The Austrian way of Wiki(pedia)!

Development of a Structured Wiki-based Encyclopedia within a Local Austrian Context

Christoph Trattner[†]
ctrattner@iicm.edu

Ilire Hasani-Mavriqi[†]
ihasani@iicm.edu

Denis Helic[‡]
dhelic@tugraz.at

Helmut Leitner[†]
hleitner@iicm.edu

[†] Institute for Information Systems and Computer Media
[‡] Knowledge Management Institute
Graz University of Technology, Austria

ABSTRACT

Although the success of online encyclopedias such as *Wikipedia* is indisputable, researchers have questioned usefulness of *Wikipedia* in educational settings. Problems such as copy&paste syndrome, unchecked quality, or fragmentation of knowledge have been recognized as serious drawbacks for a wide spread application of *Wikipedia* in universities or high schools. In this paper we present a Wiki-based encyclopedia called *Austria-Forum* that aims to combine openness and collaboration aspects of *Wikipedia* with approaches to build a structured, quality inspected, and context-sensitive online encyclopedia. To ensure tractability of the publishing process the system focuses on providing information within a local Austrian context. It is our experience that such an approach represents a first step of a proper application of online encyclopedias in educational settings.

Categories and Subject Descriptors

H.5.3 [**Group and Organization Interfaces**]: computing, Computer-supported cooperative work, Web-based interaction.

General Terms

Structured Wiki

Keywords

Wiki, local Wiki, structured Wiki, context-aware Wiki, Austria-Forum

1. INTRODUCTION

The Web is the largest knowledge base ever built in the history of mankind. Moreover, the recent movements such as Web 2.0 that allow users to publish and add content to the Web in an easy manner result in huge amounts of new Web content generated every day. It is almost self-evident that documents in this huge repository deal with almost any topic that one might be interested in. Also, modern search engines such as *Google* make it possible to find relevant Web information quickly. Therefore, it is understandable that in many areas of human work the Web content and search engines for finding relevant content for the task at hand became an integral part of day-to-day work processes.

However, simply finding relevant information does not immediately imply that the found information is of a reasonable quality [19, 15]. Typically, information on the Web, in general, and Web 2.0 sites, in particular, widely varies in quality – it is then left to the users to decide, based on their previous knowledge of a given topic or on their trust in the authors of the documents they found, which information items are of a good quality [17, 2]. However, in many cases users do not posses the required knowledge to assess the quality of information [17] (they were after all searching for that information, typically, in order to acquire new knowledge) nor, in cases of user-generated content, do the users know the real identity of the content authors and therefore cannot establish a trust relation with the authors [40]. Thus, very often the quality estimation of accessed information items is completely arbitrary [19].

There are numerous problems in different areas such as education [36, 21, 15], scientific publishing [36], journalism [12], or business [26] arising from this situation. These problems include, but are not limited to, plagiarism problem, copy&paste syndrome, use of biased information, popularity as the only measure of information quality, fragmentation of knowledge, or problems of copyright violations.

In this paper we present a Wiki-based encyclopedia called *Austria-Forum* that addresses some of these issues in an educational local Austrian context. The system aims to provide a technological infrastructure that, on the one hand, makes it possible to control the quality of published information, and on the other hand, supports learners in context-sensitive searching and browsing of available information. Thus, this

WikiSym '10, July 7–9, 2010, Gdan'sk, Poland

paper contributes to the community as follows:

- Outline of the three main problems that are currently present in applying Web and Web 2.0 in education.
- Introduction of *Austria-Forum*: a large Wiki-based online encyclopedia system addressing the previously identified problems by introducing an alternative publishing process and a range of structural and contextual concepts.
- Discussion of the system's status and problems which emerged while running *Austria-Forum* as the largest Austrian online encyclopedia available.

The paper is organized as follows. In Section 2 we discuss in more detail the current problems of the above mentioned combination of the Web and search engines in an educational settings. Section 3 presents a Wiki-based approach that is applied in *Austria-Forum*. In Section 4 the current implementation of the system is discussed. Section 5 presents the current status of *Austria-Forum* and lessons learned in applying such a novel Wiki-based approach. Section 6 provides an overview of related work. Finally, Section 7 concludes the paper and provides directions for future work.

2. CURRENT PROBLEMS OF WEB-BASED EDUCATION

It is our experience from many Web-based educational projects in universities and high schools that the following problems are still not sufficiently addressed by diverse Web-based educational systems:

- Plagiarism problem or *Google* copy&paste syndrome
- Unchecked, incorrect, or biased information on the Web
- Fragmentation of knowledge and acquiring of superficial knowledge because of lacking context(s) available

The following sections will discuss these problems in more detail.

2.1 Plagiarism problem or Google copy&paste syndrome

As recent studies suggest writing school reports, university essays, seminars, or master thesis is often supported by using *Google* and *Wikipedia* [26, 36, 12]. As Weber states: *"Many students, researchers and journalists start and finish their work with Google and Wikipedia: they use information that they found – in whatever the way and without any analysis or inspection"* [36]. As a consequence numerous plagiarism checking systems have been developed [24, 26], but such systems are limited in their achievements by definition – the systems can only make a suggestion that a particular document might be a plagiarism.

