DYNAMIC THEMATIC ARTICLES FOR ON-LINE NEWSPAPER

Sébastien Iksal
Department of Artificial Intelligence and Cognitive Science
ENST Bretagne
Technopôle de Brest Iroise
B.P. 832, 29285 Brest, France
E-mail: sebastien.iksal@enst-bretagne.fr

ABSTRACT

Today, Press institutions have to consider Internet as a new medium because of its high availability. The Press behaviour has to change, some daily newspapers have soon an electronic version on the Web. One of the new features of this medium is the opportunity to deliver adaptive information. ICCARS aims at assisting the journalist in writing adaptable thematic articles. The adaptation can be to the reader, but also to the medium used to read the article. The document is considered as a set of fragments with a particular structure in order to compose it dynamically. ICCARS has just started, so this paper is a short description of its goal and propose our first assumptions.

KEYWORDS

Dynamic composition, indexing, ontology, adaptive presentation.

1. INTRODUCTION

The Press institution began to use computers as electronic tools for publishing. At this time, all software were dedicated to printed Press documents. The new world of information and communication increased the number of computer features. The high availability of Internet modifies the organization of newspaper's offices, as well as the Press behaviour. Now journalists work with electronic mail, chat, search engines …, and use Internet as a way for accessing information and getting contact with Press agencies. Today numbers of daily newspapers in the world propose an electronic version. A lot of Web users are interested in on-line newspapers and above all those who are away from their country and cannot get access to the local newspaper. So it's easy to understand that Internet brought a new economic model for the Press, these are new actors, new methods, new tools, and requirements – for instance the delay for publishing an article is very reduced, but the most important innovation is this new medium : The digital document – in fact the hypermedia document.

Computers were first used as a tool for managing the presentation of information and after for information retrieval. The new challenge is the publication on the Web, that is why our interests concerns computer assisted reporting systems. The definition of Houston [1] clarifies the role of the reporter and the system's one : *Computer assisted reporting implies the use of a computer not only as a word-processor, but also to facilitate investigations through on-line databases, to collect a lot of data from official institutions, to analyse this information and to use the result for writing articles with more benefits of hindsight and context's precision than never before.*

New problems and issues about the role of Press arise with Internet. On this network, everybody is able to search and find oneself information. The Internet user is able to use various search engines to collect information. Nevertheless, this set of data needs to be analysed and synthesized. So, it is not sufficient for a newspaper to propose raw information (news wires of Press Agencies). The most important work of the Web journalist is to select, synthesize and analyse information and events for his readers.

Within the framework of our project, we have chosen to apply dynamic composition on thematic articles. A thematic article is a document of several pages which is an analysis of a particular event. The document falls into several parts which correspond to specific points of view. It is written by a single reporter or by a small group, and it is a single work, not a collection of different jobs. This kind of document is able to highlight the role of a journalist. With digital documents, the journalist can detail his synthesis and his arguments, so his work may be more developed than in "printed" version. With the use of new technologies of information and communication, the cost for publishing an article decreases in proportion to the growing number of readers. So newspapers can increase the cost of advertisement, that is why "digital" Press may be more interesting than "printed" Press. Because of the vast readership, the journalist can changed his manner of writing, in fact, it will be better to think about several versions of his article than a single document for numbers of readers. That is why an article has to be adapted to a reader by using some rules given by the journalist. We can noticed that the reader is interested by a presentation and a content adapted to his preferences and interests.

ICCARS (Integrated and Cooperative Computer-Assisted Reporting System) is a prototype of an authoring system for the dynamic generation of thematic and adaptable articles. The system will not replace the journalist, but it aims at assisting him to write adaptable documents. Because this project has just started, we attempt in this article to present some problems and to give some of our initial assumptions. After a short explanation about "news providers" on the Web, the ICCARS project is introduced, through the project, we considered some problems and the new features which correspond to the digital document. During a first discussion, some journalists clarified some aspects of the project. Next, some research and development orientations we shall try to explore are proposed. A document can be described through three structures which are detailed. The storage and re-use aspects about documents are also introduced.

2. PRESS AND INTERNET

Numbers of Press institutions on television, radio as well as the printed Press have today their own digital world. Internet is more considered as a new challenge of communication than an interesting tool. Some daily newspapers propose their "printed" edition and the digital one at the same time (Le Monde[2], le Télégramme [3] …). Others like monthly magazines differ their editions (Linux Magazine [4]). About adaptation to the reader, some Web sites personalize the access to information and others the search engine. Sometimes the organization of the front page is computed on the fly. Internet brought also new information providers such as "city-guides" and "Internet providers", so Press institutions have to take them into account.

