different statuses and rights to the semiotic product;
- Non trivial argumentative relationships between the document fragments: each fragment stands in a potentially complex relationship with the others (contradictory statements for example).

Various degree of structuration of DofAs

- Questionnaires/forms
- Standard plans (systemic models used in engineering, or the standard outlines used to draw up legal contracts).
- Stereotyped style of writing, such as that used in the scientific texts.
- Question-and-answer style used in internet newsgroup
- Narrative documents, in which one wishes to mask the underlying plan.

- *For example a newsgroup makes use of explicit links based on the thread of discussion structure, and seems to be less completely integrated than a technical document.*



The role of annotations in DofA

Emergent communicational transactions in cooperative activities



- To understand more clearly DofA it is necessary to analyze the dynamic of the transactions to which they give rise.
- In face-to-face cooperative activities the participants can suggest solutions to problems and put forward arguments in the framework of polylogal interactions.
- When transactions are mediated by DofA, it will be necessary to set up appropriate mechanisms to facilitate these emergent transactions.
- Annotations are one of the main mechanisms used for this purpose

Documentarising fragments



- If free fragments are not properly articulated together as soon as they are inscribed on the substrate, the DofA will not be able to efficiently sustain the emergent distributed transactions involved in the cooperative activity.
- A free fragment will constitute an annotation only if it has undergone a process of documentarisation explicitly stating the nature of the link between each fragment and the main semiotic product

Annotation



- We define an *annotation* in the strong sense of the term as a *documentarized fragment of a semiotic product, i.e., one which is endowed with specific attributes with which it can be explicitly linked up with the other components of the document.*
 - Naming of the author, specifying the date, the place, the semantic relation with the main semiotic product, etc.
- The work of annotation includes a specific efforts making it possible to re-use fragments for either individual or distributed collective purposes

The codification of documentary investment in cooperation technologies

- The semi-automatic editing and searching of annotations require *codifying* the attributes serving :
 - as means of external articulation between the fragments and the DofA (anchoring or numbering, for example),
 - as a mean of internal articulation with parts of the semiotic content, by replacing tacit links by more explicit ones (for example, using meta-data to typify the nature of the annotation).

Codification data bases for annotation

- This process of codification requires a set of predefined resources (a codification database) based on:
 - official administrative data (such as the participants' names),
 - data automatically generated by a “function” in the mathematical sense of the word (as often occurs with the date or serial numbering).
 - ad hoc modeling efforts providing list, thesaurus, ontology, etc.,

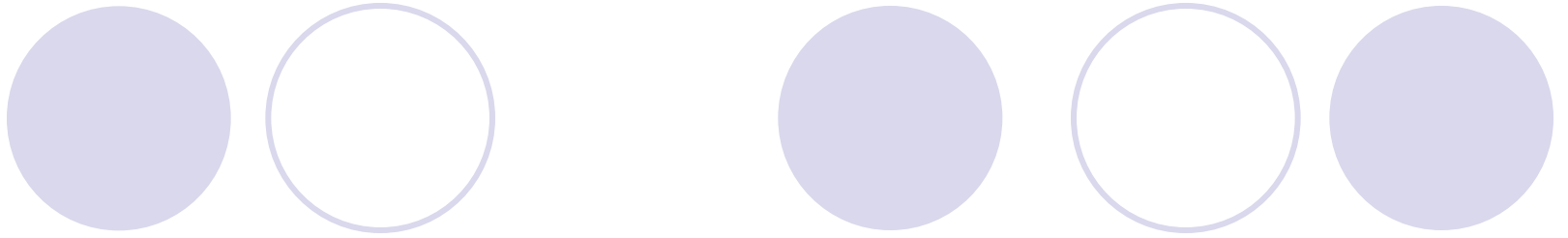
Cooperation technologies



- In some cases, the cooperation technologies will make use of this codification by partly automating some aspects of the work involved in codifying (documentarising) fragments thus decreasing the documentary investment required on the part of the users.
- For example systems of annotation available in word processor programs make it possible to automatically codify annotations by providing the name of the author, the date, anchoring data, a system of numbering, etc.

Newsgroups

- Some other cooperation technologies are not necessarily always perceived as annotative documentary technologies : electronic mail systems and newsgroups.
- In these technologies semantic articulation of the annotative fragments occurs via a process of semi-automatic codification:
 - specifying the author's name, the date, repeating the title of the original contribution in the response, hierarchically ranking the contributions in the thread of the discussion, etc.



Conclusion: new socio-technical approaches for the management of codification databases, the socio-semantic web

Indexing DofAs



- The special characteristics of DofA means that they require new, more appropriate principles of indexing and classification
- This would facilitate the “operational” information management activities, as well as contributing to the long-term “knowledge management” issues by making it possible to archive the various versions of DofA more efficiently.

Socio-Semantic Web

- To design codification databases the Tech-CICO laboratory has started a **Socio-Semantic Web (W2S)** research project.
- It is complementary to Semantic Web which focus mainly on defining universal standardized “formal ontologies”.
- Socio-Semantic Web aims to provide highly flexible codification databases which meet the changing requirements of local communities of users
- It is based on *semiotic ontologies* which set up an *interpretative* relationships between both the semiotic concepts and the situations for which they stand (HyperTopic standard)