2.2 Unchecked, incorrect, or biased information

To guarantee correctness of information published on different Web 2.0 sites is elusive, if at all possible. Such sites follow the concept of the "Wisdom of Crowds" [30], as a model of democratization of information publishing and access. Essentially, the idea behind this concept is that collective attention of many users improves the quality and corrects the errors [30, 37]. There are certainly positive aspects of this concept in regard to the correctness or even the quality of information. For example, some recent models for measurement of the quality of *Wikipedia* articles are based on the number of contributors and the number of edits for a particular *Wikipedia* page [18]. The calculated quality of *Wikipedia* contributions increases with the number of contributors and the number of edits [18]. Also, some studies compared the quality measures obtained by these and similar statistical models with the quality assessment made by experts in a particular area [2]. These studies showed that the estimated quality very often matched the expert-asserted quality. However, such studies investigated only a small fraction of Web-based articles (e.g. less than 100 articles) from areas such as entertainment or geography. Many other reports suggest that very often information found in *Wikipedia* articles was not correct and needs to be used with caution [10, 35]. For example, Waters writes that during his "History of Early Japan" class at the Middlebury College a number of students reported incorrect information on two topics. Surprisingly, all of these students used virtually the same language in the incorrect paragraphs – they simple copied&pasted information from the *Wikipedia* articles on these topics [35]. This shows starkly the whole negative consequences of copy&paste syndrome in combination with incorrect or unchecked information. Often, information published on user-generated content sites is biased towards a particular political, commercial, or ideological opinion or view on a topic and is therefore not based on the facts [10]. To remedy this problem the *Wikipedia* organizers have started a background editorial process for some of the *Wikipedia* articles [10]. However, this editorial process is still an anonymous one, i.e. the editors are anonymous and they use their Wiki user names – there is still no official authority that stands behind such an editorial process. Moreover, the sheer amount of the *Wikipedia* articles and the rapid growth of the number of articles makes it impossible to check all articles.

2.3 Fragmentation of knowledge and acquisition of superficial knowledge

Today, users that need to acquire a certain knowledge "*Google*" significant keywords and access a couple of pages shown on one of the first result pages. The decision which pages to access is typically made by the short page excerpt shown beneath the link to a page. When they access a page users skip through "irrelevant" parts of the page until they find the needed information. If the search was not successful the process might be repeated a couple of times by altering the search keywords. The result of such a "learning" process is that users read many small fragments of found documents and acquire what can be called fragmented knowledge [15, 19]. The context, correlations, or connections between parts of knowledge are not visible in such a learning process and the users miss getting a general idea or an overview of the topic of their interest. As a consequence the acquired knowledge of the topic is merely superficial [36].

3. APPROACH

To solve the problems described in Section 2, a Wiki-based encyclopedia system was developed which supports a user in his/her work in a local, structured, and controlled way. Thus, an Austrian online encyclopedia system called

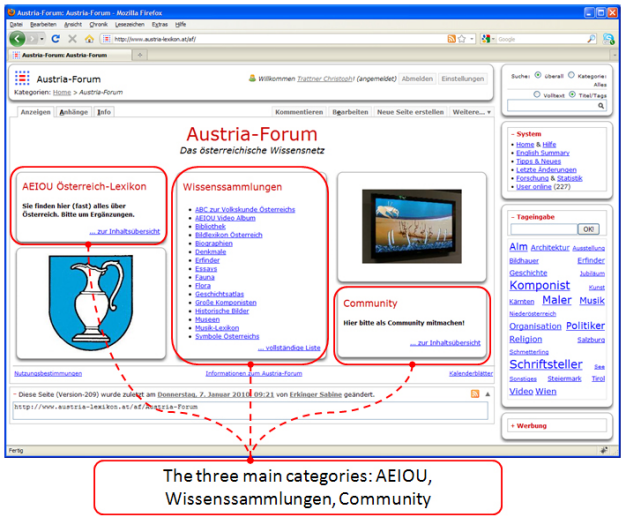

Figure 1: Screenshot of the HomePage of *Austria-Forum*. In the middle – from left to right – the three main categories "AEIOU", "Wissenssammlungen" and "Community" which are available for contribution.

Austria-Forum[1] was developed by adopting the following principles:

- Local Austrian as opposed to global *Wikipedia*-like content.
- Quality inspected content as opposed to "Wisdom of the Crowds" approach.
- Structured and context-sensitive Wiki as opposed to flat structures of the majority of Wiki systems.

3.1 Local Austrian content

Over the last decade, *Wikipedia* has established itself as the largest free online encyclopedia ever built. It contains a collection of information contributed by individuals from all around the world. What *Wikipedia* offers to a global community of users, *Austria-Forum* aims to offer on a local scale to users with an interest in a specific topic, i.e. in this case any topic related to local Austrian context. Although the main focus of *Austria-Forum* is to provide local information about Austria, it is not intended to be considered as just a national platform. As its key publishers state it: *"in times of a world globalization, Austrians should know more about their own culture"* [5].

The main point in which these two systems differ from each other is that they focus on supporting different target groups of users. *Austria-Forum* provides more specific data on Austria but on a finer level of detail; i.e. it offers fine-grained information to users with special interests in Austria. *Wikipedia*, however, contains articles from a global context. It is true that the German version of *Wikipedia* contains a certain number of articles related to Austria, but there is typically a lack of detailed information. To illustrate the point: such information may be a detailed history of a town, say Graz, and changes that it underwent during a specific period of time, say 16^{th} century. Such an information is, for sure, not interesting in a global *Wikipedia* context and therefore cannot be found in *Wikipedia*. On the other hand, for a student, or high school pupil writing a school work on the history of Graz this information is of great interest. Another example might be information on possibilities of hiking and climbing near Graz. Surely, a *Wikipedia* user will not find this information there, since considering the global *Wikipedia* context this type of information is never included in *Wikipedia*. However, a visitor coming to Graz might find this information interesting. A final example worth mentioning is also the complete collection of art work of such well-known painters as the Viennese Kurt Regcheck with digitalized version of his artistic work.

Thus, the *Wikipedia* community is not much motivated to provide fine-grained information in such global system as *Wikipedia*. On the other hand, *Austria-Forum* system has shown that it serves as an inspiration for the community to contribute with highly valued information in a local context.

3.2 Quality-inspected content

In order to overcome the problems outlined in Sections 2.1 and 2.2, *Austria-Forum* was implemented with one crucial point in mind: it should contain articles that are citable, which indicates that their correctness is guaranteed. Citable articles mean that the author of an article is known, i.e. the authors are authenticated with their real names. Moreover, once an article is "stamped" as reliable, it is assured that it will not change.

