Can the Web turn into a Digital Library?

Hermann Maurer
Institute for Information Systems and Computer Media
Graz University of Technology, Austria
hmaurer@iicm.edu

Heimo Mueller
Institute for Pathology
Medical University of Graz
heimo.mueller@mac.com

Abstract: There is no doubt that the enormous amounts of information on the WWW are influencing how we work, live, learn and think. However, information on the WWW is in general too chaotic, not reliable enough and specific material often too difficult to locate that it cannot be considered a serious digital library. In this paper we concentrate on the question how we can retrieve reliable information from the Web, a task that is fraught with problems, but essential if the WWW is supposed to be used as serious digital library. It turns out that the use of search engines has many dangers. We will point out some of the possible ways how those dangers can be reduced and how dangerous traps can be avoided. Another approach to find useful information on the Web is to use "classical" resources of information like specialized dictionaries, lexica or encyclopaedias in electronic form, such as the Britannica. Although it seemed for a while that such resources might more or less disappear from the Web due to attempts such as Wikipedia, some to the classical encyclopaedias and specialized offerings have picked up steam again and should not be ignored. They do sometimes suffer from what we will call the "wishy-washy" syndrome. It is interesting to note that Wikipedia which is also larger than

all other encyclopaedias (at least the English version) is less afflicted by this syndrome, yet has some other serious drawbacks. We discuss how those could be avoided and present a prototype of a system that does take care of many of the problems mentioned above, hence may be a model for further undertakings in turning (part of) the Web into a useable digital library.

1. Introduction

The Web has turned into the dominant source of information. Most persons use it by employing one of the available search engines, or by going directly to a site they tend to rely on. Using one of the major search engines like Google or Bing is tempting, yet one has to be aware of a number of problems: first, often hundreds of thousands hits are presented, more than anyone will ever look at; actually, even if the search engine indicates that it has found some hundreds of thousand entries only the first few thousands can be accessed by the user: this goes little noticed since even reading beyond the first few pages with search results is more rare than it should not be; second, the reliability of information found is not at all guaranteed: it is up to the user to investigate whether results can be trusted or not and this is often almost impossible; third, most current search engines are still based on a set of words, rather than understanding natural language questions; fourth, the ranking of search results is not transparent; indeed it often depends on factors that influence the user in the wrong direction. We will examine those four points in Section 2.

Rather than using a search engine one might directly go to a specialized site. In Section 3 we discuss pros and cons of going to one of the sites that are structured in classical fashion, like a specialized collection of information on art, or animals,

or minerals, etc. We exclude Wikipedia from this set on purpose since we will discuss issues concerning Wikipedia separately in Section 4. In Section 5 we present a prototype where an attempt is made to eliminate most of the weaknesses discussed in previous sections and argue that using the ideas mentioned we could finally end up with a very large repository of reliable material that we can base our work and our judgements on. We will also mention a new kind of E-Book that is part of our prototype that may go a long way to turning large portions of the Web into a genuine digital library.

2. Some aspects of search engines

One of the most obvious problems encountered when using search engines is that the number of search results is too large to be used systematically. Further, many search results are similar to others, i.e. have an undesirably high degree of redundancy; worse, some search results are contradictory.

Concerning redundancy, it would be nice if search engines in the future would try to cluster together similar search results automatically, may be even combining results in a cluster into one more or less coherent document, so that users are only confronted with a limited number of clusters, or even better with documents representing the most important “views” on a topic.

In some isolated cases this has worked quite well as pointed out in (Wurzinger, 2010): it is shown in that paper that in some cases redundancy can be cut by 75% applying fairly simple similarity recognition algorithms as used for plagiarism detection, like in (Zaka B. et al, 2008), (Maurer H. & Kulathuramaiyer N., 2007); (Maurer H. & Zaka B., 2007) and

(Kappe F. & Maurer H. & Zaka B., 2006). To be more specific, 20 essays on 50 topics where reduced to an average of 6.3 essays per topic without loss of information. The only price paid was that not all essays were as coherent and smooth to read as the originals.

