Experiences Based on a Major Information Server

Delilovic, Namik; Ebner, Martin; Maurer, Hermann; and Zaka, Bilal

Abstract: *In this paper we describe lessons we have learnt from building up a webserver with reliable information over a period of ten years. We show how we have tried to overcome often encountered weaknesses and thus encourage the community to follow our suggestions. We also point out that much research and development work remains to be carried out.*

Index Terms: Intelligent Web server; digital library; flexibility in Web servers; Austria-Forum

1. INTRODUCTION

A team involving the authors has been involved in developing a substantial "quality-controlled information server" called Austria-Forum [1], [2], [8], [33] concerning information of interest to Austrians since 2009. The aim of this paper is to correct a misunderstanding: Despite a large number of research results and recommendations for practical use, austria-forum.org, has been misunderstood by many as just "another collection of partially interesting digital information".

This is a serious mistake that needs to be corrected. Of course, the server is offering interesting information on a huge variety of topics, and has gained some respect for this. But the collection of this material has a second reason: Large quantities of information and users are needed to conduct research on how to handle huge amounts of data and users. One approach is to study how to represent reliable information and how to make access to it useful and easy for users. Another one is how to let users discuss important questions. A third one is the problem how to make sure that the material is not just attractive to one age group but to young and old, from school student to persons already having left active life for some time. And a fourth and may be most important in our times: to make sure it allows to differentiate between fake news and real news, between half-truth and objective reality. Much substantial research in those areas seems to have sometimes gone unnoticed. It is the aim of this paper to report on some of the research and development achievements that are influencing the Web, and are needed so that the Web is not just a new way into chaos [7].

2. NUMBERS

The Austria-Forum has some 3 million different (mainly German speaking) users at the time of writing (October 2019), over 1.1 million media objects and close to six million page views per year. It has been successful in showing how to improve critical areas of Web presentations and is attempting to be a kind of role-model for similar attempts by further research and development. Here is an overview of the structure of the rest of the paper:

Section 3: How to avoid the presentation of fake information by using a repository of carefully documented reliable information.

Section 4: What is the role of digitized books?

Section 5: How can one assure meaningful answers from search-engines.

Section 6: How can the software help to provide serious communication between well-defined groups and avoid outpouring of trivialities and fake information as is typical for some social media?

Section 7: How to fight fake news.

Section 8: How to reach all age groups.

Section 9: Conclusion.

All ideas and developments presented in the sections that follow are by-products of research and development of a group supported by a number of universities and sponsors. The development is supported by numerous institutions from within Austria and other countries and a board of some 200 honorary editors and researchers.

Disclaimer: Many of the results have been published in research papers, theses, keynote presentations and the like before. We cannot

discuss all issues involved and are leaving some to be studied in detail using the references provided.

3. QUALITY CONTROL OF CONTRIBUTIONS IN AUSTRIA-FORUM

One important way to assure high quality is to make sure that contributions (that are in searches presented, are written by well-known persons, usually with a CV showing their competence. This is also much supported by big efforts like [33].

Users can register (even with an arbitrary pen-name) in a way so that only the system knows the E-Mail address, but still they can contact authors without revealing their identity or E-Mail. The system allows authors to answer, without needing to know the identity of the user who contacted them.

There is a second alternative to provide feedback: Each Web-page in Austria-Forum has a green "feed-back button" [13] that allows to send, without any registration, messages to the administrator, who can act on them, including forwarding them to one of the editors, if this seems appropriate.

These two facts, i.e. that authors commit themselves to what they write rather than to hide in an anonymous group and that two ways of communication for questions, comments etc. are provided, assure a much higher standard of contributions than it is found in anonymous systems where users have no easy way to react and do not know in which direction an author or author collective is trying to color information according to own beliefs, as is sometimes argued about Wikipedia [34}. A second argument that reduces the value of Wikipedia is the fact that different language versions present in quite a few cases very opposing views.

However, in addition to above, Austria-Forum provides another important way to assure quality: It contains thousands of online readable digitized books [11], [13]. Links from any Web-page to any point in a book (and conversely, and from a book page also to another book page) are possible. Hence, an author writing about some matter can make a link to a section of a book dealing with that matter, thus adding to his credibility the credibility of the author and publisher of the linked book. Even if authors (or some member of the Austria-Forum team) have not made links to some book sections (a process to be supported by providing

suggestions for links automatically) users can consult books available in Austria-Forum to check issues themselves.

