SaferWeb - Community-Driven Collection Of Suitable Websites For Children

Sebastian Richly, Anne Goethlich, Ines Mauermeister, Michael Thiele

{sebastian.richly, anne.goethlich, ines.mauermeister, michael.thiele}@inf.tu-dresden.de

Abstract: Regardless of what you might think about the Internet and its communities, it is an undeniable fact that their importance and the number of users have grown significantly over the last years. They have also come to manage rather complex tasks, as illustrated by tagging communities and similar applications. In our *SaferWeb* approach, we use a community platform to collect suitable websites for children. Current webfilters for children use blacklist or whitelist approaches. However, blacklists only block already known websites and thus, finding and adding new sites is a time-consuming manual effort. Similarly, whitelists only allow known websites suitable for children, and again, the effort required to find and manage adequate sites is immense as well. *SaferWeb* uses a community to manage a whitelist and thus distributes the effort to many shoulders. In this paper, we present our *SaferWeb* community and its whole architecture also containing a proxy and a browser toolbar.

1 Introduction

The size of the Internet is continuously increasing and it is now said to be *the* central information and knowledge source [Isk05]. Today, the Internet does no longer only provide information. It also helps us get along in everyday life as it offers a large number of services: the ordering of groceries, the booking of flights, bank transactions, or even the casting of political votes. Furthermore, the online interaction with family and friends, such as sending e-mails and chatting, plays a major role in the usage of the Internet. According to the ”2007 Digital Future Report” of the USC Annenberg School for Communication, participation in online communities actually leads to social participation [UoSC].

Children and teenagers represent one of the fastest-growing user groups on the Internet [Raz]. But using the Internet does not always have positive effects for them. Quite the contrary: they might unintentionally visit websites with explicit or even unconstitutional content. For parents it is very important to protect their children from such experiences. Unfortunately, none of the existing approaches satisfy the wishes of parents. Today's filtering approaches using blacklists - for websites or website contents such as words or pictures - are a common technology but they can easily be bypassed. Blacklists contain only those harmful websites that are already known. New ones have to be added manually but this can take a long time and in the meantime, children already run the risk of stumbling across those sites. There also exist approaches using whitelists. These types of lists have the same big disadvantage: sites have to be collected and added manually - mostly by

pedagogues. This results in high personnel costs and small lists. The big advantage, however, is the good quality of such lists, as they are made by professionals.

Today, tagging communities are a powerful tool for social navigation in large data repositories. They are useful to organize information such as websites (del.icio.us), photos (Flickr), blogs (Technorati), and research papers (CiteULike) [Mun99], and they are only based on user input. In the *SaferWeb* project, we want to use the approach of community-driven whitelist creation. This whitelist contains only sites that are - in the eyes of adults - suitable for children. Hence, it is not a blacklist, where prohibited sites are managed. The whitelist is managed by a community but for the *SaferWeb*, this will be a special community that consists of two types of community platforms: one for adults and one for children. On the adults portal, parents can suggest websites they consider suitable for kids. They can cast positive or negative votes for every suggested website they visit in the World Wide Web. On the children's portal, children may search for websites and rate them.
A website will only be stored on the whitelist - and thus be visible for children - if enough positive votes are given. The big advantage with this idea is that nobody has to service the *SaferWeb*. To guarantee that children use only sites on the whitelist, a client-based http-proxy is used. *SaferWeb* achieves two things: it protects children from visiting websites with harmful content and it expands the joy of use by detecting appropriate websites for them. The big advantage of the community-driven whitelist creation is that manual administration is no longer needed. The community itself manages the whitelist.

The paper is structured as follows: In Section 2, we discuss communities and their usage as well as approaches for a secure Web for children. In Section 3, we present our approach of the community-driven whitelist creation and the *SaferWeb* architecture, the implementation of which is described in Section 4. The paper will be concluded by a summary and outlook.

2 Related Work

In this chapter, we want to introduce the character and advantages of communities and discuss why they continue to grow. On this basis, we describe some communities that have been developed in the connection with a childproof Web.

2.1 Communities

More than ten years ago, the book *Net Gain* by John Hagel III and Arthur G. Armstrong was published. Its central assumption was that *virtual communities* demonstrate a new business model that will bring along a global structural change: "Virtual communities are not an opportunity that executives can choose to address to ignore. They represent a profound change that will unalterably transform the business landscape - and benefit only those who confront it head on." [Hag97]

A community can be interpreted geographically as a group of people who live in the same

region or in the same place, or it may refer to people who live together and share a common interest or a common task. More than 120 years ago, the question of what a community is and which rules apply to communities spurred the emergence of sociology.

If you transfer the term *community* to the field of electronic communications, it describes a group of people who have developed common knowledge and shared experiences and who inspire their own identity. All members of a community have the same interests that allow them to join the community to discuss and create knowledge. But the most important criterion is that the members of a community get active and interact with other members of the community.