2.1. News on-line

Today, it is not rare to find a Web site for a national or local daily newspaper. They provide news through this new medium : Internet. Even if these sites are very closed to the printed organization, they propose new services more user oriented such as search engines about archives, links through the other sites of the press agency (Celtimusic or vivabrest for *Le Télégramme*; sortir.lemonde.fr, interactif.lemonde.fr, finances.lemonde.fr … for *Le Monde*), free e-mail …

2.2. Web sites for personalized information

Internet increased the need to satisfy the reader, that is why numbers of sites propose personalized answers. Someone provide all the most interesting news according to your preferences through e-mail such as e-revue[5]. For a Web site, the most interesting solution is proposed by Crayon[6] which assist the reader in organizing his own newspaper (it is possible to name it like *"The MyNewsPaper Post"* or *"The MyNewsPaper Tribune"*). The internet reader is able to modify his newspaper and select who is allowed to read it.

2.3. Personalized information

A lot of Web sites propose to personalize the access and the presentation of the information written for the printed newspaper. They don't think about adaptation in the writing of an article. Two projects work with personalized news which can be read through a Web site. *Sistemi Telematici Adattativi* [7] is a project which propose to filter and display news and ads according to user's preferences and characteristics. *KMI Planet* [8] is a kind of private on-line newspaper where all readers and writers are in a same group. It collects news through e-mail, processes and sends the result to the most interested readers. The tool is able to order articles in order to fill in gaps, and after to inform the reader when the news is ready. It propose also an advanced interface for searching documents.

2.4. Press and its rivals

During a long time, local Press agencies were the main local news providers, but with Internet, new actors like "city-guides" propose local news. At the beginning, city-guides aimed at proposing information about public services, association, weather … Today, they have their own team of

journalists ([9]) and propose national and local news. That is why, the answer of local Press agencies was to create their own city-guides such as http://www.vivabrest.com for *Le Télégramme* and http://www.brest.maville.com for *Ouest-France*.
National Press agencies are also concerned by the phenomenon because some sites such as Internet providers propose classified news (http://actu.libertysurf.fr). They receive all news wire from agencies like AFP (Agence France Presse).
It is important to notice that the job of national, local, daily or weekly Press is not limited to provide raw information. They write also special issues, feature articles, archives that we cannot find on the other sites. The main challenge for the Press on the Web is to become not only information providers but also to offer services around information. The printed Press loses numbers of readers, so Internet is a new medium useful for increasing their readership and in fact their income with advertising, e-business …

3. ICCARS PROJECT

ICCARS is a joined project between the IASC laboratory, a private company called Atlantide and a regional daily called Le Télégramme. It is funded by Brittany Council. The ICCARS prototype will be a computer assisted reporting system. Its main goal is to assist the journalist for writing a digital thematic article. Digital documents lead to new problems and offer new features for writing them and for displaying them.

3.1. New problems

As opposed to the "printed" document, the digital document has two main particularities which came from multimedia documents:

- This kind of document uses some new objects such as audio and video. These objects are useful for emphasizing a particular point of the document. But it's difficult to select and organize good media between text, picture, audio and video.
- A digital document is divided into several information units. Then, it is possible to read them in different ways. But, the reader has a fragmented view of the document which does not help him to read and understand it.

In order to access information, the reader has to navigate between different units. Navigation implies that he may get lost in the information space, then the comprehension of the document may decrease. In cognitive science, the comprehension of a document is often characterised as the construction of the mental model that represents it. The readability of a document can be defined as the mental effort spent on the construction process [10][11]. In a way of increasing the readability of the hyper-document, it is necessary to assist the user in the construction of his mental model. Improving the reader comprehension needs to have a consistent document, so it must have a global structure which has to be stable and easy to understand. This structure has to be conveyed to the reader as a model of the document. In order to reduce cognitive overheads, some efforts have to concern orientation and navigation. For avoiding the disorientation, readers have to know the overall document structure, then it is necessary to give some tracks of their moves through that structure. Indeed, comprehension and orientation are closely related [12][10].

3.2. New features

Digital document is a medium with some interesting features quite different to the "printed" ones. The size of the article is no longer limited, so the journalist can offer a huge document. This enrichment and the fact that the document is composed of a considerable set of information units allows :

3.2.1. Different ways to read the same information

The system can propose, according to the author, different navigation guides to read the document :

- Some guides are user dependent :
  - Semantic or conceptual (global structure),
  - Ontological : according to relations between objects of the domain,
  - Task oriented : according to the user objective.
- Some others are user independent :
  - Temporal : chronological order of the events,
  - Causal,
  - Spatial.

All these ways has to be planned by the journalist, and when they are user dependent, the system applies some rules based on a user model.

3.2.2. Reader interaction and document evolution

One of the main issues for the on-line Press is that an article must be "alive", that is to say accepting the readers reactions and/or to be modified according to new events. Readers can participate with polls or discussions [13]. It is very interesting in the case of thematic article, in fact, the "on-line journalist" cannot publish his document many weeks after the event. The journalist must start to deliver the

information really closed to the event, and propose his analysis regularly until the end of that event. As an example, during a court action, the journalist must open the press kit at the beginning, expand it with the main events without redoing his work. For this, the prototype needs to provide tools to enrich and manage an article.