To reduce redundancy dramatically (not by 75% but by 99%) and to retain coherent essays (i.e. to construct coherent essays from lots of snippets that have been collected) is still something that sounds like science fiction today; yet it is one of the great challenges search engines are facing, even if finding solutions will still take some major breakthroughs. Yet clustering similar documents together and automatically preparing a few sentence summary of the difference of one cluster when compared to a different cluster is possible today.

Using good visualisation techniques the relation between document clusters could be shown in an impressive way: it is a pity that the publicly available large search engines are not making use of those features to any extent. At the moment, search engines are reducing the amount of information available to users mainly by ranking and by "personalizing" them. They do usually not allow to do further searches in the large set of documents located (something that would help a lot and would be easy to offer). Recent attempts in this direction are search engines such as the slash-tags in (Blekko 2010).

A major problem is the issue of ranking. Ranking algorithms are usually not publicized, giving rise to many speculations. Like, if an item A is listed before an item B is it really because it is the better hit, or is it that whoever is behind A has a better relationship with the search engine, maybe even to the extent of paying for preferred treatment? Is the sometimes heard

rumor that sites using Google Analytics are on purpose better indexed by Google as bonus for using another Google product true or not?

It is probably of interest to some readers to note how ranking can have a negative effect for them. Let us explain this in terms of a (very realistic) example. When trying to book a hotel in some city one often finds easily some booking agency that allows you to book a suitable hotel at ease. Although all kinds of information on the hotel is provided, like how to get there, amenities available etc. a valid phone number or E-mail of the hotel is often not included for obvious reasons: one does not want the customer to be able to make direct contact with the hotel.

Although this is usually not of concern, it may be, if a last minute change becomes necessary, if one is lost and wants to phone for directions, if one has to cancel or re-book, etc. Seasoned travellers overcome this barrier with a trick: they use the booking agency to locate a hotel that looks good; then they use the hotel's name for a search with a search engine. This will usually lead them again to a or the same booking agency, but now comes the trick: going some ten pages further in the search results gives a good chance to find the website owned by the hotel, with all necessary parameters. With some luck, one can even get rooms cheaper, one can certainly negotiate for some small extras that one usually would not be able to mention. ("My room should not be adjacent to the elevator shaft" might be a valid concern by persons who are disturbed in their sleep because of the audible movement of the elevator.) It is interesting (disturbing?) to notice that some agencies are starting to refuse to accept hotels for bookings if the hotels have their own homepage!

Concerning correctness, let us quote from (Wurzinger, 2010):

"We all accept that no information obtained is reliable (except if know we can trust the source of information), yet how dramatic the unreliability is can be shown with numerous examples. Searching for "boiling point of Radium" with Google two entries retrieved Aug.25, 2010 are shown in Fig.1

Fig. 1: Boiling point of Radium

One entry shows 1737 degree Centigrades, the other 1140. How should we know which one is correct?

May be life does not depend on this particular answer. However, consider a case we have been confronted with when we picked a type of wild mushrooms recently that we could definitely identify as "Echter Ritterling" (Gruenling). When we wanted to check if it was edible or not we found five entries on the first search page, three telling us that it is a delicate edible mushroom, one informing us that it is deadly poisonous and one simply that it is poisonous!"

How is it possible that even in what seems reliable sources such wild discrepancies and contradictions occur? There are two main reasons: one, often definitions differ: if you look for the "largest cave in Canada" do you mean largest by length, by volume, by height, or by which other criteria? If you want

to know the height of a mountain on the moon do you mean the relative height compared to the deepest point “near” it, or do you mean the height above a hypothetical sphere giving the average height of the moon (as we sort of do on Earth when we compare heights to sea-level); second, the discrepancies are often due to the fact that information comes from different times: it is very unfortunate that documents on the Web are rarely dated!