The motivation for such interaction is found in a common interest or task, such as tagging. In the classic community in the spirit of newsgroups or forums, this interaction happens only by exchanging messages in text form. However, there are also variants of the idea of a community where members dispose of other options to stay in contact with one another.

An online community gives people the opportunity to contact each other and exchange information via Internet. A community on the Internet provides the basic tools for communication such as forums, chat systems, newsboards, swap meetings, matchmaking and so on. These are the most popular tools that allow communication between members. Depending on the target group, the individual functions will be co-ordinated and matched with the interests of the users. In this sense, it is necessary to base the definition of a community not on the technology it uses but on the content that brings the members of the community together. It looks like a social phenomenon. Communities can also be used to intensify the bonds between the user and the website and to enhance their identification with the topic.

An online community can expand quite successfully when its driving force is not some idea of a marketing company but the direct wish for growth of the members of the community. Examples for well-functioning communities are Usenet and MySpace.

2.2 Secure Web for Children

The issue of how to keep children safe when they surf the Internet is not a new discussion. The best way is that parents take the time to accompany their children when they explore the World Wide Web. Unfortunately, they often just do not have the time to track each and every step. However, parents should at least talk with their children about the risks and dangers of the Internet. Children should know that they are never allowed to share personal information with strangers or give away passwords, and they should immediately report to their parents when they feel uncomfortable [Saf].

Not only parents are interested in a solution for the security risks for children in the World Wide Web. There are also government initiatives to create a secure Web, such as the US project *Dot Kids* or the German project *fragFINN*.

In June 2003, the domain *kids.us* [] went online. The US White House describes it as ”much like the children’s section of the library, where parents feel comfortable allowing

their children to browse". It will "be a safe place for children to go" [Rel]. Third-level domains such as *News.kids.us* or *NewYork.kids.us* can be registered and are then checked by NeuStar to ensure that their content is appropriate for children. Unfortunately, not more than forty third-level domains have yet been ordered. The reason may be the high registration fees and the requirement that all external links be removed [Was05].

The idea of *fragFINN.de*, supported by the German government, is a list that consists of appropriate websites for children. This so-called whitelist was compiled, extended, and screened by pedagogues who are specialized in the field of media. With the aid of a particular toolbar, Microsoft's Internet Explorer can be set up in a way that only websites on the whitelist can be visited [Fra]. Criteria to be stored in the whitelist include, for example: Does the page contain explicit content? Does anybody regularly care for the content of the website?
Glubble [Glu] is another project that uses filtering. It does not filter websites. It re-creates and filters search results of Google, Yahoo etc. It also allows parents to activate and deactivate sites for their kids (black- and whitelisting).

There also exist non-governmental projects such as *Imbee* and *Blinde Kuh*. *Imbee* calls itself the "first free social network designed for young people". On *imbee.com*, you can predominantly do personal things like creating a blog or uploading pictures. The website is approved by parents and endorsed by teachers [Imb]. The first German search engine for children under the age of 14, *Blinde Kuh*, was created in 1997. The project is divided into different subject groups like nature, history, environment, and sports, which are then further subdivided into concrete issues. *Blinde Kuh* has been supported by the German government since August 2004 [Kuh].

All of the presented solutions have one disadvantage: someone has to screen the websites and control the content. This leads to a lot of work for which staff is needed. The question arises as to whether or not there is a solution that reduces (or eliminates) the need for staff while being a very secure solution at the same time. One could imagine that parents and their children work together in a community where the parents decide which websites are appropriate. This idea would not only have the advantage that there is no staff needed but also that parents are involved in what their kids do in the World Wide Web. The community-driven approach makes the secure Web for children flexible and dynamic, a fact that should not be underestimated in the age of Web 2.0.

3 The SaferWeb Approach

The *SaferWeb* is a community with the aim to develop a childproof Web experience for kids based on a whitelist. In contrast to other approaches, the *SaferWeb* attempts to avoid editorial costs. Tagging communities show how well Internet users can work together to reach a global aim. The *SaferWeb* community consists of two groups: (1) parents who manage the lists and (2) children who rate sites on the whitelist.