3.2.3. Adaptation to the reader

ICCARS is concerned with a particular readership, thematic articles will be composed for Internet users that is to say a vast and heterogeneous audience. Readers may have different goals, backgrounds, interests, knowledge and consequently they are not interested in the same pieces of information. It is not conceivable to propose all fragments because they are too numerous. The information space of a digital thematic article is so vast because it has to satisfy the major part of readers. So dynamic composition will use adaptive rules done by the journalist in order to adapt the content for a particular user. All of the article has to be accessible by the reader, the system will only reorganize the access. It is necessary to adapt also the presentation of a document according to the reader - interests, level of comprehension, objectives or preferences.

A well known method to adapt a system to the user is done through user modelling. A user model is able to represent various characteristics of a group or a particular user. Stereotypes, introduced by Rich [14], are an important element of user modelling and it has been extensively used because they give a simple but powerful way for adaptation [15][16]. Overlay models are powerful, but very difficult to obtain [17]. According to ICCARS, stereotypes would be used as the basis for building a kind of overlay model, in fact the global structure. An overlay model allows the user to have an adaptable and evolutional structure. The reader will have a personalized content and presentation. In order to not censure any information, some methods will be studied and will offer all that the journalist chooses to show. Some methods are stretchtext – to develop some fragments, contextual layers - to present some definitions ...

3.2.4. Adaptation to the medium

Everyday, more people and companies are interested by Internet as a new medium. And this kind of medium has modified in depth some strategies and practices in Press environment. With ICCARS, Internet is considered as a new way to distribute the information, but another issue is to propose information on various media such as PDA – Personal Digital Assistant -, Television, cellular telephone … Often, adaptation is made according to the user himself, another way of adaptation can be taking into account the device used by the reader in order to access the information.

All these assumptions will not be considered in the project, only someone will be favoured according to journalists requirements.

3.3. Results from first discussions

All these assumptions are the basis of our project. During the first interview, journalists clarify these points :

- They want to manage everything in the final document. The journalist is the author of the article, that's why he wants to define the global structure, the domain ontology and the adaptation rules. As opposed to the project *Sistemi Telematici Adattativi* [7], the journalist composes himself the document. So it's important to distinguish the effective role of the journalist in one side and the system's one in another side.
- They want to write only a single article - called "virtual article".
- No censure : Readers must be able to read all the document. It's the job of a newspaper to give all the information, they must not hide something.

4. RESEARCH AND DEVELOPMENT ASSUMPTIONS

ICCARS Project is planned for three years, and is organised into three phases. At the end of each phase, a prototype is given for validation to some journalists. They will use the system and give us a feedback about it. In order to be really closed to journalists, the design of ICCARS is user's oriented.

For reducing the publication delay of an article, the editing environment will be on-line. This solution allows the journalist to write his document through the Web, and then from every computer connected to the network.

In order to adapt the digital document, it's important to distinguish the document itself from its included objects. These units are called *documentary units* [18], *fragments* [19][20], or *information bricks* [21]. We choose to use the word fragments for this article to avoid confusion with existing terms. A fragment is a part of a document, but it may also be a part of another fragment. So it is useful to distinguish two kinds of fragments. The first one is the atomic

fragment, that is to say a "minimal unit of meaning" which is used in the process of dynamic composition. Its main characteristic is to be used directly without changes. A unit of meaning can be considered as a container to group several objects which are not separable – a graphic and its caption, a video and its comment … The abstract fragment is the other one. It is composed of several fragments and an organization between them given by the author.

A document can also be described through three levels, called structures [22] :

- The semantic structure (also called conceptual) which organizes the meaning of the document content. A solution like RDF (Resource Description Framework) [23] is considered for representing this structure and for fragments indexing.
- The logical structure (or syntaxic) corresponds to the author defined organization of the document in order to increase its readability. XML (extensible Markup Language) [24] is considered because of its ability to separate the content and the presentation. So, XML allows the system to display the document on various media.
- The physical structure corresponds to the rules for presenting the document on a particular medium. It is possible to use stylesheets with XSL (extensible Stylesheet Language) [25] for instance.

Content and presentation are completely separated by the use of these structures. About thematic articles, the logical structure is approximately the same for all documents as opposed to the semantic structure which reflects the author side of the journalist. In ICCARS, the journalist will have to create the semantic structure with a very high level of details. Writing an article will consist of filling an instance of this structure. The system will use the instance in order to compute adaptive structures for various readers.