This, by the way, is the reason for the different judgement of the edibility of the mushroom mentioned above: it was eaten without known side effect for thousands of years; in 2002 suddenly two deaths seemed to be linked to the consumption of a dish made out of the mushroom. Whether the death of two persons makes a substance poisonous is very doubtful in itself: after all, we have people with peanut, fish, milk etc. allergies (that can be deadly) yet we do not consider either peanut, nor fish, nor milk poisonous! Anyway, the isolated cases mentioned have caused newer entries on the mushroom to call it poisonous.

What can be learnt from this: (a) if various definitions are possible, the documents should make this clear: this is NOT a job for search engines but for authors of documents; (b) all documents should be clearly dated; (c) the date should be considered as part of the ranking algorithm in search engines. Note that if I search for a meeting, an algorithm, a tool, etc., I am likely to be more interested in more recent ones than in ones ten years or further back!

There is even a darker side to it. We often warn (young) people today that they should not put up too much personal information on social sites like Facebook, because it can be used against them sometimes in surprising ways. But this is

not a problem of Facebook alone: if you have ever left a picture of yourself with a girl/boy friend on a photo site, that site may not allow you to ever remove it again. Yet that picture may prove an embarrassment when you have married someone else at a later stage. Although some search engines do allow to ignore entries that are a number of years old (Google has such a parameter), not all search engines allow the deletion of old entries and those that do, do not advertise it much, so that the average user is not aware of them.

There is one other issue concerning search engines: web search engines traditionally work with a group of input words, connected by “or”, “and” or “not”. A more linguistic approach (natural language queries) was already taken in (Brockhaus 2006). Natural language queries have been allowed in this electronic dictionary (which in its full form remains fairly expensive, unfortunately) now for over 5 years. One of the easy tricks was to observe the word at the beginning of the query: “Who” is clearly asking for a person, “Where” for a location, “Why” for an explanation“, etc. By delving more deeply into language understanding the linguistic group in Saarbruecken, Germany and the implementation group in Graz, Austria, came up with fairly decent results.

Despite this fact most search engines are still based on words, albeit more and more cleverly. Inputting “Who was the inventor of the toothbrush?” is (in Google) turned into “Who invented the toothbrush” (i.e., some linguistic analysis is employed). The Google result gives what it seems is a reasonable answer (“No exact date known…” but then continues to give lots of additional information, like that the first mass production was started by a William Addis in 1770). That linguistic analysis is taking place rather than just using

the important words is evident: when inputting “toothbrush inventor” into Google we get somewhat different answers.

The search engine Bing with input “Who was the inventor of the toothbrush?” finds William Addis in 1770, but gives very different answer on “Who invented the toothbrush”. This seems to indicate that less language analysis is applied in Bing! The results above show that there are (mushroom case!) not just discrepancies within a search engine, but results also differ a lot depending on how a question is formulated. Further, discrepancies between different search engines can be quite serious! Often there is no easy way to determine who is right. In essence, one can trust the result of a query only if one can trust the source.

The question “Who was the physicist born in Vienna and died in Italy?” does not work well with Google. The reason is clear if one looks at the search results: the search is text based, so Google finds all Vienna physicists. Since Schrödinger worked (but did not die) in Italy at some stage his name pops up quite early, i.e. the very “die” is ignored. To be fair: Boltzmann as the correct answer to the question posed is also found, but on a later page. Bing actually finds Boltzmann better than does Google, and provides interesting further information, yet its search is also clearly word-based. In the prototype system which we will briefly mention in the last Section of this paper, since documents have meta-date associated with them, Boltzmann is found immediately. Since general search engines cannot assume the presence of systematic meta-data, they either have to work with words or have to dig deeper into natural language understanding! But even if they do, how can we trust the result (see toothbrush example).

Summarizing this section: It is apparent that the major search engines do not employ deep language-analysis tools, are generally not good in allowing to narrow down large query sets, do not seriously try to reduce redundancy and do not taking dates ("timestamps") sufficiently into account: hence, much remains to be done to satisfy users. Due to the importance of search engines further progress can be expected, however. From a user point of view it is important that Google has now with Bing, with the completely different Wolfram-Alpha, with slash/dot and others some competitors. It will not be able to rest on its laurels. The authors wonder when the first search engine will become public that only searches sites with semantic data and a guaranteed level of reliability: it could turn the Web form valuable but doubtful resource into something of much greater value than is offered to us today.

