A semiotic approach to building a domain-independent discourse model for Intelligent Multimedia Presentation Systems.

Abstract

This paper argues that RST alone does not have the capability to efficiently support the development of a domain-independent discourse model for IMMPS. This will be achieved by providing, in the introduction, a scenario that highlights current usability and theoretical problems of the World Wide Web, by envisioning an evolution of IMMPS to solve them, by redefining, in semiotic terms, the problem of developing a domain-independent discourse model for IMMPS and, finally, by evaluating the capability of RST to provide solutions to such a problem.

Introduction

The steady increase in the amount of information available on the web, the quantitative and qualitative broadening of the range of possible web users and usage environments (display devices, connections, social and physical environments) and the current atomized structure of web information are major characteristics of today's World Wide Web. This section describes how these result in a set of usability and theoretical problems that affect authors and end users.

The authoring effort
In order to keep up with the qualitative and quantitative usage contexts broadening, authors are forced to an extra authoring effort due to the explicit, manual authoring process required for every possible context of use (users' characteristics, usage environment, task) content modification and domain of the web application.

Information filtering
The increase of redundant information makes finding and selecting among similar resources on the web increasingly hard for users, making current search engine technologies seem absolutely inadequate. Huge research efforts are carried out (by W3C), in order to semantically annotate web documents/media objects and constitute ontologies to semantically relate them, to allow a more intelligent semantics-based retrieval. In order to turn this new technology into a real improvement for the user, though, an equivalent research effort is required in developing semantic web enabled tools capable of tackling the current usability problem of filtering information resources.

Knowledge reproduction
Web information is structured by authors, that explicitly define spatial, temporal and linking relations among media objects/information chunks. Since authoring is a hermeneutic process , it implies the interpretation of the meanings of the media objects and of their semantic relations. The author, by giving a structure to media objects/information chunks, conveys his interpretation of their meanings and semantic relationships to users, that, being used to give meaning to spatial and temporal relationships (linking turns into a mixture of these two) of any communicative artifact, turn syntagmatic relations back into paradigmatic and reconstruct the original intent of the author.

The structure of web information is atomized, in the sense that it is scattered throughout a multitude of web sites, which are in turn structures of nodes, which are in turn structures of media items and so on. Web sites are the biggest information units to preserve coherence, and they can thus be considered the primary elements (atoms) of this structure. It is possible for the user to directly jump from one to the other (e.g. through search engines results pages or external links in web sites). This activity, though, is currently strongly limited by the scarce affinity to users' query of the documents retrieved by search engines, by the fact that the external links have to be provided for by the author, and, most of all, by the fact that coherence of discourse is not preserved during this process.

For all these reasons, using the term “hypertext” in the sense of non-linear semiotic text, the web is definitely not one enormous hypertext, but many hypertexts linked together. Ted Nelson's dream-hypertext XANADU, allowing anybody to access any information in his own personalized way that would reflect his cognitive structure, and avoid the arbitrary imposition of teacher’s cognitive structures through traditional linear media, is, thus, still a dream. These considerations highlight that the current state-of-the-art of the web, like the majority of current educational computing, is based on knowledge reproduction rather than knowledge building learning strategies. The former are based on the concept that the information out there exists and learning consists of reproducing it in the mind of students, and lead to a *transmission model* of discourse. The latter, on the other hand, sees learning from a constructivist perspective as conceiving new ideas. It is the way researchers and scientists learn, by creating innovative connections between concepts, and requires a different kind of discourse that, instead of simply allowing the representation of pre-existing knowledge, would enable, through *conjectures and refutations*, to actually produce new knowledge.

The evolution of IMMPS

This section describes how IMMPS should be capable of solving the issues raised in the previous section. Intelligent Multimedia Presentation Systems are systems that are able to flexibly generate various presentations for one and the same information content in order to meet the individual requirements of users and situations, the resources limitations of the computing system.

Although aesthetically not perfect, generated presentations, provide a quick means of providing selective access to massive amounts of data, alternative to pages generated by Web search engines.

These systems could retrieve semantically related media-objects from the whole web, automatically binding them in a meaningful and (from a communication perspective) effective and efficient way, adapting them to contexts of use and domains.

Automatically binding media objects retrieved from different information resources in one single presentation capable of preserving coherence of discourse, IMMPS would provide the user with a completely new and freer way to access information present on the web. Spatial, temporal and linking relations among media objects, in fact, would be automatically created on the basis of many variables (user profile, device/connection characteristics, query, task) and media objects' semantic annotation and not on the original intent of the author, thus allowing the creation of new, but nevertheless semantically significant, associations between information chunks/media objects. Allowing an easier, tailored to context of use and domain, simultaneous access to information from different resources/domains, bound together in a coherent discourse, IMMPS would, thus, get the web closer to a real 'open access' hypermedia, enhancing the use of the web as a knowledge building technology.