It is understandable that a strictly controlled content has its drawbacks, e.g. scalability problem to mention just an obvious one. Also, the success of a wiki-based system depends on its openness and a collaborative contribution of users. But how trustworthy and reliable is this user-generated content? In the case of *Wikipedia*, techniques were developed to estimate the trust of its articles. Worth mentioning is a trust system [1] which computes and displays trust values for text in *Wikipedia* articles in order to show how reliable a *Wikipedia* text might be.

In the *Austria-Forum* case, the above mentioned issues were approached by combining the openness and collaborative concepts of Wiki systems, such as *Wikipedia*, on the one hand and controlling the content on the other hand. For this purpose a specific publishing procedure was applied. This process is facilitated by a committee of four key publishers and a board of editors which consists of around 60 well-known Austrian experts in different fields. Publishers and editors are politically independent [5].

The publishing procedure is characterized by three pillars, respectively three main lexica (see Figure 1) within *Austria-Forum*. The first one, the so-called AEIOU - Austrian lexicon, presents the predecessor of an already existing lexica which was integrated into *Austria-Forum*. This area is characterized by a collaboration between editors which is a many-to-many communication via comments. Here articles are reviewed by editors, checked for correctness and quality, and finally "stamped" as a citable resource. The final step is also named a freezing process, which means that the articles cannot be changed any further. However, they become fully citable, which means that a unique URL describing the citation is provided (highlighted in Figure 2). In this way, it is known *who* has contributed in these documents and *when*

[1] `http://www.austria-lexikon.at`

the contribution has been done.

Next, a more restricted space presents the knowledge collection of special lexica about Austria, where the articles are published mostly by key publishers. Normal users cannot publish in this area but they are allowed to comment on different articles. Thus, this pillar is characterized by one-to-many communication between publishers and editors or users. The articles of this area are also reviewed and then marked as citable.

Last but not least, there is the community area, where users can contribute by creating new articles or editing the existing ones. The contributed articles undergo a review procedure and if they meet certain criteria, they are accepted as highly valued articles. These articles are then accessible as fully-citable ones. Here again, a many-to-many communication and collaboration schema is introduced.

3.3 Structured Wiki content

As often criticized, Wikis of the first generation (see original Wiki invented by Ward Cunningham *WikiWikiWeb* [9] for instance) were wild and unstructured [22]. Thus, in general they did not support any features organizing schemes like categories, sub paging, name spaces etc. to structure information within the Wiki. But the Wiki philosophy – quick and simple – changed over the years and Wikis include more and more functional parts these days. Nowadays, Wikis are more seen as a kind of *Web Content Management System* [6], whose strengths are their simpleness and community components such as collaborative editing or discussions. Another trend that can be investigated is that Wikis tend to specialize more and more in a particular field (eg. semantic Wikis, geographical Wikis, etc.). One typical advocate of such a "special" Wiki system which organizes its content in a structured way, is a so-called "structured" Wiki (cf. [32]). These Wikis integrate special structural features like structured data schemes, category mechanisms, semantic annotation methods, semantic search methods, etc. just to name a few, by default within the system to bring structure into the "wilderness" of a Wiki.

The idea of *Austria-Forum* was, more or less, to implement such a system, but in a further developed way. Thus, a structured and context-aware Wiki system based on the Open Source Wiki software *JSPWiki* [20] was implemented. The idea of structured Wiki content was adopted mainly to reduce the problem of *fragmentation of knowledge and acquisition of superficial knowledge*. The idea was to make the users more aware of the context they are surfing in (see Section 2.3) and provide them with a better overview of the information items they are searching for. Thus, a couple of special structural features were developed within *Austria-Forum*. These features include for instance an automated document clustering and categorization module maintained by an editorial board, a context-aware browsing mechanism via hierarchical categories, a hierarchical bread crumbs module, etc., to give the users a better general understanding of the content/context they are searching for. To help the user to get a better "networked" understanding, a recommender system was implemented which provides related document terms and links over resource specific (context-aware) tag clouds. These tags are created by the community. Finally, to help the user during his or her search process the search functionality of the original *JSPWiki* system was enhanced by means of a context-aware search mechanism that

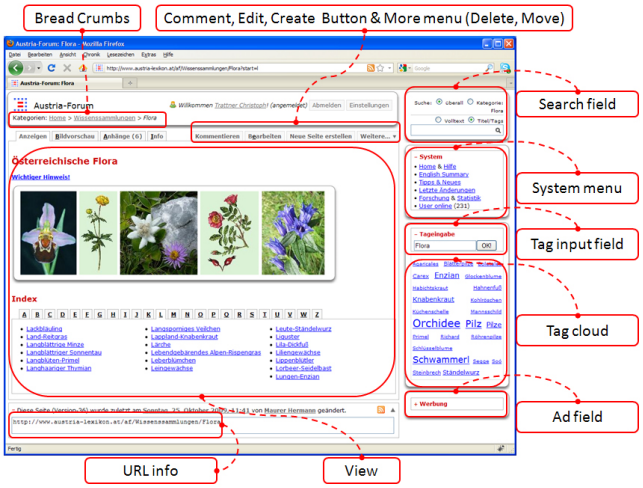

Figure 2: Search, structural and context-aware elements within *Austria-Forum*: Hierarchical bread crumbs and systems's menu on every page for easier navigation and context-awareness. Tagging module with tag input field for annotating documents and tag cloud for navigating/retrieving related documents. URL info on every page for easier page referencing. Tag/title and full text search with context-aware search option "Kategorie" or "Überall" for better search results.

allows recursive category sensitive search, and a tag/meta-data driven search mechanism, that allows querying the system with a search string based on tags or a meta-data information.

4. IMPLEMENTATION

As already mentioned in the introductory part of this paper, *Austria-Forum* is based on the Wiki software *JSPWiki 2.8*. The following section gives a brief overview of novel functionalities that were implemented within *Austria-Forum* (see Figure 3).

