Attention Capturing Story-based Individualized Information Integration: A New Way to Browse the Web

Goran Rakocevic[1], Herman Maurer[2], Veljko Milutinovic[3]
[1]*Mathematical Institute, Belgrade, Serbia*
[2]*Institute for Information Systems and Computer Media, Technical University Graz, Austria*
[3]*School of Electrical Engineering, University of Belgrade, Serbia*

Abstract - *Web search has developed from simple information retrieval to complex tools. The same is not yet true for Web information management. Heterogeneous databases are still state of the art and make thorough understanding of content hard. However, small well curated locally focused servers of cultural collections emerge throughout Europe, offering a perfect basis for making a leap in Web information management. Therefore, we propose an innovative way of presenting and accessing the information on the Web. The central concept is that of a personalized storyline offering an organized "tour" through resources. Based on content metadata and on information about uses of a resource, a system would automatically generate storylines fitting users' requests. But to capture attention story telling alone is not sufficient: humanity is starting to be unable to concentrate deeply onto any one single aspect for any length of time. In addition to the story telling issue we are going to use ideas to capture attention, like showing a cultural artefact with an error or asking a question after a video-clip. We also will tie in the Community deeply into building storylines.*

1. INTRODUCTION

The Web has turned into a rich repository of information. It has become clear that each piece of information has to be enriched with extensive meta-data, and indeed thousands of projects trying to provide meta-data are (fortunately) currently under way. Thus, we can assume that at some not so far distant point in the future, all (important) information will come with rich meta-data. But now comes the decisive question: SO WHAT? Yes, we will be able to find pieces of information much better than right now. But what is really the point if we extract a piece of information without being able to see it in context with other information, let alone in the context we are interested in? It is clear that a further layer is needed on top.

Note in passing that the lack of what we are wanting to achieve is the reason why information repositories (from Wikipedia to Europeana) suffer under the "short visit syndrome": users enter the resource, usually by means of a search engine, read two pages, and leave again. They are not even helped, let alone encouraged, to explore and learn more about the knowledge available.

1. BASIC CONCEPTS OF THE PROPOSED APPROACH

What we want to get back is an interesting story that is already personalized for us, but that, in addition, basically consists of snippets that we can choose to consume or not. To address this issue, we propose a novel concept in access to online data: the story line. A story line consists of a sequence of points, each point with some short and easily consumable information. The points shown may already depend on the personal profile. Each point has the option to branch out into topics we are interested in, where topics fitting our personal profile are offered more prominently. Each such branching out ("detour") means entering a new story line that eventually leads back to where our detour started. Each detour can have further detours, etc. Users following a story line more than once may be alerted to detours they have already exhausted.

There is infinity of story lines. Yet one can define a number of classes that can be generated automatically or almost automatically. Samples are: Each city or region gives rise to a story line (exploration of that city or region based on the personal profile), each stretch of river or road (exploring things along they way), each famous person (exploring it and related persons, events and such), a period of time (possibly restricted to some area, hence taking us e.g. through France from 750 to 1700). Very important story lines also come from tourism, from museums, from publishers (a server giving tours of big museums or tours for publishers: a story line about their collection of history books, thrillers, cooking- books, you name it.)

Everyone is talking today about personal profiles (as we have done above), yet once more we need a jump in quality. Among the items that we believe has not been considered enough is personalisation based on the vitae of the person involved. To be specific, if a person has lived for some time in places a, b and c, information relating to those places are likely to be more interesting than something relating to other places. If a person has become professional in a certain area, information on that area is likely to be of

interest (but watch out: only at an appropriate level of expertise: if we e.g. know that this person is interested in flowers, we better know whether we are dealing with a 5 year old child or a professor in botany!).

Assembling a story line with its detours in a coherent fashion is clearly a formidable if not science fiction task. The larger the repository of information we are concerned with, the more difficult the task will be: Much similar information on the same item might be found (based on available meta-data), and even different pieces may not fit together because of reasons of style or such. Hence we are convinced that we cannot fully automatically offer top story lines with attention-grabbing detours: to achieve this we have to use the wisdom of the crowds, we have to encourage the community to put together pieces (that our system has found) into a coherent fashion, involving additional personal experiences, anecdotes and such. We also believe that the integration of new versions of Web Books will be of substantial help, simply because snippets of such books may be exactly what can be used for detours, if the meta-data description of the books is of fine enough granularity.