Writing an article implies to have a collection of objects, the journalist will use internal or/and external databases. Internal database refers to documents built with ICCARS such as fragments, global structures, adaptive rules, and composition rules … The external one refers to documents collected through the Web or other information sources. In order to help the journalist, the access to these different databases must be the same. The journalist will be able to store his documents as well as those he collected to write them, so the information written by the journalist will be able to be re-used in the same article or in others. In order to re-use documents or fragments, the journalist will have to index them, a good way is to use the domain ontology and the logical structure of the object for indexing. This will facilitate the retrieval process for the reporter as well as for the composition tool, for instance to find bibliographies, personal cards … A single domain (sailing, politics …) will be selected by *Le Télégramme* for reducing the domain ontology. They will choose a domain according to the availability of their journalists and events planned for the three next years.

5. CONCLUSION

The high availability of Internet brought a lot of disruptions particularly about the diffusion of information and new practices concerning its management. Every Internet user is able to create his own Web site, everybody can give his information. The second disruption is that Internet can be viewed as a kind of new encyclopaedia with a lot of search engines. But the result of these tools is vast, without organization, analysis and absolutely not synthesized. Press institution has a new role in this community, ICCARS will assist the reporter in collecting, and filtering information, and in writing his analysis.

In order to build a user-oriented system, an analysis of journalists needs and usage has to be done. To adapt the document, and also to propose some features for the reader, we will have to study usage of Internet users. Knowledge about readers will lead our choices concerning user models, adaptation and features of the system. Knowledge about journalists and readers will help us in the conception of a method for writing adaptive thematic documents. We will have to define a relevant global structure, and composition methods to apply on this structure for displaying adaptive documents. The first phase will end at the beginning of 2001, a simpler version of the prototype will be given to journalists in order to receive feedbacks.

REFERENCES

[1] B. Houston, *Computer-Assisted Reporting : A practical Guide*, St. Martin's Press, 1996, P.3.

[2] http://www.lemonde.fr

[3] http://www.bretagne-online.com

[4] http://www.linuxmag-france.org

[5] http://www.e-revue.com

[6] http://www.crayon.net

[7] L. Ardissono, L. Console and I. Torre, *Exploiting user models for personalizing news presentations*, 2nd Workshop on Adaptive Systems and User Modeling on the WWW, AH'99 and UM'99, 1999.

[8] J. Domingue and E. Motta, *A Knowledge-Based News Server Supporting Ontology-Driven Story Enrichment and Knowledge Retrieval,* Submitted to the 11th European Workshop on Knowledge Acquistion, Modelling, and Management (EKAW '99), 1999.
[9] http://www.webcity.com
[10] D. Kaheneman, *Attention and Effort*, Englewood Cliffs, Prentice Hall, 1973.
[11] M. Thüring, J. Hannemann, and J. Haake, *Hypermedia and cognition: Designing for comprehension*, in Communication of the ACM, volume 38, pages 57-66, 1995.
[12] J. Conklin, *Hypertext: An introduction and survey*, IEEE Computer, 1987.
[13] *The future of the printed press, challenges in a digital world*, European Journalism Centre, 1998.
[14] E. Rich, *Stereotypes and user modeling,* in Users models in dialog systems, pages 35-51, Springer-Verlag: Berlin, 1989.
[15] A. Kobsa, *User modeling: Recent work, prospects and hazards,* dans Adaptive User Interfaces: Principles and Practice, North-Holland, Amsterdam, 1993.
[16] J. Kay, *Pragmatic user modelling for adaptive interfaces*, Adaptive User Interfaces, pages 129-147, 1993.
[17] P. Brusilovsky, *Methods and techniques of adaptive hypermedia*, User Modeling and User Adapted Interaction, 1996.
[18] C. Michel and S. Lainé-Cruzel, *Profil-Doc: Un prototype de système de recherche d'information personnalisé selon le profil des utilisateurs*, Atelier DVP, IHM'99, 1999.
[19] S. Tazi and Y. Altawki, *Création de documents virtuels : Cas des supports de cours*, Atelier DVP, IHM'99, 1999.
[20] S. Garlatti and S. Iksal, *Documents virtuels personnalisables pour des systèmes d'informations en ligne*, Atelier DVP, IHM'99, 1999.
[21] S. Ranwez and M. Crampes, *Conceptual Documents and Hypertext Documents are two Different Forms of Virtual Documents,* Proceedings of the Workshop on Virtual Documents, Hypertext Functionnality and the Web, Eighth International World Wide Web Conference, Toronto, Canada, May 10-15, 1999
[22] C. Vassilis, *Electronic document management systems: An Introduction*, UCH/FORTH, 1998.
[23] W3C Recommandation : *Ressource Description Framework (RDF) Model and Syntax Specification*.
[24] W3C Recommandation : *Extensible Markup Language (XML) 1.0*.
[25] W3C : *Extensible Stylesheet Language (XSL)*.