3. Special purpose encyclopaedias and dictionaries

There are thousands of free encyclopaedias and dictionaries on the Web. Some give only limited access free of charge but ask for payment for "premium use" or such. Some (typically medical encyclopaedias) are only available for closed user groups (certified physicians). One of the first such medical Web encyclopaedias was (Info-MedAustria 1999), offered by Bohmann Company Vienna for a number of years free of charge. However, like even much larger encyclopaedias (e.g. in Germany Brockhaus and Meyer, the latter available online free of charge for many years) most universal electronic encyclopaedias have disappeared or are only offering limited information for free, due to the pressure of free information, particularly from Wikipedia. For a while it seemed that Britannica would also give up completely, yet the current (Encyclopaedia Britannica, 2010) electronic premium version is quite remarkable although only parts are free.

However, the (List of encyclopaedias 2010) shows that both general purpose (“universal”) free encyclopaedias still do exist and that there is also a very large number of specialized encyclopaedias and dictionaries.

Clearly, Wikipedia has eroded the commercial basis of general purpose high quality encyclopaedias to some extent, at least for the time being. While this has been deplored by some critics like (Keen 2007), claiming that this is the beginning of a road to mediocre material a vast number of persons believe that Wikipedia is such a valuable and also high quality tool that the demise of commercial products is quite acceptable.

Although the authors of this paper have some points of criticism concerning Wikipedia they also are critical of traditional encyclopaedias for a reason that is often overlooked: the typical encyclopaedia of the 20 th century was an alphabetic arrangement of topics in an “objective “ way, thus reporting the “truth” about an event, a person, an idea, whatever.

We believe such a concept is basically flawed. Everyone agrees that if we look at a material object (as sculpture, a mountain, a house, any object you can think of) we can get a proper impression of the object only by seeing it from different views. This does also apply to non-material objects such as ideas, or personalities, etc., yet in general this is less explicitly acknowledged. But if we can only understand a complex person, a complex idea or a deep concept by getting very much opposing views a single “compromise” or “wishy-washy” description of the issue will not be helpful. What is needed are a number of different reports on the same subject with pointedly different views.

Traditional encyclopaedias have tried to live with this by having pro and contra views, yet there was always an author or team of authors behind each entry with a certain point of view, colouring the presentation. It is our belief that in future collections of encyclopaedic type this has to be avoided.

In a similar vain it has to be avoided that encyclopaedias present an issue form a single point in time, since this often hides important issues.

Let us explain this with one simple example. In the eighties of the last century Europeans were so worried about the extinction of interesting varieties of tropical wood that the import of certain types and objects made thereof was forbidden. A typical European encyclopaedia of 1985 would report this fact with some pride, showing the concern of Europe for maintaining variety in nature. However, since the import of tropical wood was not possible any more that type of wood lost its inherent value. Because of this, large forests of threatened species of tropical wood were burnt down to make room for rice fields that would yield at least a bit for the local population. Thus, the well-intended effort to protect tropical wood produced exactly the opposite of the desired effect. And a European encyclopaedia of 2002 reported (a) that certain types of tropical wood are endangered and (b) that local population was continuing to destroy it. The reason for this was (often) not mentioned.

This leads us to a critical analysis of Wikipedia. It turns out that Wikipedia might well be a step in the right direction, but that some changes would indeed increase its value still further.

4. Wikipedia

Wikipedia is certainly one of the big successes of the "Wisdom of the Crowd" paradigm as described in (Surowiecky 2005). According to the (Wikipedia Foundation 2010) some 400 million persons are using Wikipedia nowadays.

Over time, many weaknesses have been pointed out: in addition to inadvertent errors there have been cases of deliberate spreading of false information including defamation of persons, blown out of proportion description by paid or unpaid fans of some notion or person, hidden advertisements, or discrepancies in numbers reported: In some report on some country the population of city A would be mentioned at a number x, while the report dedicated to city A would mention a number y, potentially because census data from different time periods had been used. Another troublesome aspect is that the same event might occupy much space in some language version of Wikipedia, but may be quite short in other languages versions. Worse, the inventor of some device D might be person A in one country, and Person B in another country.