Parents manage the main part of the *SaferWeb* by evaluating different websites. The result is a whitelist of websites containing only content appropriate for children. Every parent

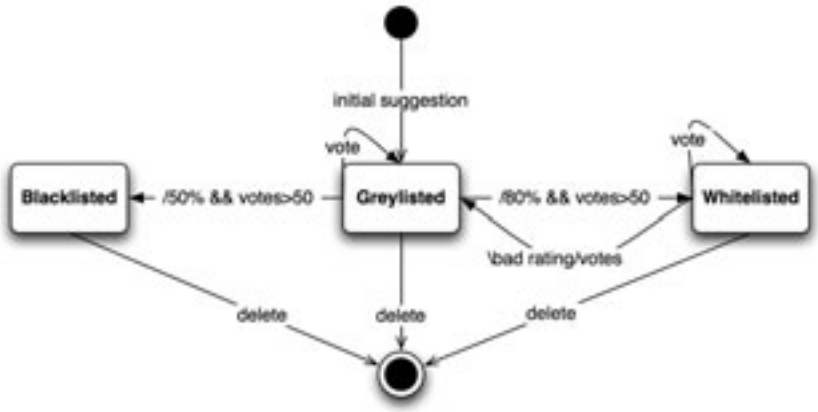

Figure 1: SaferWeb Decision Process

can help by adding more sites or removing old pages via a predefined procedure. After registering with the *SaferWeb*, every parent can suggest a website that is childproof from his or her point of view. This site will be stored in the greylist with an initial positive vote and some describing tags. It will not be stored in the whitelist because the community decides about what eventually ends up on the whitelist.

Now, other parents get the opportunity to evaluate this site. The frontend displays a choice of greylisted websites. They can browse these sites and form their own opinions about them. Finally, they can cast a positive or negative vote for a site. In the case of a positive vote, they can also add tags for the respective site. This process has to be repeated until the site gets enough votes from the community members. The suggestion will be stored in the whitelist if the following condition is met: 80 percent positive votes out of at least 50 votes. Suggestions can also have a high percentage of negative votes. In that case, they will eventually be stored in the blacklist under following condition 50 percent negative votes out of at least 50 votes. Figure 1 shows this process.

This process ensures a good collection of childproof sites from the parents' point of view. But children should also have an influence on the whitelist. In the *SaferWeb*, kids cannot vote for greylist entries but they can rate whitelist entries. If they visit a page from the whitelist, they can evaluate it with a school mark system. This influences the *SaferWeb* in two ways. (1) If a site gets many bad marks, it will be moved from the whitelist to the greylist and the votes will be reset. The page has to pass the community process once more. (2) The children build up a page ranking, which can be displayed on the *SaferWeb* site to help other kids find the best childproof sites.

This two-way process requires no editorial costs and is only based on the input of the community. But to ensure the childproof Web experience to a larger extent, more than just the community is necessary. For this reason, *SaferWeb* consists of four essential components: (1) websites for both community groups. They provide community processes described above and also rankings and news of the community. (2) The *SaferWeb* server, which

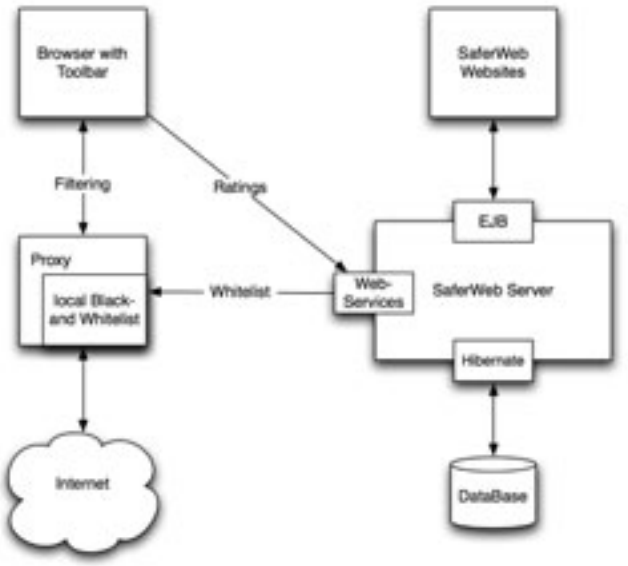

Figure 2: SaferWeb Architecture

stores the lists and user information and which processes all incoming votes and ratings. It also provides the whitelist to the (3) *SaferWeb* filter. This is an http-proxy that filters every page request and lets only sites from the whitelist pass. It also allows to manage local black- and whitelists. The (4) additional browser toolbars for parents and children enable the easy suggestion of Web pages and children ratings. Figure 2 shows the complete *SaferWeb* architecture.

4 Realization

In the following section, the *SaferWeb* components will be explained in detail. The central element is a server on which all user information, like given votes and tags, are saved and can be requested by the other components. The filter installed at the client's computer gets the whitelist from the server and prevents children from visiting websites that are not on this list. At the *SaferWeb* website, parents can vote for or against a specific website, change their user data and inspect the whitelist. Another website for kids also displays the whitelist. Elements from this list can be selected through the choice of tags. The toolbar simplifies the evaluation process. All components and their communication are described in the next sections.