4. WHAT IS THE ROLE OF DIGITIZED BOOKS IN AUSTRIA-FORUM

As mentioned in the previous section books can be used by linking from a contribution to a relevant spot in a book, or users can check up on statements made on some Web-pages using the books provided.

Since almost all books are full-text searchable, this is often easy. Note that books are readable online (but usually not downloadable). Due to the fact that links from Web-pages lead users to "stumble" across books (of which they might not even have been aware of) many publishers see the incorporation of their books into the digitized library of Austria-Forum not as a competition to selling the books in print or as e-Books, but rather as showing the existence of the books to the public.

Digitized books in Austria-Forum offer much more than usual collection of digitized books, or individual e-Books, or books on Kindle and such. Usually, digitized libraries offer books in "complete isolation", i.e. neither with links from the Web to a spot in a book, nor supporting a section of a book by e.g. providing additional pictures, or linking to audio or video-clips, or even to some interactive material, potentially turning the book even into an E-Learning system. Even more is true: with new tools, it is intended to link sections of different books (and material on the Web) dealing with similar topics together on demand, thus allowing users to look at a topic in various sources and from various points of view easily, rather than having to laboriously search in all kinds of collections for relevant contributions.

In addition, a new generation of digitized books we are developing is almost ready: It will be compatible with the International Image Interoperability Framework™ (IIIF) [30]. This is expected by most major libraries since it offers many additional features as described in [5].

5. HOW CAN ONE ASSURE MEANINGFUL ANSWERS FROM SEARCH-ENGINES

There are three different important issues to deal with:

(a) How one can easily find what one wants
(b) How answers are to be interpreted

(c) How one can assure that answers obtained are correct

Concerning aspect (a), search-engines have evolved quite a bit over the last years. If ten years ago you entered "Enzian" (the German word for gentian, a flower often associated with the mountains in Europe) into any good search engine you got a long list of the many types of the flower Enzian, but also the Schnaps (brandy) called Enzian, the rocket Enzian built during world War II and possibly all kinds of "Lederhosen" or hotels or restaurants using the word to advertise that they are close to alpine traditions.

We tried to avoid such a multitude of answers by structuring Austria-Forum into "categories" and "subcategories". Thus, when you look at the entry page of Austria-Forum and you are interested in the flower Enzian, you probably should not immediately use the search function provided, but first choose the category "Nature" and then the subcategory "Flora", to narrow the scope of your search. This idea of narrowing the scope of search was picked up in various ways by other systems. In Wikipedia typing a word that can mean more than one object results in showing you alternatives from which you then choose. Google eventually ended up with two parallel approaches: either you can enter a full query sentence describing in some detail what you are looking for or you will get a few hits, but then a set of categories. When you e.g. search for London, you get, in addition to some pictures and links a list of suggestions:

Searches related to London

London **city**
London **uk**
London **attractions**
London **wiki**
London **population**
London **map**
greater London
London **facts**

It is not clear how queries will develop. Some believe that text/language understanding will get good enough to allow very complex queries. We tend to believe that a multi-step process might be better in the end. Typically, the query "London" might result in a few obvious hits (Google-like), then presenting alternatives, after you choose further alternatives, and so on, until the system has determined in a dialogue with you what you are really interested in. Much information on this can be found in [14] and [17].

It is worth mentioning that the categories we use have still another reason: Material in different categories may well come from different sources. The entries in "Biographies" are much based on the huge collection of biographies the Austrian Academy of Science [31] is offering, yet those biographies are very technical and not suitable for the general public for which the Austria-Forum is intended. The category "culture" (Kultur) is really a collection of various special topic encyclopedias compiled under the supervision of one of the editors, the category "geography" is based on reliable data from different e.g. UN servers, etc. The category "pictures" (Bilder) forced us to merge a number of reliable data-bases. Since title and metadata may differ for the same picture, and the pictures themselves may be a bit distorted, this is not a trivial effort [3], [16]. Encouraged by chatbots and "automatic reporting", we have also applied this to some extent to geography [15] by using interactive approaches and producing graphs on demand [32].

Aspect (b), how answers are to be interpreted, is often not really understood. However, consider a query like "Number of Nobel Prize winners in the UK". Most search engines will give you a number, but leave you at a loss, since you don't know: Is this the number of Nobel Prize winners born in the UK? Or is it the number of Nobel Prize winners currently living in the UK. Or is it the number of Nobel Prize winners who got the Nobel Prize for research they did in the UK, etc. We believe that most search engines ignore the fact that answers without the definition on what the answers are based on, are fairly useless. We have tried always to provide this information in Austria-Forum, particularly in our interactive section on geographic information.