Allowing coherence throughout the whole web, instead that only inside the boundaries of every web site, and an infinite number of dynamic and unstable relations between media objects, Web information accessed through IMMPS could effectively be represented by the metaphor of a sea of information particles (media objects), continuously coming together and parting, instead of a web made of sites connected through static and stable links.

Semiotic approach to a domain-independent discourse model for IMMPS

A semiotic framework

A presentation is any composition of media objects created for the purpose of communicating information to a user. Presentations are communicative artifacts, vehicles for human thoughts, i.e. “texts” in the semiotic sense.

Hypertexts or multimodal presentations, in fact, are not evolutions of traditional linear text, but just another manner for the text to express its potential. Being characterized by an

organizational structure and a particular communicative function, they are indeed texts. They are examples of the broadening of the range of forms that text can embody. The broad category “text” includes, thus, also movies, books, paintings, commercials and so on. In particular, hypertexts are non-linear texts, i.e. texts that are meant for a use through more than one sequence. Multimodal texts, on the other hand, are texts that rely on more than one modality, i.e. a combination of a semiotic “code” and a channel, to accomplish their communicative function. Just like the concept of text, the definition of hypertext or multimodal text does not imply the electronic support. Presentations generated by IMMPS are hypertexts and multimodal texts at the same time.

The organizational structure of any text is composed by a set of communicative elements and a set of relations between them. The overall message conveyed by a text is, in fact, not derived by the sum of the single messages conveyed by each communicative element, i.e. as if each one of them was abstracted from the text and considered alone. It is something very unique, that can also be radically different from the messages of the individual communicative elements.

Discourse models for IMMPS

Presentations generated by IMMPS are complex interactive discourse structures. Discourse structures can be divided into six discourse categories: narrative, argument, description, procedure, simulation, and association. Causal links are the fundamental relations between information chunks that characterize narratives and that allow them to be comprehensible. Arguments have to maintain relations between conclusions and supporting evidence. Descriptions present relationships between the parts and the whole, while procedures clarify relations between conditions and actions. Simulations are descriptive and procedural at the same time, presenting parts-whole and condition-action relations. Associative discourse, such as lyrics and poetry, uses symbolic links. All these relations guarantee the meaning of the whole text and have to be preserved also in hypertexts.

The multimodal character of presentations generated by IMMPS, though, raises some new issues. Each medium, for example, should be used in a presentation if it is suitable to convey and process that kind of information, and the relations between media have to be carefully considered. Modalities, being characterized by an output modality (e.g. visual, auditory etc..) and a code (linguistic, iconic etc...) are intrinsically better suited to convey a certain type of information (concepts) rather than others. The modality "music", for example, is not suited to convey the concept "mouse", whereas an image would probably be the most effective modality in this case, even more than text. In fact a nice icon representing the animal would probably avoid having the user disoriented by the ambiguity of the word "mouse" that could also refer to the input device. On the other hand, text and speech, using linguistic code, are inevitably more suited to convey abstract concepts. How many images would be necessary to convey the concept "tyranny"? Would they convey it precisely and immediately as just one single word does? Would they convey the broad concept or only several instances of tiranny? The characteristics of the code generally strongly influence the capability of modalities to convey certain kinds of information rather than others.

Besides the communicative act, the selection of a particular modality should take into account for the selection also other *domain variables* like context of use, preferences, performance parameters (e.g. more efficient), learning parameters (e.g. learning overhead) and cognitive properties (e.g. attention load).

Furthermore, not only the system has to correctly choose which modality is more suitable to convey a certain part of the global information, but must also consider relations among the messages (information) conveyed by each modality. E.g. Besides knowing that music is suitable to act as “context”, and image and text to convey the main information, it should also consider whether the message conveyed by the text is complementary or analogous to the message conveyed by the image. If they are complementary, for example, the system should differently consider how to relate the two media objects in the presentation dimensions and never consider the possibility of having one without the other, not to corrupt the overall message.

IMMPS require levels of abstraction higher than final presentation, allowing authors to write specifications at these higher levels, and lower-level presentations being automatically generated to meet these specifications. Presentation-independent authoring abstractions currently in use include presentation constraints and rhetorical structure. Constraints define spatial, temporal and navigation limits and relationships of a presentation, Rhetorical structure, on the other hand, represents the conceptual flow along which a presentation progresses. Abstractions of document presentation, provided by constraints and rhetoric, provide several benefits, like removing much redundant encoding expected from the author, and making the presentation adaptable to a wider variety of use cases . It has been shown how Rhetorical Structure Theory is a higher level of abstraction than constraints , and can be used to derive constraints. Rhetorical Structure Theory has, thus, already been employed to support IMMPS discourse models.

Limitations of Rhetorical Structure Theory

Tree structure

RST argues that any text can be described by a tree structure. The tree structure is inadequate if we acknowledge that multimedia presentations are multimodal texts, because every text (we are talking about semiotic text), is composed by elements that are all interrelated. Each media item of a text, in fact, is semantically related to all the other media items of the text, through relations that, along with the messages of the single media items, compose the message of the text.