However, having said all this it is also clear that the average quality of contributions is quite good, that the control of many readers is working to a high degree. It also must be understood that editing, censorship and correction procedures can be quite different between various language versions of Wikipedia, and that rules are not carved in stone, but keep being improved. Here is an example: in the English Wikipedia it was initially possible to write completely anonymously. After a famous slander case this was given up.

We quote from Wikipeida itself:

*"The **Seigenthaler incident** was a series of events that began in May 2005 with the anonymous posting of a hoax article in the online encyclopedia Wikipedia about John Seigenthaler, a well-known American journalist. The post fabricated statements that Seigenthaler had been a suspect in the assassinations of U.S. President John F. Kennedy and Attorney General Robert F. Kennedy. The 78-year-old Seigenthaler, who had been a friend and aide to Robert Kennedy and a pallbearer at his funeral, characterized the Wikipedia entry about him as "Internet character assassination".*

The hoax was not discovered and corrected for more than four months, after which Seigenthaler wrote about his experience in USA Today. The incident raised questions about the reliability of Wikipedia and other websites with user-generated content that lack the legal accountability of traditional newspapers and published materials.[3] After the incident, Wikipedia co-founder Jimmy Wales stated that the encyclopedia had barred unregistered users from creating new articles."

Thus, today, at least some versions of Wikipedia do not allow to write contributions unless some screening of the writer has taken place.

We have criticised that traditional encyclopaedias have only one entry for even the most complex topic, even if that topic cannot be presented by someone "claiming to have the truth" but only by presenting different points of view. Wikipedia is doing the same, yet it does allow to examine the thread of discussion that has lead to the current result, thus giving much more insight than is the case in traditional presentations.

However, we feel that a number of crucial improvements are still missing to make Wikipedia to what it is now trying to be: the ultimate source of reliable information on any subject whatsoever. To prove our point we are in the process of establishing an undertaking where we try to reproduce what is good in Wikipedia, yet where the introduction of a number of additional features will help in achieving a new kind of quality. It would be crazy to try to do this on the scale of Wikipedia, so we have restricted the scale dramatically by only collecting information on a single small country and issues involving it.

5. The Prototype

The system (technically a JSP WIKI with many plug-ins) has been officially in operation since October 2009. It covers only "Austriaca", i.e. items that involve Austria or Austrians in some way. At the time of writing the prototype that can be tested at www.austria-forum.org comprises some 170.000 "objects", an object defined as text-file, picture, panoramic-, audio- or video-file. Completion of the desired functionality and a first solid foundation information-wise is planned for first half of 2013, at which point Austria-Forum will contain some million objects.

It is important to understand the main differences between Wikipedia and the Austria-Forum:

(i) In the Austria-Forum the domain is restricted to Austriaca as described; it emphasizes information that has a high degree of stability. Thus, a biography of a former poet or the description of an event in history is well suited, a biography of a rising new star in politics is acceptable, results of the rescue

of the Chilenian miners in October 2010 (despite the fact that Austrians were involved in it in a critical way) or sports events of the last month have no place in Austria-Forum: “news type” information is left to the media. One reason is to avoid competition, the other is pragmatism: we cannot muster the resources to also cover all those items, and the third is maintainability. Once all important historic facts about Austria, all mountains, flowers, animals, minerals, stamps, coins, etc. etc. are collected maintenance is comparatively easy: the biography of a poet like Stifter needs no updates, nor does the description of building the first road over some alpine pass; and although flowers and animals might change a bit, the emphasis is a bit, i.e. keeping this up to date is a manageable effort, particularly since all news reports by the Austrian Press Ageny (APA) are analyzed automatically and lead to alerts if applicable.