Aspect (c), how does one know if an answer obtained is correct, is clearly very important but also quite hard to answer. After all, more than one well-founded opinion may exist concerning a topic. Thus, in general, the best one can do is to offer a set of answers as explained at the end of Section 4. However, if answers are numbers, one can try to do better. We have checked certain geographic facts in Austria-Forum by using a number of data-bases: if all give the same value: fine. If they differ, the best we can do is to show the difference and point to the sources the numbers come from. So, if you want to know the area of France, we cannot tell you exactly, but we just tell you this:

Factbook: 643801
DBpedia: 674843
Geoname: 547030

Infoplease: 547030
Britannica: 543965
Wolfram: 551500

Here is an attempted explanation for the differences: For France, Britannica gives the smallest area. This agrees exactly with the area in the largest German Encyclopedia Brockhaus (2014): It follows the French Land register data that excludes lakes, ponds and glaciers larger than 1 km² and the estuaries of rivers. (This is, by the way, very much in contrast to how the figures are arrived e.g. for Finland that not only includes freshwater lakes like many databases do, but also ocean channels!). The situation of France is particularly complicated also due to overseas departments (by law, genuine parts of France). But should their area be counted? Or how about still more subtle cases like New Caledonia or French Polynesia, with special agreements with France? Or the French Antarctic section that is claimed by France but never accepted in the Antarctic treaty?

Or, not so relevant for France but very much for low lying islands, do you measure the size at low tide or high tide? How about political contentious areas: is the Crimean Peninsula now part of Russia or Ukraine?

What we have briefly mentioned concerning the area of countries applies to many other facts, be it the population of some city, the number of mountains in a country, etc. Some more on this is found in [14].

The idea and results in this section show very clearly that much of further research will be required to solve most issues satisfactorily, but we are proud to say that we have successfully begun to tackle some of them.

6. THE ROLE OF THE UNDERLYING SERVER AND COMMUNICATION FACILITIES

Section 5 has made it clear that search facilities on large servers or sets of servers still need much improvement. This certainly includes Austria-Forum; yet we want to mention a number of issues that we have pursued or are pursuing.

The fact that our information is structured in categories, subcategories etc. makes searching a bit easier, particularly because we allow categories to overlap. Thus, e.g. essays on mobility can also be found under traffic, but only one physical copy exists.

We allow full-text search in each (sub) category: This is essential, since a full-text search over the whole server is unrealistic. The full-text search for digitized books is restricted to one book, but a feature to dynamically define a set of books for full-text search is in preparation.

In a contribution, linking to others is often desirable. Yet linking a word to some other page, just on the basis of the fact that an entry exists for that word, yields a multitude of links on every page, destroying the appearance and readability. We have tried to tackle this problem in a number of ways: A link for a word is only created when the destination is more than a definition of that word; often, a link to one entry does not make sense, hence we have collected contributions into "topics" ("Themen"), as a first try of collecting all relevant contributions together as described at the end of Section 4, and we can link to such collections.

For user-friendliness we allow to initiate a search in Austria-Forum for a word by just double-clicking at it.

Searches (if not full-text) are carried out on the basis of the URL of the contribution, its title, its major headings and associated meta-data. Meta-data, at the moment, is mainly compiled manually and non-structured, what is a serious problem whose solution we have been working on: basically, for each Web-page, meta-data based on textual analysis should be generated and added, or at least proposed for addition. The addition of metadata/keywords for pages of digitized books is planned. This will allow additional search facilities within books and it will also make it easier to find books dealing with a topic of interest.

As has been explained before, registered users can contact authors of contributions. They can also add comments visible to the public to each contribution: Such comments are always verified by the administrative team of Austria-Forum to avoid misuse, but are also a simple tool for an online discussion that we will extend to more sophisticated discussion facilities, often only for well-defined user-groups in [5].

There are other aspects of communication among users and, also, between users and the administration. For this, it is crucial that the identity of all contributors is assured. An aspect of the reliability as addressed in [14].

It is remarkable that as early as 2007 the question was raised whether the Web / Internet,

which are taking over more and more of our cognitive work, are not making us stupid, see [10]. In [4]. A reasonably positive answer is given, yet if one looks at [7], a contribution written by one of the Internet pioneers, it appears that developments have to be observed carefully.

There is one aspect that has not received the attention it deserves. A link in a document is like a "goto" in a programming language, leading to a new place without informing if and when to return to the place where one came from. Already in 1987 this problem was discussed [18]: Ted Nelson suggested "transclusions" rather than links (in programming terminology procedure calls rather than gotos) since he and others [19] tried to convince the community to not use gotos, i.e. links, but few took up the idea [6]: In Austria-Forum a tool we call "InsertPage" does do, however, most of what is expected.