4.1 Structuring Content

Currently, there are two popular concepts emphasized, which are applied by Wiki systems in order to structure their content by means of categories. The first concept which shall be discussed here is the so-called (semantic) tag based categorization approach. It invites the user to add a special "category" tag (which is a special Wiki-code-block) to every (related) document the user wants to categorize. The definition of the category is thereby written in the Wiki code-block itself. A special category-page is then typically used to give the user an overview of the categories available in the Wiki system (cf. [38]). *MediaWiki* has extended the basic flat category approach, to be able to create a hierarchically category structure over category tags [34].

The second categorization concept is known as "subpaging". Unlike to the tagging approach subpaging requires an existing category before creating a new document = (sub)page beneath a certain category. Moreover, inter-linking between category-pages and sub-pages is done automatically. There are a couple of Wiki systems like *MediaWiki*, *PHPWiki* [28]

or *TWiki* [32] which support such a feature but limit it to just one level. However, whether the tag based approach or the subpaging concept is the optimal method for creating a category-based hierarchy system within a particular Wiki, is hard to say, since both concepts have their certain advantages and disadvantages [7, 22, 39].

In *Austria-Forum* a subpaging concept was implemented because of the following reasons: 1.) Categories in *Austria-Forum* should only be created and maintained by persons from the editorial board. 2.) Categories in *Austria-Forum* should provide a hierarchical structure like a taxonomy. 3.) Categories in *Austria-Forum* should reflect a sub encyclopedias and sub categories. 4.) The category mechanism should follow a popular and easy to handle concept. 5.) Categories, subcategories, and documents should be inter-linked automatically. 6.) Categories should be movable. 7.) Categories should be lockable (recursively). 8.) Categories in *Austria-Forum* should be hierarchically browsable.

The feature that distinguishes *Austria-Forum* from typical 1-level approaches used in *MediaWiki*, *TWiki*, or *PHPWiki* is that *Austria-Forum* supports almost unlimited levels of categories and sub categories in a hierarchical form. Since *JSPWiki* (Version 2.8) did not support such a feature by default (Version 3.0 will support subpaging, but it was not available at the time *Austria-Forum* was developed), new *FileProvider* modules were implemented. These two modules are: *AustriaForumFilesystemProvider* for page content and *AustriaForumAttachmentProvider* for handling media files, since *JSPWiki* handles media content as separate files attached to a particular site. The great advantage of *JSPWiki*, as compared to popular Wiki systems such as *MediaWiki*, *PHPWiki* or *TWiki*, is that the *JSPWiki's* storage module is well abstracted. Thus, it was possible to implement two simple structural file system provider modules, that replace the existing ones with hardly touching other modules. As one might know, disk access often becomes a bottle neck since *I/O* processing is often slow. To handle this problem, extended versions of *FileProvider*- and *AttachmentProviderModul* were additionally implemented, which enhance these modules via a dynamic caching mechanism, i.e. in *Austria-Forum* only content is cached which has already requested/viewed before.

4.2 Retrieving Content

As described in Section 4.1, one of the requirements of *Austria-Forum* was that categories are hierarchically browsable, i.e. retrievable. Since a subpaging concept was implemented on the file system level it was obvious to profit from such a concept also on the presentation layer. Thus, unlike to the CamelCase approach typically used [22] within Wiki systems, *Austria-Forum* implements an URL addressing and information retrieval schema based on the concept of structural URLs and link [31]. The following URL notation is used within *Austria-Forum*:

```
/<category-page>/<sub-page>
/<category-page>/<category-page>/<sub-page>
/<category-page>/<category-page>/<category-page>
```

category-pages (`<category-pages>`) provide an overview of the category and structural links to subcategories and documents belonging to that category. Technically seen, there is no distinction made between category-pages and sub-pages on the presentation layer. The Wiki treats all category-pages as "normal" Wiki-pages. The only difference that can be drawn is that category-pages have one or more sub-pages attached. Thus (for example), in order to retrieve a contribution about "Konrad Lorenz" out of the category "Biographien", the following URL is constructed:

```
http://www.austria-lexikon.at/af/Biographien/
Lorenz_Konrad
```

To retrieve all contributions from category "Biographien", the following notation is used:

```
http://www.austria-lexikon.at/af/Biographien
```

The corresponding structured URLs are referenced within *Austria-Forum* with the following structured links:

```
[Biographien/Lorenz_Konrad]
[Biograpien]
```

Note that a relative linking schema is per se *not* supported by *Austria-Forum*. The reasons for such a behavior were the following: 1.) We did not want to confuse our users too much with many different types of URL addressing schemes. 2.) We wanted our users to always know in what context they would find a contribution within *Austria-Forum* 3.) We did not want to implement a relative URL addressing schema since it would have made page renaming and automated reference adaptations complicated. Thus, instead, a page filter was implemented that allows the user to insert a relative (`[Lorenz_Konrad]`) or full referenced URL (`[http://www.austria-lexikon.at/af/Biographien/Lorenz_Konrad]`) but rewrites the relative URL to an absolute one when a *save* action occurs (`[Biographien/Lorenz_Konrad]`).

Since this hierarchical addressing scheme is rather popular (see the Open Directory Project *dmoz* [11], the *Google Directory* [14] or *Yahoo! Directory* [42] service for instance), whenever it comes to addressing structured data items in the Web, and since one should always be able to address the right categories or sub-page, it is assumed that users will get used to this kind of notation and information retrieval behavior within the *Austria-Forum* system rather quickly.

4.3 Creating, Editing, and Controlling Content

One of the most important parts of a Wiki system are the create/edit functionalities and the possibilities of controlling/observing edited content. In *Austria-Forum* the process of creating and editing content is more or less similar to the concepts typically used in a Wiki-based system, i.e. clicking on a "broken link" or looking-up a non-existing page leads to the opening of an editor to create a new page. The only real difference at that point is the fact, that a "Create new page" button was additionally implemented and attached to every page editable since usability studies showed that users had some troubles with creating a page the Wiki way. To edit a page, a simple "Edit" button shows up on every page that is allowed to be edited. Note that the same paradigms to "Create" and "Edit" are used for category-pages as well as sub-pages within *Austria-Forum*.

