Principles of Classification for Writing Hyperdocuments

Maria Luiza de Almeida Campos
Hagar Espanha Gomez

ABSTRACT
Hypertext has introduced a new technology in document structure that makes non-linear reading possible, but is not involved with semantic content. Principles for organization of subject content were established in the fields of Faceted Classification and of Concept Theory. They constitute sound theoretical and methodological bases for planning and writing hyperdocuments for educational purposes. Faceted Classification method consists, on the one hand, in identifying general or facet concept classes within which are specific mutually exclusive classes and, on the other, in principles for their organization and relationships. Concept Theory provides orientation for the organization of node content. These are the main points for the organization of documents/information. Such theories provide conditions for coherent nodes and contribute to avoide cognitive overload.

General Considerations

Gutenberg's invention made possible greater distribution of texts, thus providing greater access to sources of knowledge records. However, while causing many transformations relating to access, it did not bring any significant modification in books which refer to the support of information. According to Roger Chartier, the printed book depends greatly on primitive forms, as it imitates pages, writing, appearances, in the same way that previously

book pages were sewn. From this point of view, his invention does not produce 'the book'. (CHARTIER, 1997, p. 134).

Nowadays, an even greater revolution than that of Gutenberg is taking place and it is altering not only techniques of text reproduction, but also structures and forms of support that enable a particular form of communication to readers with electronic documents.

The technology known as Hypertext has brought about alterations in writing, with consequences for the production of educational documents. It makes possible the building of hyperdocuments, a product that requires new forms of writing and affects the process of traditional linear reading.

The production of non-linear texts using hypertext technology also requires a complex structure named hyperstructure, consisting of graphs pointing to nodes which are chunks of information, and of arches which link these chunks (CONKLIN, 1987, p. 31). The main feature is its ability to link content concepts within a document or within several ones. In essence, hypertext is a

> computer based medium for thinking and communication. The thinking process does not build new ideas one at a time, starting with nothing and turning out each idea as finished pearl. Thinking seems rather to proceed on several fronts at once, developing and rejecting ideas at different levels and on different points in parallel, each idea depending on and contributing to the others. (CONKLIN, 1987, p. 32)

Hypertext materializes and emphasizes the associative and non-linear reading process and this requires a rethink of the processes related to the arranging of ideas and how they affect modes of writing, i.e., new processes for new forms of writing. In this context, we will investigate a type of text named hyperdocument.

THE HYPERDOCUMENT

Hypertext technology brings together processes of both writing and reading in non-linear form. When writing a text on a sheet of paper, the author follows a linear process where logical chains of reasoning must have a beginning, a middle and an ending. Attempts to

adapt this to the reading process cover thousands of years, but this process is naturally non-linear. How often does a reader need to search other documents to help, for instance, with the definition of a term, with the historical context of the topic presented in a document, with the identification of the school of thought of the author?

With the technology of hypertext the author presents readers with many possibilities of reading directions. In hyperdocuments these possibilities evidence the conceptual dimension employed in their construction. Every possible 'navigation' is conditioned to this conceptual dimension because of its nature as a closed hyperdocument developed within a field of knowledge, with explicit linking between nodes.

In addition to all this, technological possibilities have also brought about, in this specific case, discussions related to the construction of hyperdocuments and the search for the solution of problems inherent to this new object; they are cognitive overload and disorientation.

Cognitive overload may occur when authoring a hyperdocument as well as when reading it. For the author, it can be seen as a mental overload, because he/she needs to name nodes and to define semantic relations among them. For the reader, this overload occurs as a result of the need to make frequent choices and different ways to track. Disorientation occurs when reader feels lost, i.e. he does not know where he is in the network, or does not know how to get to some other place that he knows or thinks that exists in the network, losing track of the route he followed (CONKLIN, 1987, p. 38).

The readability of a document means the mental effort required for the process of its construction. So, if we wish to increase the readability of a hyperdocument, we must help readers in the construction of their mental models, strengthening those factors that support this process and weakening those that hinder it. This is done by strengthening coherence and hindering cognitive overload (THURING et al, 1995).

To assure coherence at node level one needs to rethink the processes of writing presented in the structure of the text within the conceptual node. As hypertext is a new textual means of information, one needs to think about a new rhetoric and style when producing a

text. Some authors are developing research in this direction (LANDOW, 1987; MOULTHROP, 1992), but it is important to consider writing issues at the moment when the conceptual node is being written.