7. How to Fight Fake News and Statements that are Half True.

The comprehensive report [12] has at the beginning a statement that describes the situation very well. It essentially says that we are in danger of developing from a rational (fact) based society into an emotion (feeling) based society simply because one cannot trust facts found on the web. For every statement one can find counter arguments, sometimes solid ones, sometimes on purpose subtly wrong statements, motivated by either political or economic interests.

Hence [12] recommends to have servers in every language with reliable information, one of the attempts of Austria-Forum, but also an attempt by others. Together with the team of the Great Norwegian Encyclopedia [33] the Directorate-General for Parliamentary Research Services of the EU parliament organized a meeting of representatives of all countries involved in building such reliable servers in Brussels Oct. 9-11. The exchange of information between a total of 17 such efforts proved very valuable. It made clear that closer collaboration will be essential to cover all relevant topics. Notice that there are two kinds of major topics of concern to all beyond the usually expected information of mainly national interest.

One set are important developments from energy generation to mobility, from climate change to protecting the environment, just to mention some examples. It is really essential that if one e.g. wants to find out "what is the future of hydrogen energy" (in general or for mobility), one can find a reliable contribution on this topic. This clearly applies to a large variety of topics. It is our feeling that this can only be handled by having groups of experts for specific topics assigned to some country, the outcome translated into all languages (probably semi-automatically with some polishing afterwards).

Another set is information on political decisions on a national or European level. It is unacceptable, as it happens today, that after some (parliamentary) decision subtly modified versions (by sloppiness or on purpose) are distributed over social networks, influencing many receivers without them being even aware that they are manipulated. We believe the proper way to do this is to set up a discussion forum for each such topic, let everyone voice their opinion, but extract from the minutes of the meetings the exact decision that was taken, and if the vote was not anonymous maybe even showing who voted in which direction. In the soon to be released software [5] this facility will be available.

8: How to Reach all Age Groups

Running servers with reliable information is important for all age groups. Yet many servers tend, by the large amount of textual material, or by their design, to appeal more to one age group than to another. It has to be accepted, like it or not, that reading long coherent pieces of texts or full books is done significantly less by younger generations that grew up with smart phones and information from SMS, from social media, from YouTube, etc. Thus, servers will have to be designed to appeal not just to one age group, but to all, i.e. incorporating more pictures, clips and less text in some version and more traditional ways of information in others. Or to put it more generally: whether servers or digital libraries, new developments of IT technology and their effects have to be taken care of.

Encouraging active participation in discussions, even competition (e.g. on the level of schools or classes), maybe quizzes with even some small rewards, or interactive experiments that can be carried out, will appeal to some, may even turn the traditional information server into an instrument supporting learning [9], [20].

9. CONCLUSION

We have tried to argue in this paper that a general knowledge server, offering high quality information, has to support a number of ways to assure such quality; one of the most important ones is to add digitized books with very flexible ways of linking them to other material and to provide functions for feedback and potentially moderated discussion groups.

It should also incorporate different ways of accessing material to make it attractive to more than one age groups and replace the paradigm of only retrieving static web pages by accessing data in such a way that reports, as the user desires, are generated dynamically. We have implemented much of this in Austria-Forum [1] as a prototype and are about to roll out an advanced version (particularly concerning digitized books).

We believe that such a server is not just of interest for the public in general and a valuable tool for teaching and learning, but can also help much in revealing fake news as such and give fair presentations of all the many global problems facing mankind, hopefully in various modes corresponding to different groups of users.

REFERENCES

[1] Maurer, Hermann: Austria-Forum and Beyond; IPSI BgD Journal, Transactions on Internet Research; (2014), vol. 10, no. 2 (ISSN 1820-4503), 3 - 8

[2] Kulathuramaier, Nara; Maurer, Hermann: A Survey of Communications and Collaborative Web Technologies; CIT- Journal of Computing and Information Technology vol. 23, no.1 (2015), 1- 18

[3] Mehmood, Rizwan; Maurer Hermann: Merging image databases as an example for information integration; Central European Journal of Operations Research, vol. 23, no. 2 (2015), 441-458)

[4] Does the Internet Make Us Stupid?; CACM, (2015), vol: 58, no. 1, 48-51

[5] Zaka, Bilal: Investigating interaction activities in digital libraries: The Networked Interactive Digital Books Project. (In this Issue)