Since controlling edited content within a Wiki system is often a hard job and a considerable number of edits are done in a short period of time, a structural controlling and observation mechanism was implemented within *Austria-Forum*. Unlike the paradigm usually used in common Wikis such as *Wikipedia* where *Special_Pages* are implemented to show

up *RecentChanges* or *BrokenLinks* (just to name a few) to help the editorial community to control edited content in an easier and more convenient form, *Austria-Forum* implements such a feature for every single page available within the system. In other words, in *Austria-Forum* one can follow *In-* and *OutboundLinks*, *Broken-* and *UnknownLinks*, *Diff*-information, *Version*-history and *RecentChanges* on every single sub or category-page. Note that on category-pages links are checked recursively as well, i.e. one will get an overview of the whole link quality structure by clicking on the *Info*-tab of a category-page. The same feature will be available for the *RecentChanges* functionality in the near future.

4.4 Organizing Content

Since category-pages within *Austria-Forum* can be seen as directories on file system layer and container on presentation layer, a range of plug-ins was implemented to equip the user with a couple of simple tools to handle the structural concept used within *Austria-Forum* in an easy way.

One of the points that is often criticized [39] when running a Wiki system based on the concept of subpaging is that documents are limited to just one category whereas contributions categorized by a tagging approach may be available in any category with just a simple category tag assignment. To overcome this problem, a "transclusion" plug-in was adopted that allows the inclusion of any Wiki document by a single line of Wiki code (`[{Insert page='<sub-page>'}]`).

For an easier category handling and for organizing content within a certain category within *Austria-Forum*, a couple of structural plug-ins were invented (see Figure 3). For instance, a plug-in called `[{CategoryIndexPlugin}]` was invented, which generates a simple/paginated alphabetically sorted link list of the sub-pages available within a certain category, or plug-ins called `[{GlossaryPlugin}]`, `[{Tabbed GlossaryPlugin}]` were implemented which generate a simple glossary based and alphabetically sorted link presentation of the sub-pages available within a certain category of *Austria-Forum*. Overall, nearly 20 so-called "structural" plug-ins were developed to handle content such as documents and attachment files in a structured way. A full list of all plug-ins with a detailed description is available online[2].

Note that all resource specific plug-ins are so-called "reference save", i.e. any plug-ins that are dealing with any type of resources within *Austria-Forum* are indexed by the reference manager module provided by the *JSPWiki* engine, i.e. when moving/deleting/renaming a category, these plug-ins will not break the link structure of the system, since referring links are also renamed or deleted.

4.5 Tagging Content

Recently with the emerge of modern Web 2.0 applications such as *flickr*[3] or *delicious*[4] social tagging systems have gained tremendously in popularity [25, 41]. In such systems people use free form vocabulary to annotate a resource (URL) with a special keyword (tag) [16]. The weighted set of keywords (tags) assigned to a defined set of resources (URLs) by a set of users within a system and visualized as a navigation or organizational tool is called a "tag cloud" (cf. [3]).

[2] `http://www.austria-lexikon.at/af/Hilfe`
[3] `http://www.flickr.com`
[4] `http://www.delicious.com`

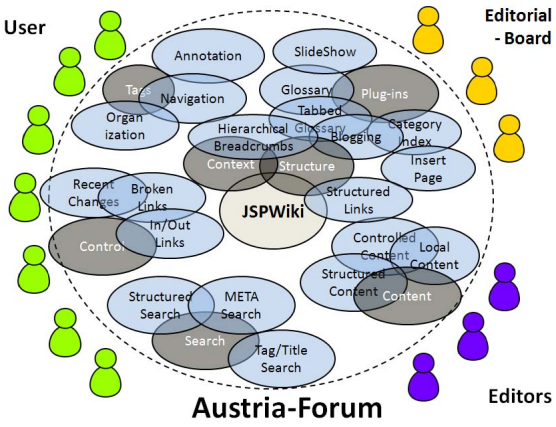

Figure 3: Component diagram of Austria-Forum (with new core components marked as gray clusters).

In *Austria-Forum* a built-in "tagging" approach was implemented [33]. The motivations for developing and integrating such a system within *Austria-Forum* were the following (cf. [3]): 1.) Provide the user with a simple to handle tool for organizing information items within *Austria-Forum*. 2.) Provide the user with related terms and links to related documents via a tag cloud presentation on every single page within *Austria-Forum* to get a better networked understanding since related documents are tagged with similar terms/tags. 3.) Provide the user with a navigational tool since related documents are connected over related terms/tags. 4.) Provide the user with a tool for enriching documents on semantic level by means of free keyword annotations that are indexed by the system's search engine module to provide a title and tag based search mechanism.

The system works as follows: With every document retrieved by the user a tag field shows up (see Figure 2). Thus, annotating a resource within *Austria-Forum* simply requires a term to be filled into the tag input field and a completion of process by clicking the "Ok" button to store the tag. Tags are stored in a special tag database over a special tag service routine [33] which requests the built-in search engine module to index the tags as well. The transfers to the tag service routine are done asynchronously via *AJAX*.

To profit from this tagging approach also in means of organization and navigation information items within *Austria-Forum*, a tag cloud module was implemented. The module itself works as *JavaScript* module that renders a tag cloud presentation out of a *XML* data file provided by the tag service routine. The *XML* data file contains user or resource specific tags. User specific tags are used to render a personal tag cloud, which helps the user to get a quick overview of the information items "bookmarked". The resource specific tags are used to render a resource specific tag cloud, i.e. this tag cloud takes all tag assignments from all users into account that were made to one specific resource. By clicking on a tag one gets a list of resources that were tagged by the same tag within the system, i.e. one can navigate related documents by means of related terms and resources with the help of such an approach.