What is being done to a great extent is the conversion of texts produced for linear reading using hypertext. Generally, this is not the best option to warrant coherence at the node level (NILSEN, 1993).

Interpretative action is an issue to be thought of in the role of both author and reader. In this respect, linguistic phenomena[1], according to French authors on discourse analysis, may facilitate or not the meanings in textual construction.

A text is a unit of meaning which in discourse analysis is seen as the materiality of the speech. One needs to think of the conditions of text production, in this case, the text as a hyperdocument.

Hypertextual writing opens up a new authoring process. When writing a hypertext the author must master both the subject which he/she wishes to write on and this new intellectual technology. A hyperdocument possesses a new speech materiality which does not use pencil and paper and linear thinking in its form of writing, but explores technological possibilities of man-machine interface as well as possibilities of the very speech construction and order within a knowledge domain.

Authors should limit the fragmentation of text, which characterizes hyperdocuments, in order to achieve coherence between nodes. This aspect seems to be endemic in hyperdocuments and as result one has information segmented in disjointed nodes presented in separate windows. This fragmentation may lead to lack of significant context, so giving the impression that a hyperdocument is an aggregate of pieces of information loosely joined together instead of a coherent whole (THURING et al.).

---

1 Linguistic phenomena such as polyphony, presupposition, negation, direct and indirect speech, words within inverted commas, metaspeech, paraphrases, irony, imitation, pastiche. (Maingueneau, 1998)

A means of reducing this impression is the explicit representation of relationships between nodes. Besides that, it is fundamentally important that the author has consistent guidelines when planning the text and the browsing/reading possibilities.

Planning of Hyperdocuments for Long Distance Teaching

The planning of hyperdocuments aims to establish a text structure to create mutually exclusive nodes. This is a measure that may avoid disorientation. Therefore, it must be conceived within a systemic approach which naturally provides *a priori* indications for some kinds of relationship useful to browsing.

The planning may also consider the different levels of the information needs of readers whom the educational hyperdocument is devised for, thus avoiding the problem of cognitive overload, i.e., giving the reader the possibility of easily finding the information on the desired level, being it only an indication or detailed information on the concept[2] he/she is searching for.

The structure of the hyperdocument consists, then, of establishing conceptual units/nodes and relationships aiming at consistency and coherence in the document as whole. At a certain point, hyperdocuments can be compared to a system of concepts. An underlying classificatory action exists, as relationships between nodes are created from a network of associations.

In hyperdocuments, a node is characteristically a knowledge unit, a concept that corresponds to the smallest unit of information in the subject field considered. This is one of the elements that allow the reader to track his/her way through the text. One of the first issues relating to the nature of hyperdocument node contents is the use of a verbal tag appropriate for a given subject content. If not appropriate

2 Concept is the term used here, as each node is a concept (Rumbaugh et al. 1994)

it can lead the reader to cognitive overload and it may damage his/her understanding of the text.[3]

The theory of Faceted Classification provides basic principles for mutually exclusive chunks of texts and for logical relations. One of these principles, that of exclusiveness, means that one must be sure that all the characteristics of a class of concepts be analyzed. In other words, a given concept should belong to that class and to no others. A node must be built in this way, as a cohesive whole, so that it may have different further uses.

As each node is a conceptual unit, one finds in the Theory of Concept the elements to guide the development of the content of the fragments of text.

In short, we may state that the planning of the hyperdocument consists of identifying general and specific classes as well as generic and semantic relationships between them, according to principles of Faceted Classification. For the development of the content of each concept/node, the Theory of Concept provides the elements.

Theoretical Principles for Writing Hyperdocuments

Faceted Classification

The Theory of Faceted Classification was developed by Shiyali Ramamrita Ranganathan in the thirties, to make clear the principles he adopted when developing the Colon classification, the scheme he devised to arrange books in the Madras' University Library, in India (RANGANATHAN, 1967) .

3 This may seem inconsistent within the field of discourse analysis, but it is necessary to make clear that the hyperdocument is being analyzed from the perspective of the author preparing a teaching text and not of the reader .The author has "to fight" to ensure that his message is understood. It is up to the reader to interpret it from his/her own viewpoint and this is not a constraint for the author as he/she pursues the principle of univocity, for he/she writes aiming at a virtual reader.