(ii) The Austria-Forum distinguishes between approved main entries and general entries in the community section. In the latter, rules similar to Wikipedia apply, yet contributions can be upgraded and moved to the main entries section if the editorial board so decides. Main entries have an author who has been screened and whose CV is available to users, so that they have background information on who is writing what. Main entries are also taken from books and archives: in each case the aim is to provide a clear source.

(iii) It is also attempted to also associate a date with each entry: not the last date of a minor update, but the date when the main entry was created. Note that this has two aspects: we hope to be able to e.g. show pictures with sliders that allow to view the change of a city, a glacier, a river, or other items, over time. We even hope to have a slider showing different points of view on various subjects. The “time-stamp

paradigm" also means that if someone wants to do a major edit to an approved entry, this is not considered desirable. Rather, a new entry with the same name is created. Thus, ideally, you will not find an entry on "nuclear energy" but a sequence of entries like "History of nuclear energy in Austria", "Why nuclear energy is important", "Why nuclear energy is dangerous", etc: pointed and provocative contributions about nuclear energy from various points of view. Ideally, you should not find a picture of our city Graz, or an essay about Graz, but photos of Graz at various times, and essays describing Graz at various times. Thus, quite in contrast to Wikipedia, an essay on Graz should no be updated, but retained as time capsule, and another time capsule added later.

(iv) Since contributions have a source and a date, it is possible to quote them in scientific contributions, an open issue with Wikipedia contributions. Austria-Forum is interactive in as much as anyone can add comments to a contribution: many comments may lead some editor to even write a new version of the essay, leaving the old essay with all its idiosynchrasies intact. Other communication facilities are also provided to hopefully strengthen the spirit of community, and visualization tools are being developed to follow the doctrine of information consolidation explained in (Wurzinger, 2010).

(v) We do not believe in providing a single encyclopaedia, but a substantial set of them covering various topics. The reason is that the search in Austria-Forum allows to not only be narrowed down to one area (a very desirable feature) but to use available metadata. Note that Fig. 2 shows a form filled out with entries typical for a biography and indeed the search finds immediately the person at issue (Boltzmann). But the form (metadata) required to find a lake, a building, a flower, etc. would clearly have to look very different.

Fig. 2: Searching in Austria-Forum using meta-data.

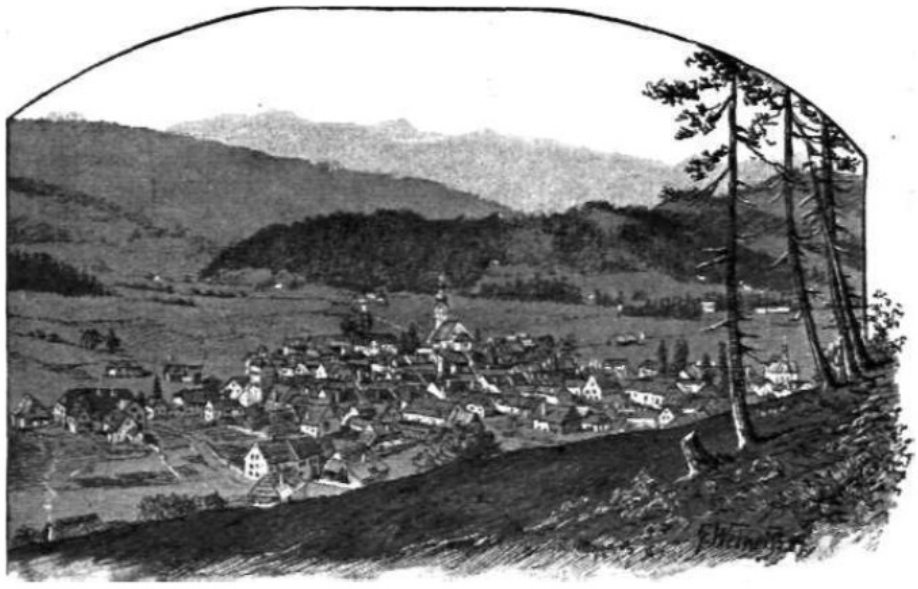
(vi) We have added to the Austria-Forum a new kind of object akin to an e-Book. Those books are stored in a kind of bookshelf: the first two rows with some historical books are shown in Fig. 3.