4.6 Searching Content

As already mentioned in the introduction, *Austria-Forum* should enable context-sensitive searching and browsing of content. In order to fulfill this requirement, new mechanisms were implemented which complement the existing functionality of the open-source search engine *Apache Lucene* [4] applied in *JSPWiki*. Context-aware and meta-data search mechanisms provide the means for an enhanced search within the system. In this way, users are spared excessive retrieval results; they can perform search in specific areas of the system. Moreover, retrieval results are refined through meta-data.

A search in *Austria-Forum* is performed by using the search field visualized in Figure 2 in the right upper corner. When a user searches for general information, the option "everywhere" ("Überall") should be activated. A more sophisticated option presents the search within categories which can be activated by clicking on "category" ("Kategorie"). The system is aware of in which category the user is currently navigating, so that by activating this option, search results found only in the corresponding category are returned. Furthermore, a full-text search and a tag/title based search are also provided.

The concept of meta-data search is implemented by using the *Lucene* field-data structure. Each field object is characterized as a key-value pair which conforms to the meta-data mapping. The defined meta-data differ among categories. For instance, in the biography category typical meta-data are: date of birth, place of birth, or area of work; whereas in the alpine flora category such meta-data would be: region, season, or color. [5] This approach ensures an efficient, precise, and specific information retrieval within the system.

5. DISCUSSION

This section gives a short overview of the systems's status and the problems emerged while running *Austria-Forum* as the largest Austrian online encyclopedia system available over a period of four months.

5.1 Current status of Austria-Forum

Officially, *Austria-Forum* was launched at a press conference[5] in Vienna on October 19^{th}, 2009. At that point the system comprised around 90,000 information items (30,000 documents and 60,000 multimedia content such as audio, video or PDF files) and was facilitated by a committee of four key publishers and a board of editors which consisted of around 40 well-known Austrian experts in different fields. There were five members in the editorial team at that time. Their task was to edit or digitize existing Austrian contributions from well-known Austrian authors. The technical team, responsible for implementing and running the Wiki, consisted of two developers at that time.

Currently, four months later, *Austria-Forum* has established itself as the largest Wiki-based online encyclopedia available in Austria. It attracts around 3,500 different users each day and serves almost 121,000 information items (42,500 documents and 78,500 multimedia content such as video, audio or PDF files). Thus, over 30,000 new contributions were made within a period of four months. The size of the editorial board grew from 40 to 60 experts from different fields. The number of employed editors stayed at a size of five, while the number of Wiki developers increased to a team of around six people. The number of community-users is at the moment of around 750 (registered) members.

[5] `http://www.austria-lexikon.at/af/Tipps_und_Neues/Pressespiegel`

5.2 Problems encountered

The following section discusses shortly the main issues which encountered while running Austria-Forum the last four months (from 2009-10-19 until 2010-02-18), viewed from a community and a usability perspective.

5.2.1 Community issues

One of the most surprising issues that arose while running *Austria-Forum* was the fact that even if around 30,000 new contributions (overall) were made since the system was launched in October 2009, the number of contributions in the community area was proportionally vanishingly small. Thus, around 1% of all contributions were made by users who were registered in the system but who were only allowed to publish content into the community area of *Austria-Forum*. Much better was the output of the group of voluntary editors who were allowed to contribute in the two closed categories "AEIOU" and "Wissenssammlungen". They produced nearly 40% of the new content since October. The rest (59%) of the contributions were made by the group of employed editors.

Although the overall numbers are rather promising for the group of editors who volunteer in the *Austria-Forum* project, the number of posts in the community area is rather lacking. Thus, we plan to push this "category" a little bit further by opening it for anonymous editing and collaboration in the future. Moreover, we plan to open the comment- and tagging-feature within *Austria-Forum* for anonymous accessors as well.

5.2.2 Usability issues

Since the system's usability was always one of the most important claims while implementing *Austria-Forum*, a usability study was conducted during the post-development phase of the system in summer 2009 to fix the "biggest" usability issues before the system was released in October 2009. Thus, a heuristic evaluation [27] by a group of 20 experts (graduated students of computer science, 22 years of average age, 60% male and 40% female) was conducted in a first step [13]. The goal of the expert users was to record all positive and negative findings during the evaluation phase of the system and to give additional feedback via a feedback questionnaire after the test. Interestingly, the experts recorded about 60 points which they found annoying, but correlated just with three of their findings. As the top 3 negative findings, the experts pointed out: the poor German localization of the system (*JSPWiki* provides a German localization file by dafault, but most plug-ins are not well localized), the bad Quick-Help-Page (with a short overview of all common Wiki commands) and the missing (overall) Help-Page of the system.

Since detailed feedback about the editorial processes was requested before going online, i.e. we wanted to know how feasible the system is regarding the process of searching, creating, and editing documents within *Austria-Forum*, additionally, a thinking aloud [23] test with a group of 25 test users (92% students, 8% pupils, 60% male and 40% female, 21 years of average age) was conducted [13]. The test itself

was split up into 5 tasks ranging from quick (max. 2 minutes of time) and easy to "hard" and time intensive (max. 10 minutes of time). The tasks ranged from *logging into the system*, *searching for a particular article* to *creating/editing a document and uploading/displaying an image file*. Surprisingly, one task produced fatal problems among the test users, i.e. none of the 25 test users could solve the task. It was the task of uploading an image file and including it into a document. Further, the test users had problems with the relevance of the search results, since this functionality was not fully implemented at that time, and the creation of hyperlinks, since they were not used to the Wiki syntax and could not use the editor's toolbox, since it was closed by default at that time. Of course, all of the findings were investigated and problems could be solved before the system was launched in October 2009.