Ranganathan's theory consists of classification of ideas into Facets, which are general classes within a field of knowledge. These are, in turn, manifestations of fundamental categories, which are valid for any field of knowledge.

He identified five fundamental categories: Personality, Matter, Energy, Space and Time. They are, in fact, means to guide thinking when structuring any field of knowledge and they are an easy way to analyze its constituent concepts. This method is deductive as it provides the bases for the identification of general classes within which are chains and arrays of concepts of the field of knowledge analyzed.

In a hyperdocument, categorization is very important when building conceptual nodes. Modular writing brings together chunks of information, and then, it is necessary to select these chunks. For example, if a certain node deals with the properties of milk, the industrial process must be in another node in oder to archieve logical coherence when distributing the contents among the different nodes. This coherence must already be envisaged when preparing the hyperdocument, through the analysis of the selection of contents. As one knows, these contents result from a logical perspective of classification and their relationships within a category or among different categories.

The Classification Research Group of England has identified categories that most frequently occur and which are, in fact, a breakdown of the five fundamental categories of Ranganathan (VICKERY, 1966, p. 46-47). They were developed for the construction of thesauri but they fit equally when authoring hyperdocuments:

	(examples)
Things, Entities	
Naturally occurring	Minerals, Animals, Plants, Soils
Products	Bridges, Engines, Fiber
Mental constructs	Equations, Rectangles, Formulae
Input to the System	
Parts, components, structure	Beam, Wheel, Wing
Organs	Heart, Seed
Materials, constituents	Metal, Glass, Nitrogen
Attributes	

Qualities, properties	Cohesion, Color, Solubility
Processes, behavior	Vibration, Inflammation
Operations	
Experimental	Cutting, Breeding
Mental	Calculation, Reasoning
Operating agents – any thing or entity can act as an agent	
Place condition	

The content of the facets together their specific classes can be considered as fragments of text – the concepts – which are hierarchically related. The analysis of each fragment may reveal other kinds of relationships.

The above categories may be seen as a guideline for the analysis of a given topic of a subject field.

Theory of Concept

The Referent-Oriented Analytical Concept Theory provides a sound basis for the determination and understanding of the concept for information representation/retrieval. Dahlberg developed her theory in the seventies, aiming at building terminologies in the field of Social Sciences (DAHLBERG, 1978). In this same period, she explained the relationships between Concept Theory and Classification Theory (DAHLBERG, 1978a). Later on, she used her theory in the field of verbal documentary languages, specifically for building thesauri (DAHLBERG, 1980). In this paper those principles of the theory that are useful when authoring hyperdocuments must be emphasized.

According to Dahlberg, a concept is a "knowledge unity, comprising verifiable predications about a selected item of reference, represented in a verbal form" (DAHLBERG 1983, p. 17). Concept Theory is 'referent-oriented' because 'it presupposes that each concept refers to something and it is called 'analytical' because, by predicating the referent and thus generating and identifying its concept characteristics, a concept is constructed in an analytical mode' (id. p. 18).

Dahlberg's Concept Theory may prove useful to ensure consistency within a node and among them.

Another issue in modeling nodes is how to separate "chunks of information" so that they can be further interlinked by the reader when

navigating through the document. If in the beginning it was possible to identify themes dealt with in hyperdocuments as a whole and to define what we call the structure of the work using principles of Faceted Classification, now it is necessary to establish criteria to relate contents consistently.

Each node contains information on a general concept which represents a concept class. So, when defining the scope of a content to be dealt with in a conceptual node one must first of all explore all the elements that contribute to the understanding *of what a concept is* and what are *its constituent elements* are. In hyperdocuments, a node needs to be worked this way, as a cohesive whole, so that one may use it later in different ways. These elements are present in conceptual definition, according to Dahlberg's Theory of Concept.

Dahlberg considers three kinds of definition, namely: nominal, ostensive and conceptual. The latter one provides elements to establish links/relations among fragments of text. It may be subdivided into generic definition, one that produces hierarchical relations; partitive definition, one that produces whole-part relations; definition by negation, which is rare and produces relationship of opposition; and definition by function, which produces a functional relationship.