Our long-term vision is that resource servers like Europeana, Wikipedia, Flickr, etc. etc. and other publicly accessible repositories are used to produce a network of hundreds of local (regional) servers allowing Attention Capturing Story-based Individualized Information Integration. We believe that the integration of new versions of Web Books will be of substantial help, simply because snippets of such books may be exactly what can be used for detours, if the MTES description (meta-data) is fine grained enough. Our current project is to be seen as a very first step in this direction. Clearly, it need not in the long run be restricted to cultural heritage, but could also include scientific material. (Like, why not use story telling about advances in Natural Sciences, Engineering or Medicine! Yet this is not within the scope of this project).

2. EXISTING SOLUTIONS AND THEIR CRITISISM

A number of solutions and on-going research and development initiatives do exist, that already focus on concepts that are similar to the ones that are proposed on this paper. In this section a brief summary will be presented, along with a discussion on how the here proposed solution differs.

Project *MOSAICA: Semantically Enhanced Multifaceted Collaborative Access to Cultural Heritage* [1] offers the possibility for user created storylines from the available content. The stories in MOSAICA (called Virtual Expeditions) are manually created, unlike the automated generation in our proposed approach.

A Slovenian national project DEDI II (www.dedi.si) was envisaged by some of the researchers in the MOSAICA project, where the storytelling was implemented with use of the lightweight wiki-like syntax. The approach was tested with primary and secondary schools and was found to be very good at keeping the pupils interested in the conveyed story.

CultureSempo [2] is a Finish national project with the aim of integrating access to the Finish national digital heritage. The system is based on a Semantic Web 2.0 system, and allows for an organized display of the content retrieved based on a conceptual search. Items can be displayed on a timescale, a map, or sorted by topics. While the CultureSempo system does offer a level of organization of the retrieved content, the actual order and choice of access to items is still left to the user. Our proposed approach goes one step further, by guiding the users' access to the results.

Project BAMI is an Italian national project with the goal of making a set of digitized documents from the cultural institutions of Milan available on the Web. Like CultureSempo, the system offers the display of the retrieved content along timelines and visualization of links between items. BAMI also leaves the choice and order of access to items to the individual users, like the CultureSempo system.

To summarize, the major contribution of the concept presented in this paper will be the possibility to automatically generate storylines from the content best matching the search request and the user's profile. Surplus material will be offered as marked detours, guiding the user back to the original storyline.

3. DETAILS OF THE PROPOSED APPROACH

The central concept of the idea advocated by this paper is therefore the personalized storyline offering an organized "tour" through the large volume of available resources. Based not only on content metadata, but also on information about possible uses of a resource (so-called semantic linkage anchors) together with how resources have been used before by other users, the ACS-III system will be able to automatically generate and assess best matching storylines upon a user's request, based on the query he/she submits, but also his/her recent browsing behaviour and known areas of interest (e.g., with respect to history, arts, or economy). The major story a user follows will be constructed from the best matching and the most suitable content, and organized in linear manner, without forcing users to manually seek their own path through the content, thus alleviating the "Lost in Hyper-Space" problem [3]. To avoid long storylines that are difficult to follow, story navigation will based on several concepts:

- *"Thin" storylines:* The major storyline forming the backbone of the user navigation will contain only the most relevant and suitable content.

- *Detours*: If users wish to extend their insights beyond the content in the major storyline, they can take *detours.* Those detours can either be sub-stories that cover this additional content, or are based on the t*ransclusion of document elements* [i.e. the most relevant segments of selected documents will be inserted into the content presented to the user, following the concepts of Transclusion [4], [5],] both leading the user back to the major storyline upon completion. Some detours will be automatically suggested by the system, while the user may explicitly request others.

- *Storyline reintegration: W*hen users returns to a storyline, after taking a detour, they will be shown further optional detours if such are available.