6. RELATED WORK

The tremendous success and expansion of Wikis lies on their simplicity and efficiency. Moreover, most of them are available as open source and their syntax is easy to learn. In particular, the most rapidly grown Wiki system – *Wikipedia* has raised a great interest of many researchers. A group of researchers of Palo Alto Research Center [29] investigated the publishing process of *Wikipedia* and concluded that recently the growth of *Wikipedia* has slowed down. Suh et al. [29] analyze the overall activities in *Wikipedia* with the focus on editing activities such as new page creation, adding/modifying/removing contents in existing pages, and reverted editing. Firstly, it is showed that there has been a general slowdown in *Wikipedia* editing activities in the last two years. Editors were divided into classes based on the number of their contributions per month and editing activities within each class were investigated. In *Wikipedia*, a small group of "elite" editors contribute the large amount of edits while the larger group of "normal" users contribute the rest. It is proved that editing activities of the editor's class with the most edits per month did not decrease in the last two years, whereas the editors belonging to other classes decreased their overall editing activities. The researchers explain these new *Wikipedia* developments by *"decreased opportunities for sharing existing knowledge and increased bureaucratic stress on the socio-technical system itself"* [29].

Citizendium [8] – a project of Larry Sanger, a co-founder of *Wikipedia* presents an approach which is mostly related to our work. However, the rapid growth of *Austria-Forum* is not to be compared with the one of *Citizendium*. When this paper was written, *Citizendium* contained 13,161 total articles, of which only 121 articles where approved by editors.

Problems encountered by using free encyclopedias in education (see Section 2.2) present also the main focus of the article written by Waters [35], where he states that a rigorous publishing system should enhance the applicability of citable resources from *Wikipedia* in educational settings. One of the main concerns of Waters is the anonymity of the *Wikipedia* articles, which could be improved if the authors provide their real names instead of their wiki-usernames. This issue and also issues regarding the voluntarism of editors and their field of expertise are successfully addressed in the case of *Austria-Forum* and presented throughout this paper.

7. CONCLUSIONS AND FUTURE WORK

In this paper a large Wiki-based encyclopedia called *Austria-Forum* was presented that aims to combine openness and collaboration aspects of *Wikipedia* with approaches to build a structured, quality inspected, and context-sensitive online encyclopedia in educational settings. To ensure tractability of the publishing process the system focuses on providing information within a local Austrian context. This work is relevant to researchers or developers who are interested in running a large Wiki systems that work in a more local, controlled, and structured way, than current Wiki-based systems such as *Wikipedia*.

Future Work: Even if the overall numbers of *Austria-Forum*, such as number of daily visitors, growth/number of new documents, growth/number of user edits or accounts, just to name a few, are rather promising for a Wiki-based system that has been online for such a short period of time, there is a lot of work and time currently invested to popularize *Austria-Forum*. For instance, one of the biggest projects which is currently planned is a campaign which aims to integrate and evaluate *Austria-Forum* as an e-learning tool in all Austrian schools and universities. In preparation of this campaign, a big usability study is planned to be conducted in the summer term 2010. This usability study will focus on evaluating the system's usability in more detail, i.e. a detailed investigation of modules/features such as annotating and navigating related documents via resource specific tags (tag clouds), retrieving/structuring information items via structural links and plug-ins, context and meta-data driven search, hierarchical browsing, and context-awareness. In addition to this we are currently working on a *NetBeans*[6] and *SVN*-based framework that allows pupils, students, and teachers to access the content and source-code files of *Austria-Forum JSPWiki* and compile/deploy and test them as a user specific instance on one of our test servers. Thus pupils, students, or teachers will be able to implement new features and compile/run their version of *Austria-Forum* as a separated instance on one of our test servers. Last but not least, we are currently working on a geographical extension of *Austria-Forum* that aims to integrate geo-spatial information into existing contributions available in *Austria-Forum* and to offer this feature as a collaborative e-learning tool to pupils and students.

8. ACKNOWLEDGMENTS

We would like to thank Prof. H. Maurer for his fruitful discussions during this work. Furthermore we would like to thank P. Diem, S. Erkinger, C. Fressel, D. Kaiser, M.L. Lampl, H. Mikl, I. Schinnerl, and K. Ziegler for their valuable inputs and discussions during the development phase of the system. Additionally we would like to thank K. Trummer for setting up the necessary hard-ware parts and for his fruitful technical discussions over the last few months. Last but not least, we would like to thank all the volunteers that have helped to make *Austria-Forum* such a big success. This work is funded by Technical University of Graz, the government of Styria and the non-profit organization "Friends of Austria-Forum".

[6] `http://netbeans.org`

9. REFERENCES

[1] B. T. Adler, K. Chatterjee, L. de Alfaro, M. Faella, I. Pye, and V. Raman. Assigning Trust to Wikipedia Content. In *WikiSym '08: Proceedings of the 2008 international symposium on Wikis*, New York, NY, USA, 2008. ACM.

[2] B. Amento, L. Terveen, and W. Hill. Does "Authority" Mean Quality? Predicting Expert Quality Ratings of Web Documents. In *Proceedings of the 23rd Annual international ACM SIGIR Conference on Research and Development in information Retrieval*, pages 296–303. ACM, 2000.

[3] M. Ames and M. Naaman. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM.

[4] Apache Lucene. Apache Lucene. `http://lucene.apache.org/java/docs/`, 2010. [Online; accessed 2010-03-01].

[5] Austria-Forum. Austria-Forum – Das österreichische Wissensnetz. `http://www.austria-lexikon.at/`, 2010. [Online; accessed 2010-03-01].

[6] P. Baumgartner. Wie man das richtige Content Management Tool für ein bestimmtes Lernmodell auswählt. *Fernuniversitaet in Hagen, Institute for Educational Science and Media Research (IfBM)*, Mai 2005.

[7] M. Buffa and F. Gandon. SweetWiki: Semantic Web Enabled Technologies in Wiki. In *WikiSym '06: Proceedings of the 2006 international symposium on Wikis*, pages 69–78, New York, NY, USA, 2006. ACM.

[8] Citizendium. Citizendium. `http://en.citizendium.org/wiki/Welcome_to_Citizendium`, 2010. [Online; accessed 2010-03-01].