In writing hyperdocuments it is fundamental to understand the nature of relations. They must be governed by classificatory principles; connection among nodes without principles based on logical and ontological relationships leads to links that may be inappropriate. So, the following kinds of relations may guide authors of hyperdocuments.

Hierarchical relation – A relationship that makes possible the organization of a line of reasoning where ideas are interconnected by logical succession. This means that this relationship represents a sequence of ideas of the same nature, forming logical chains and arrays of concepts. For instance the concept 'thesaurus' and kinds of thesaurus, such as 'multilingual', 'monolingual', 'microthesaurus', among others.

Partitive relation – A relationship that provides the organization of nodes in ontological coordination. It presents the elements or parts that constitute an object or the stages of a process. Similar to the hierarchical relationship, the partitive relation establishes a certain

precedence among nodes of this kind and this is useful for documents with teaching objectives. For example, in a document on thesauri, the node that deals with planning implies several parts such as subject delimitation, target public, survey of sources of terms, forms of presentation, etc., and these constitute partitive relations.

Functional-Sintagmatic Relation – This is a relationship that establishes links between conceptual nodes of a different nature, not considered as parts or stages of a process. For example, a concept denoting a process or operation leads to concepts that supplement them, as in the following sequence: production – product – producer – consumer. The adoption of principles to identify how to relate concepts of different natures, will render hyperdocuments that will not give the impression of a mere aggregate of information 'loosely' gathered. On the contrary those relationships are built under a logical, coherent perspective.

FURTHER CONSIDERATIONS

A tutorial for thesaurus construction was developed according to the theories presented here and may be accessed at http://www.conexaorio.com/biti/tesauro. Figure 1 shows the scheme produced in the planning/classificatory process.

Browsing is the space of the reader. The objective of hypertext technology is browsing. The theories proposed here contribute to some extent to browsing that the author can offer students in a didactic sequence. But they are not the only ones.

The reader must find orientation as he/she browses, accepting or not the options offered. Even this may not suit the reader. So, other mechanisms should be offered, such as a map which is indeed the systematic plan identified when building the system of concepts, which provided the development of conceptual nodes. This plan provides the basis for an alphabetic index as well.

It must be emphasized that these two types of index –alphabetic and systematic– are relevant devices to avoid becoming lost or disoriented. The alphabetic index is useful when the reader already knows

the topic that he/she wishes to access. The systematic approach gives the reader the overall plan of the hyperdocument and provides access to topics not necessarily previously thought of by the reader.

It is believed that the theories described here should be discussed with subject specialists so that consistent and coherent content can be present in future hyperdocuments for distance learning.

REFERENCES

CONKLIN (1987) Hypertext: an introduction and survey. *IEEE Computers*, sep., p. 17-41.

DAHLBERG, I. (1978) A Referent-oriented analytical concept theory of interconcept. *International Classification*, v. 5, n. 2, p. 142-150.

DAHLBERG, I. (1978a) *Ontical structure and universal classification*. Bangalore, Sarada Ranganathan Endowment. 64 p.

DAHLBERG, I. (1983) Terminological definitions: characteristics and demands. *In: Problèmes de la définition et de la synonymie en terminologie*. Québec, GIRSTERM, p. 13-51

LANDOW, George P. (1987) Relationally encoded links and the rhetoric of hypertext. *Hypertext '87 Proceedings*, p. 331-338.

MAIGUENEAU, Dominique (1998) *Termos-Chave da Análise do Discurso*. Belo Horizonte: Editora UFMG.

MOULTHROP, Stuart (1992) Toward a rhetoric of informating texts. *The Fourth ACM Conference Hypertext*, p. 171-180.

NILSEN, J. (1993) *Hypertext &Hypermedia.* New York, Academic Press.

RANGANATHAN, S. R. (1967) *Prolegomena to library classification*. Bombay, Asia Publishing House. 640 p.

RUMBAUGH, J., BLAHA, M., PREMERLANI, W. et al. (1994) *Modelagem e projetos baseados em objetos*. Rio de Janeiro: Campus.

THÜRING, M.; HANNEMANN, J.; HHAKE, J. M. (1995) Hypermedia and cognition designing for comprehension. *Communications of the ACM*, v. 38, n. 8, p. 57-66.

VICKERY B.C.. (1966) *Faceted classification schemes*. New Brunswick, N.J., Rutgers State University.