Fig. 3: Part of one of the book-shelves

However, not only does the bookshelf look similar to a real bookshelf, also the books themselves behave more like real books than e.g. PDF-files, yet they do offer advantages like searches as one would expect from electronic substances. Books are also heavily cross linked with other information within the Austria-Forum and beyond, allowing trips through time. Here is an example:

Opening the first book on the top row of Fig. 3 and going to the village of Obdach shows Obdach around 1895 (Fig. 4).

in einem adeligen Hause, die Falb bald hierauf antrat, verschaffte ihm eine unabhängige Existenz. Falb gründete in dieser Zeit die populär-astronomische Zeitschrift „Sirius", welche er bis 1877 dirigierte. Bald darauf publicierte er das Werk: „Grundzüge einer Theorie der Erdbeben und Vulcanausbrüche", worin er zum erstenmale seine Theorie über den Einfluss der Mondnähe auf die Erdbeben-Erscheinungen aufstellt. In den



Obdach.

nächsten Jahren finden wir Falb an den Sternwarten und Hochschulen zu Wien und Prag, woselbst er die Vorlesungen Hochstetters, Hausteins, Dureges und Mays über Geologie, Physik, Mathematik, Planetenberechnung etc. frequentierte. F. trat hierauf zum Protestantismus über. Die Erdbeben von Belluno 1873 und der Ausbruch des Ätna 1874, welche Ereignisse Falb mit großer Bestimmtheit vorausgesagt hatte, verbreiteten rasch den Ruf Falbs und bestimmten denselben, seine Theorie wie folgt zu präcisieren:

Fig. 4: Obdach some 115 years ago

The small elliptical icon on the left indicates that there is a panoramic view of today of the same village available. A click at that icon results in the picture shown in Fig. 5:

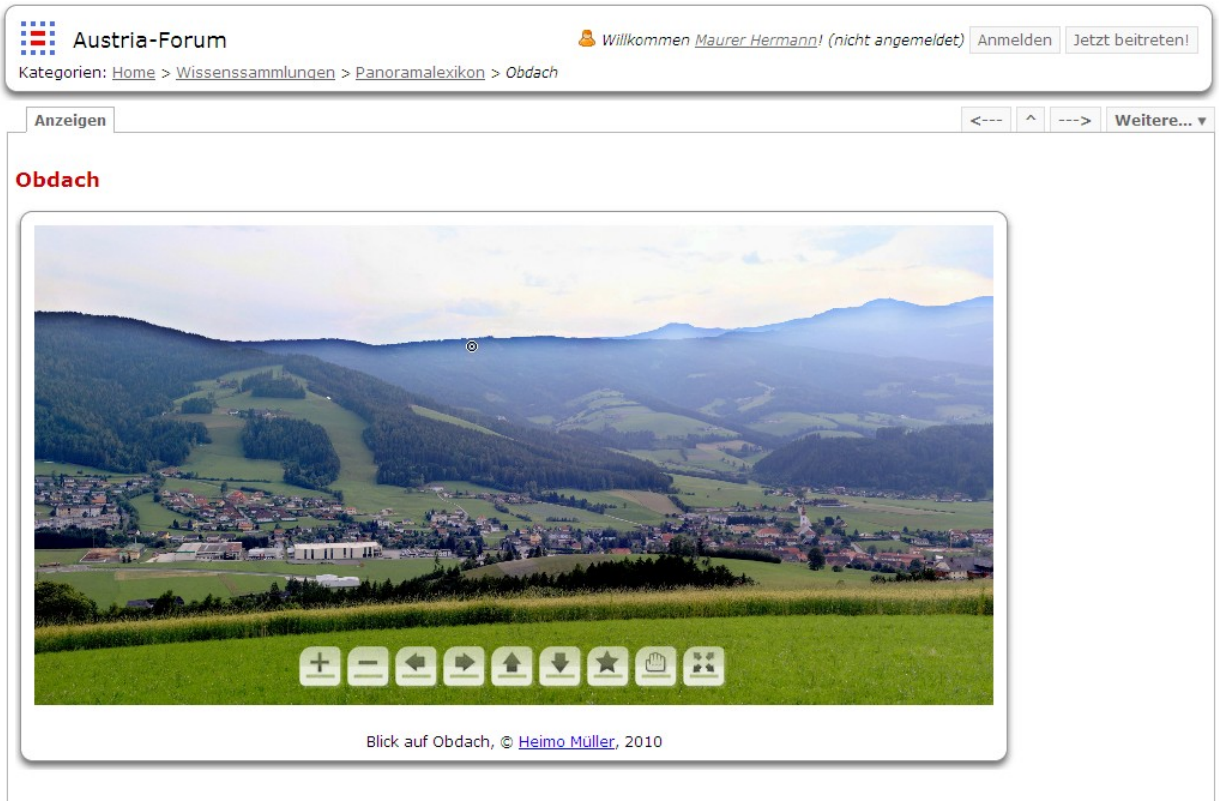


Fig. 5: A panoramic view of the village of Obdach

The panoramic view does allow, as one would expect, to zoom in or out, to pan, and to tilt the picture.

Thus, Austria-forum allows to jump backward and forward in time, a feature that will be dramatically expanded over the next two years.

Allowing users to add their personal (or public) remarks and links will turn this new kind of object (Mueller, H. & Maurer, H., 2010) into a valuable tool and into a solid basis for well-grounded discussions. Most important, it will have much more the feeling of being a genuine digital library than most

previous effort, where the “emotional book feeling” has been largely ingnored.

For readers eager to try this out look at e.g. http://www.austria-lexikon.at/ebook/bookshelf/ and click at the first book on the shelf!

6. Conclusion

It is an accepted fact that we are going to use material on the web more and more. In this paper we have analyzed a number of ways how to retrieve reliable information. We have argued that no approach is without its flaws. We have further explained a substantial prototype that is currently developed that we hope will be a major contribution to handling the flood of information in what will look more an more like a physical library, yet comes with all functionality expected from an electronic corpus.

References:

Encyclopaedia Britannica, (2010) Electronic Version, http://www.britannica.com/ last checked Nov. 30, 2010

Wurzinger G. (2010). Data consolidation in large bodies of information; JUCS vol. 16 (to appear)

Zaka B., Kulathuramayer N., Maurer H., Balke W.-T. (2008). Topic-Centered Aggregation of Presentations for Learning Object Repurposing; Proc. of E-Learn 2008, 3335-3342

H. Maurer, N. Kulathuramaiyer (2007). Fighting plagiarism and IPR violation: why is it so important?; Information Services & Use, vol.27, no. 4, IOS Press, 185-191

Maurer H. & Zaka B. (2007) Plagiarism - a problem and how to fight it; Proc. of ED-MEDIA 2007, 4451-4458

Kappe F. & Maurer H. & Zaka B. (2006). Plagiarism - A Survey; J.UCS 12, 8 (2006)

Blekko (2010). http://blekko.com last checked Nov. 23, 2010

Brockhaus (2006) Der elektronsiche Brockhaus, Mannheim, Germany (2006)

Encyclopedias (2010). http://www.encyclopedia.com/ last checked November 25, 2010

Keen A. (2007). the cult of the amateur. Double Day 2007.

Surowiecki J. (2005). The wisdom of the crowds. Achor Books 2006.

Wikipedia Foundation (2010). http://wikimediafoundation.org/wiki/Home last checked November 27, 2010

InfoMed Austria (1999). Medizinische Informationen, Bohmann, Vienna

Mueller, H. & Maurer, H. (2010). A new approach for e-books for teaching and learning (to appear)

The Austrian Ecnyclopaedia (1995). AEIOU. http://www.aeiou.at, last visited November 30, 2010.

List of Encyclopaedias(2010). http://en.wikipedia.org/wiki/List_of_online_encyclopedias, last visited Nov. 30, 2010.