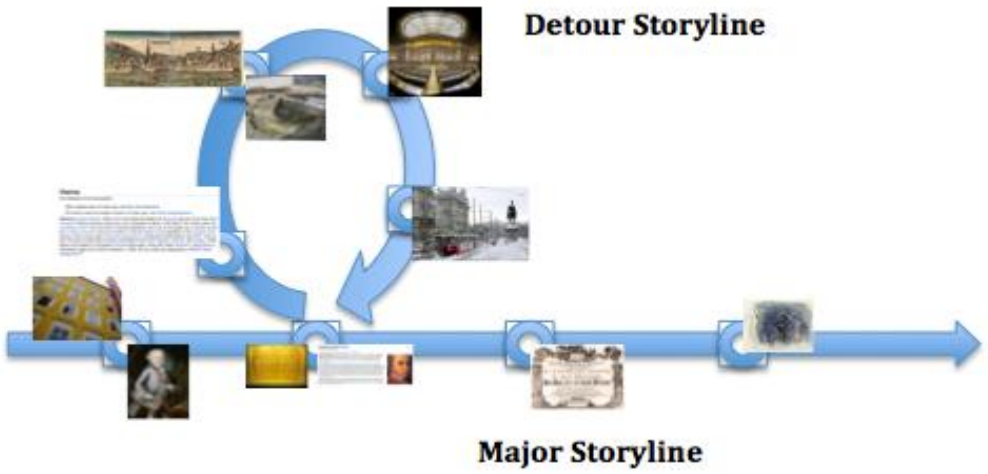

Figure 1: *Structure of a storyline and a detour*

Beside the information available in the digital repositories, a future Web system should incorporate, build on, and support sharing of a wider body knowledge existing within the community [6]. When exploring a storyline, and the selected detours, users will be able to add further material and even add detours, or denote different parts of the content they would prefer in/closer to the major story, or further away, separated in another detour. The results of these activities will be strengthened/weakened connections among entities, ultimately giving the community the ability to shape how different stories are generated, and enrich the system with the intrinsic notions of which facts are important to specific topics.

The proposed system should also explicitly foresee the option of manually editing, storing, and sharing a successful storyline, thus providing support for simplified organization of exhibitions of cultural and historical content, but also forming the building block of innovative educational tools. In addition to the novel concept of storylines we are going to incorporate a set of paradigms, which will capture the user's attention. To make this less abstract, let us provide just two examples. We might show an old cultural artefact but it will contain one error; or, after the end of a video-clip the viewer has to answer a question or two. In both cases the user will be proud to have succeeded.

4. USE-CASES OF THE PROPOSED APPROACH

To better illustrate the vision and the benefits of this proposal, the following use cases are presented. Let us consider three typical users of the system: Peter and Lisa are using the system for information gathering, whereas Alexander is using the system for the presentation of cultural heritage:

Peter and Lisa are interested in Roman settlements in the area of modern day Serbia. Using a standard web search engine yields a messy variety of sources from which it is difficult to create a holistic picture. After the first few, yet helpful results (Wikipedia entries for Roman heritage in Serbia and different sites like Sirmium and Singidunum), there is a wild mixture of archaeology articles, historical descriptions, special interest pages on metallurgy, tourism pieces and vacation offers, even down to shopping Web sites within the first two result pages. Neither the user intent is taken into account (like planning a vacation vs. writing an essay, differences and similarities to Roman settlements in

other countries, etc.), nor is the resulting presentation easy to digest. What is the most significant settlement? Were there many? Which sites have been excavated? Which are still in use? All three users realize that selecting and organizing the information would take a lot of time and effort. Thus, they decide to use the new Web browsing system. A search for Roman settlements located in Serbia retrieves a number of articles, the first one regarding Sirmium, one of the most important sites in the ancient times. The article provides access to a rich body of information sources:

Peter wishes to expand his research into related close-by Roman locations. Rather than navigating/backtracking blindly through different search results, Peter starts from the Sirmium article and asks the system to automatically generate a *storyline*. He may decide to look only for content related to artefacts and events dated three hundred years around the creation of Sirmium, and located over a larger area, covering the central and western parts of the Balkan Peninsula. The storyline should be *focused on locations and ordered in a spatial way*, the next article is the one related to physically closest location to the current one in the current article. Going through the articles gives an impression of travelling across the country. The content selection, and the suggestions for additional reading are enriched by learning from the actions of those taking a similar path through the system in past. This gives Peter an even better view and experience of the topic.

Lisa finds out that the starting article links ancient Singidunum to today's Belgrade, her hometown. She starts to wonder how Belgrade changed through the ages to gain its present look and importance as capital. The *ACS-III* system also helps to answer this question. Lisa requests a storyline, this time covering a large time span, from Roman times to present day. But spatially, she asks for articles restricted to a ten-kilometre radius around Belgrade. This time, the storyline should be *focused on economic change and ordered by time*, presenting the evolution of the city. Besides specially written articles, Lisa goes directly to maps, illustrations and texts in old, digitized books. Lisa finds her reading fascinating and decides to store her individual storyline linked to the most interesting text passages, pictures, and maps for future use and sharing with her friends.