[6] Krottmaier, Harald.; Maurer, Hermann: Transclusions in the 21st Century; Journal of Universal Computer Science (2001) vol. 7, no. 12 (1125–1136

[7] Lanier, Jaron: Ten arguments for deleting your social media accounts right now; (2018), Penguin, Random House, UK

[8] Maurer, H., Mueller, H. (2013) Can the web turn into a Digital Library; Intl. Journal of Digital Libraries, Springer Pub. Co. vol. 13, no.2, 65-75

[9] Scerbakov, A., Ebner, M., Scerbakov, N.: Using Cloud Services in a Modern Learning Management System, Journal for Computing and Information Technology (CIT) (2015), vol. 23, no.1 (1), 75-86

[10] Brabazon, Tara: The University of Google (2007)

[11] Delilovic, Namik; Zaka, Bilal; Maurer, H.: Libraries of Interactive Books as Powerful Tool for Information Communication; AACE- Proceedings of ED-Media (2019), 1353- 1359

[12] Bentzen, Naja: Europe's online encyclopaedias, European Parliamentary Research Service (2018) — PE 614.657

[13] Delilovic, Namik; Maurer, H.: A Note Concerning Feedback and Queries for Web pages, Journal of Universal Computer Science (2019), vol. 25, no. 7, 733-739

[14] Azfal, Tanvir; Glatz, Mathias; Maurer, Hermann: Finding Reliable Information on the Web Should and Can Still Be Improved CIT- Journal of Computing and Information Technology (2018), vol.26., no.1, 1-6.

[15] Maurer, Hermann; Mehmood, Rizwan: Semi-automatic Story Generation for a Geographic Server (2017), Journal of the Croatian Cartographic Society, vol.16, no. 27, 12-25

[16] Mehmood, Rizwan: Techniques for Consolidation and Exploration of Information on a Webserver; Ph.D. Thesis Graz University of Technology (2016)

[17] Glatz, Mathias: Access to Knowledge on the Web; M.Sc. Thesis, Graz University of Technology (2018)

[18] Nelson, Ted: Literary Machine (1987)

[19] Rubin, F. (1987): GOTO considered Harmful; Communications of the ACM 30 (3), pp. 195-196

[20] Maurer, Hermann: Problems and solutions for using computer (networks) for education; Journal of Research in Innovative Teaching & Learning (2017), vol. 10, no. 1, 63-78

[30] https://iiif.io/

[31] https://www.oeaw.ac.at/

[32] https://austria-forum.org/af/Geography/Lab.

[33] https://meta.snl.no/Great_Norwegian_Encyclopedia

[34] https://www.wikipedia.org

[35] austria-forum.org

Delilovic, Namik. Attended the Polytechnical High School in Zavidovići (2003 – 2007), got his B.Sc. in Mobile Computing from the Applied University of Hagenberg, Austria and is currently finishing his M.Sc. In Software Development for Business/ Informatics at Graz University of Technology. Diu an Erasmus stint at Bosphorus University - Istanbul (2009 – 2010), worked for a number of companies including Samsung Mobile Austria and was Samsung Best Mobiler - Seul (Korea) . He has good technical skills in IT and WWW, and already a number of publications.

Ebner, Martin. Priv. Doz. Dipl. Ing. Dr. is the official head of Austria-Forum, but is officially mainly responsible for networked learning at Graz University of Technology. He has numerous publications on e-Learning and multi-media and other topics and is author of the standard book on e-Learning in German , the "Lehrbuch für Lehren und Lernen mit Technologien" (http://lt3.eu). He is considered a leading expert in his field.

Maurer, Hermann. Got his Ph.D. from the University of Vienna and spent a number of years at Canadian, USA and New Zealand Universities He was founding Professor and later Dean at Graz University of Technology. He has obtained numerous national and international awards and has publications numbering in the hundreds. . More than you ever want to know about him you can find e.g. under www.iicm.edu/maurer .

Zaka, Bilal. He is a veteran IT professional with over 17 years of experience in industry and academia. Bilal managed an offshore consultancy venture with Eletel Inc. USA. He headed a team of developers in Pakistan who worked on a product range for mobile devices. Besides regular offshore software development activity Bilal provided wide consultancy services. Bilal earned HEC's overseas scholarship for PhD studies in 2005 and proceeded for a Ph. D, in Austria. Bilal joined Institute for Information Systems and Computer Media (IICM, now ISDS) at Graz University of Technology Austria. Bilal worked on a number of industrial projects under the supervision of Prof. Dr. Hermann Maurer, and Prof. Dr. Frank Kappe.