[9] W. Cunningham. WikiWikiWeb. `http://www.c2.com/cgi/wiki/`, 2010. [Online; accessed 2010-03-01].

[10] P. Denning, J. Horning, D. Parnas, and L. Weinstein. Wikipedia Risks. *Communication of the ACM*, 48:152–152, 2005.

[11] dmoz. ODP - Open Directory Project. `http://www.dmoz.org/`, 2010. [Online; accessed 2010-03-01].

[12] F. Fedler. Plagiarism Persists in News Despite Changing Attitudes. *Newspaper Research Journal*, 27:24–37, 2006.

[13] C. Fressel. Standard-compliant Web Design Exemplified by 'Austria-Forum'. Master's thesis, KMI, Graz University of Technolgie, 2009.

[14] Google. Google Directory. `http://directory.google.com/`, 2010. [Online; accessed 2010-03-01].

[15] P. Haber. Google–Syndrom. Phantasmagorien des historischen Allwissens im World Wide Web. *Geschichte und Informatik/Histoire et Informatique. Vom Nutzen und Nachteil des Internet für die historische Erkenntnis*, 15:73–89, 2005.

[16] T. Hammond, T. Hannay, B. Lund, and J. Scott. Social Bookmarking Tools (I): A General Review. *D-Lib Magazine, 11(4)*, April 2005.

[17] C. Hölscher and G. Strube. Web Search Behavior of Internet Experts and Newbies. In *Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications netowrking*, pages 337–346, Amsterdam, The Netherlands, The Netherlands, 2000. North-Holland Publishing Co.

[18] M. Hu, E.-P. Lim, A. Sun, H. W. Lauw, and B.-Q. Vuong. Measuring Article Quality in Wikipedia: Models and Evaluation. In *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pages 243–252, New York, NY, USA, 2007. ACM.

[19] S. Johnson. *Everything Bad Is Good for You: How Today's Popular Culture Is Actually Making Us Smarter*. Riverhead Books, New York, USA, 2005.

[20] JSPWiki. JSPWiki – a feature-rich and extensible WikiWiki engine built around the standard J2EE components (Java, servlets, JSP). `http://www.jsp.org`, 2010. [Online; accessed 2010-03-01].

[21] N. Kulathuramaiyer and H. Maurer. Why is Fighting Plagiarism and IPR Violation Suddenly of Paramount Importance? *Series on Innovation and Knowledge Management – Proceedings of International Conference on Knowledge Management, Innovation, Technology and Cultures*, 6:363–372, 2007.

[22] B. Leuf and W. Cunningham. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2001.

[23] C. Lewis. Using the "Thinking-aloud" method in cognitive interface design, 1982. Research Report RC 9265. T.J. Watson Research Center, Yorktown Heights, NY.

[24] C. Liu, C. Chen, J. Han, and P. S. Yu. GPLAG: Detection of Software Plagiarism by Program Dependence Graph Analysis. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 872–881, New York, NY, USA, 2006. ACM.

[25] C. Marlow, M. Naaman, D. Boyd, and M. Davis. HT06, Tagging Paper, Taxonomy, Flickr, Academic Article,ToRead. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 31–40, New York, NY, USA, 2006. ACM.

[26] H. Maurer, F. Kappe, and B. Zaka. Plagiarism – A Survey. *J.UCS*, 12(8):1050–1084, 2006.

[27] J. Nielsen. Finding Usability Problems through Heuristic Evaluation. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380, New York, NY, USA, 1992. ACM.

[28] PHPWiki. PhpWiki: a PHP WikiWikiWeb. `http://phpwiki.sourceforge.net/`, 2010. [Online; accessed 2010-03-01].

[29] B. Suh, G. Convertino, E. H. Chi, and P. Pirolli. The Singularity is Not Near: Slowing Growth of Wikipedia. In *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 1–10, New York, NY, USA, 2009. ACM.

[30] J. Surowiecki. *The Wisdom of Crowds*. Anchor, 2005.

[31] P. Thistlewaite. Automatic Construction and Management of Large Open Webs. *Inf. Process.*

Manage., 33(2):161–173, 1997.

[32] P. Thoeny. TWiki – the Open Source Enterprise Wiki and Web 2.0 Application Platform. `http://www.twiki.org`, 2010. [Online; accessed 2010-03-01].

[33] C. Trattner and D. Helic. EXTENDING THE BASIC TAGGING MODEL: CONTEXT AWARE TAGGING. In *Proceedings of IADIS International Conference WWW/Internet 2009*, pages 76–83. IADIS Digital Library, 2009.

[34] J. Voss. Collaborative thesaurus tagging the Wikipedia way. *Arxiv preprint cs/0604036*, 2006.

[35] N. L. Waters. Why You Can't Cite Wikipedia in My Class. *Commun. ACM*, 50(9):15–17, 2007.

[36] S. Weber. Das Google-Copy-Paste-Syndrom. Wie Netzplagiate Ausbildung und Wissen gefährden. *Heise*, 2006.

[37] A. Weiss. The Power of Collective Intelligence. *netWorker*, 9(3):16–23, 2005.

[38] Wikipedia. Categorization — Wikipedia. `http://en.wikipedia.org/wiki/Wikipedia:Categorization`, 2010. [Online; accessed 2010-03-01].

[39] Wikipedia. Subpages — Wikipedia talk. `http://en.wikipedia.org/wiki/Wikipedia_talk:Subpages`, 2010. [Online; accessed 2010-03-01].

[40] I. H. Witten, M. Gori, and T. Numerico. *Web Dragons: Inside the Myths of Search Engine Technology (The Morgan Kaufmann Series in Multimedia and Information Systems)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2006.

[41] H. Wu, M. Zubair, and K. Maly. Harvesting Social Knowledge from Folksonomies. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, pages 111–114, New York, NY, USA, 2006. ACM.

[42] Yahoo! Yahoo! Directory. `http://dir.yahoo.com/`, 2010. [Online; accessed 2010-03-01].