Using RST tree structures to describe texts for IMMPS has consequences on the content selection process that is carried out (for every media item) on the basis of only one relation to another media item or to the query. The consequences of this are tolerable when IMMPS use a single and coherent database, since the relations that are not provided for by the tree are implicit in the database author intent, who has previously already made a selection among which media objects the database should hold. This previous selection/interpretation of the media objects done by the database author is what currently avoids having IMMPS come out with hilarious results like a picture of "Martin Luther King" close to a text that talks about ancient African slaves, just because of their individual relation to the concept (keyword) "racism". The temporal contrast that would strongly affect users' comprehension, could be avoided, for example, by providing another semantic relationship that would, at least, set another constraint among the years which the text and the image refer to, and, thus, avoid this association. This same operation, though, would also change the tree structure into a more complex structure. For all these reasons, the tree structure is not adequate for a domain-independent discourse model.

RST relations

RST argues that any text can be described by a set of RST relations. These relations, though, connect two media elements and are, thus, not capable of representing how every element is related to the overall message conveyed by the whole “multimodal text”.

Moreover, Nucleus-Satellite relations associate a piece of knowledge with another less important piece of knowledge. Contradicting the assumption that every element of a text is equally important in conveying the overall message, importance is implicit in every RST relation. It is determined a priori, regardless of domain and context of use, and in a completely arbitrary way. E.g. The RST relation “Elaboration” relates a Nucleus composed of “basic information” and a Satellite composed of “additional information”. But there's no other RST relation reverting this order and making “additional information” be more important than “basic information”.

The difference of importance set by RST nucleus-satellite relations is what many systems use to decide which of the two elements to remove, for example depending on the display device characteristics. In some cases, though, for example with expert users looking for insight information, the decision of removing the “additional information” instead of the “basic” might definitely not be appropriate.

Conclusions

Taken for granted that RST has provided great improvements to IMMPS, allowing clear structural definition of texts, it seems to fall short of the objective of supporting the development of a domain-independent discourse model for IMMPS. In particular it does not seem capable of efficiently accounting for the multimodal nature of IMMPS presentations. In fact, it provides a general representation of the flow of the presentation that allows coherence to be preserved, but does not seem to be able to efficiently account for the *media selection process*. Whether RST would need improvements, or to be backed by other theories, or should be completely substituted is out of the scope of this paper.

References

M. Bordegoni , G. Faconti , S. Feiner , M. T. Maybury , T. Rist , S. Ruggieri , P. Trahanias , M. Wilson, A standard reference model for intelligent multimedia presentation systems, Computer Standards & Interfaces, v.18 n.6-7, p.477-496, Dec. 1997

Davida Charney. *Comprehending non-linear text: The role of discourse cues and reading strategies*. In Proceedings of Hypertext '87 (The University of North Carolina, Chapel Hill, North Carolina; November 13-15, 1987), pages 109--120, November 1987.

Linn Mark Collins. Structural issues in Multimedia Design, in Proceedings of the CHI '96 conference companion on Human factors in computing systems : common ground: common ground, 1996 , Vancouver, British Columbia, Canada .

Marlene Scardamalia , Carl Bereiter. Technologies for knowledge-building discourse, Communications of the ACM , May 1993, Volume 36 Issue 5 .

Popper, K.R. and Eccles, J.C. The Self and its Brain. Springer-Verlag, Berlin: 1977.

Lloyd Rutledge, Jim Davis, Jacco van Ossenbruggen, and Lynda Hardman. Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure Proceedings of the International Conference on Multimedia Modeling 2000 (MMM00) (pages 89-105), November 13-15, 2000, Nagano, Japan.

S. J. Morris , A. C. W. Finkelstein , Development of multiple media documents, Proceedings of the conference on Organizational computing systems, December 1993

Lloyd Rutledge, Brian Bailey, Jacco van Ossenbruggen, Lynda Hardman, and Joost Geurts. Generating Presentation Constraints from Rhetorical Structure. Proceedings of the 11th ACM conference on Hypertext and Hypermedia (pages 19-28), May 30 -- June 3, 2000, San Antonio, Texas, USA.

J. Krause. A Multilayered Empirical Approach to Multimodality: Towards Mixed Solutions of Natural Language and Graphical Interfaces. In Intelligent Multimedia Interfaces [20], pages 328 352.

Oscar Rosell Martinez. Design dependencies within the automatic generation of hypermedia presentations, Technical University of Catalonia, June 30, 2002, Published as CWI technical report INS-R0205

N.O. Bernsen and L. Dybkjoer, A theory of speech in multimodal systems,.....

Frank Nack and Lynda Hardman. Denotative and Connotative Semantics in Hypermedia: Proposal for a Semiotic-Aware Architecture. (technical report INS-R0202), March 2002.

William C. Mann and Sandra A. Thompson. Rhetorical structure theory: A theory of text organization. Technical Report ISI/RS-87-190, USC/Information Sciences Institute, Marina del Rey, CA, June 1987.