Both Lisa's and Peter's storylines consist of linear building blocks, with limited, but concise selections of information, most relevant for the topic and the development of the story. However, the *ACS-III* system provides pointers to additional information in a form of suggested detours, complementary sub-stories offered at appropriate locations along the major storyline. The detours are introduced in such a way that after the sub-story is finished, users end up back in the original, major storyline.

Alexander works for a Belgrade museum and is responsible for maintaining and updating the Museum's collection within a dedicated content server enabled for the *ACS-III system*. He finds that the high degree of automation the system offers is absolutely essential. When uploading a new digital book or new information about some cultural artefact, much of the necessary meta-data is automatically extracted and the pages are annotated to create anchors for possible links in a future storyline, leaving him with only the fine-tuning to do. Digital items like images, sounds, videos, and event 3D models can be seamlessly integrated into several possible Web storylines as standalone content, requiring only a small amount of manual meta-data to be entered and few possible metadata anchors and links to other items to be created. Moreover, Alexander can create, edit, and share storylines, thus creating topic-centred virtual exhibitions and guided tours, available to the general public. Recently, he found another use for this option. Occasionally, Alexander substitutes the history teacher in the local elementary school. Through *ACS-III* storylines he can extend the traditional lessons and textbooks in exciting ways. To encourage students to go through the assigned storylines, Alexander inserts quizzes and riddles between articles to capture the students' attention.

5. CONCLUSION

The main innovation of the concepts presented in this paper is the integration of user and content in what can basically be seen as a new layer on top of digital information collections, a layer on top of the semantic web. This integration is already well known from the curating of exhibitions in museums, which always creates stories focused on important aspects, topics, or developments within a broader context. But in contrast to such manually derived and static storylines, the future Web system should provide exciting and attention-capturing narrative storylines automatically derived per query and personalized for each individual user.

In the area of business intelligence and linked open data, information extraction today plays an important role for searching, mining and managing information. In the area of cultural heritage the assignment of metadata has up to now only been performed in a descriptive sense, with the notion of general usefulness. In our vision we innovate the metadata assignment by extracting possible semantic anchors from the digital items that allow the basic anchoring within a storyline. Best matching anchors will be chosen for forming storylines with respect to the degree of narrative intelligence that a pairing within a common storyline promises. Following writing theory, this degree can be measured in several aspects like the strength of development or potential conflict.

Moreover, the inclusion of small attention capturing tasks, subtle changes to be detected, and surprising conclusions to

be drawn will allow to involve users deeper in cultural storylines than traditional content presentation. At any point in a story users are enabled to provide individual feedback about the current storyline, leave storylines for detours, or join storylines with other users.

ACKNOWLEDGEMENTS

The work presented here was supported by the Serbian Ministry of Education and Science (project III44006).

REFERENCES

[1] Barak, Miri, Herscoviz, Orit, Kaberman, Zvia, Dori, Yehudit. MOSAICA: A web-2.0basedsystem for the preservation and presentation of cultural heritage. In Computers & Education, Volume 53, Issue 3, November 2009, pp. 841–852

[2] Hyvönen, Eero, Ruotsalo, Tuukka, Häggström, Thomas, Salminen, Mirva, Junnila, Miikka, Virkkilä, Mikko, Haaramo, Mikko, Mäkelä, Eetu, Kauppinen, Tomi and Viljanen, Kim: CultureSampo-Finnish Culture on the Semantic Web: The Vision and First Results. In: Klaus

Robering (editor): Information Technology for the Virtual Museum - Museology and the Semantic Web. LIT Verlag, Berlin., November, 2007.

[3] De Jong' T, Van Der Hulst, A, The effects of graphical overviews on knowledge acquisition in hypertext, Journal of Computer Assisted Learning Volume 18, Issue 2, pages 219–231, June 2002

[4] Nelson, T. H. Literary Machines, Mindful Press, (1981).

[5] Nelson, T. H. 'Transliterature: A human- ist format for re-usable documents and media'. (2007), http://transliterature.org/.

[6] J. Surowiecki The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations, Doubleday, 2004