Figure 3: Frontend for the Children's Site

4.1 Remote Components

The *SaferWeb* uses the J2EE JBoss application server as its central communication instance between all components. We use the MySQL database as data storage. The data access is realized through EJB3/Hibernate. The database contains all given votes and added tags that are associated with a user and a URL. With the help of this information, the whitelist is computed at runtime. This means that whitelist, greylist and blacklist are not stored explicitly. Each request delivers the currently correct list. To ensure that the lists can be used on every platform and by other applications, it can be retrieved via a standard Web service [(W3a] [(W3b]. This makes it possible that other similar services can use the results of our community. The communication with the Web frontend for the children and parents is realized with RMI/JNDI [Wut01].
The frontend (see Figure 3) itself uses ICEfaces 1.7 components [ICE], which are based on JSP (Java Server Faces 1.2)[Wut01], both for the parents' site and the children's site. ICEfaces is an integrated Ajax application framework that enables Java EE application developers to easily create and deploy powerful thin-client internet applications in pure Java. In addition, facelets are used to create reusable templates in order to reduce redundancy in the written code.

4.2 Local Components

To ensure that children can only visit the sites on the whitelist, an http-proxy is used. However, this approach requires some additional settings on the local computer. Since only the children and not the parents should be restricted in their use of the Web, it is necessary

Figure 4: Toolbar for the Children

to use different user accounts on the computer. In contrast to the parent account, the child account should not be enabled to change the proxy or other settings of the browser, or the proxy itself; it should also be impossible to run different browsers. This can be realized on all modern operating systems and for all common browsers. Thus, it is easy to ensure that the child visits sites of the whitelist only. The proxy itself is a standalone Java-based application that filters all incoming requests. It runs as a service on the client computer, updates the whitelist periodically, and automatically receives it from the *SaferWeb* server. Since it is based on Java, it can be used on every platform that supports Java. The proxy supports http 1.0 and 1.1, so every common browser can work with it.
When a client requests a website, the filter tries to find a matching whitelist entry. If at least one entry is found, the request is forwarded to the browser. To find a matching whitelist entry, an efficient algorithm is used that performs a search on an ordered whitelist.
The proxy can also manage a local white- and blacklist. This makes it easy for parents to configure the sites for their kids depending on their age.

An additional component is the toolbar (shown in Figure 4). It is designed for the Web browser Firefox to ensure that it can be used independent from the platform. It is a utility that facilitates the voting process of the children and the rating process of the parents. Two separate versions - one for parents and one for children - have been developed. Parents can vote for or against the currently open website, while children can rate the website. Every time a website is downloaded, a query is sent to the *SaferWeb* server to check whether or not the current user has already voted for this site. If this is not the case, the parent can vote via a positive or a negative button. Additionally, a random website from the greylist is presented to the parents. They can visit the site and vote for or against it. The children's toolbar has similar functionality but they can only rate whitelist entries.

5 Summary and Future Work

The idea of the *SaferWeb* cannot be compared to any other existing project. The unique feature is that the system works on its own because the community explores new websites and votes for or against them. Editorial costs are no longer necessary. No staff for servicing the *SaferWeb* is needed. It is flexible and dynamic, an important fact in the age of Web 2.0. The *SaferWeb* is based on a unique set of components: the server, Web frontends for kids and parents, a http-proxy and a Firefox toolbar. They all ensure a childproof Web experience for kids.
But even if the *SaferWeb* has now fixed the disadvantages of existing projects, there are some open challenges to make the *SaferWeb* more secure and more comfortable. These

visions may be realized in future versions of the *SaferWeb*.

To achieve a more secure system, *SaferWeb* ought to use a verification system to verify the age and identity of registered parents. This is important, especially when creating accounts for adults to avoid multiple registrations. In Germany, there are third-party services that verify the identity of a certain person. An example is the so-called "Post Ident" method of the German federal mail. Unfortunately, this verification cannot be part of the current *SaferWeb* version as it is not free of charge.
A current problem is represented by so-called "dummy" or fake accounts. These would allow to collect enough (fake) votes to push a website to the whitelist. In order to avoid such abuse, we extend the *SaferWeb* with IP filtering techniques and with the concept of "Super Adults". These are parents with more powerful voting rights, enabling them to immediately ban suspicious sites and send them to the grey- or blacklist, just like administrators would do.

Another important goal is to increase the joy of use as well as the identification with *SaferWeb*. The extension of the *SaferWeb* website is planned for future versions. Then it shall include features like lists of new websites, the presentation of a "website of the month", and it should also offer the latest news taken from appropriate children news websites. Parents and children can have their own voting lists, in which all websites are collected with the associated votes. Another possibility might be the integration of moderated discussion forums for different issues. It would be great if children would not only regard *SaferWeb* as a service to go online but if they would also enjoy using *SaferWeb*, talk about its possibilities with friends, and continue to learn about new websites and new features. Perhaps a mascot might help to achieve this vision. Future versions of the *SaferWeb* are to be used in other countries as well. This assumes that the interface is arranged in a specific way for each country.

6 Acknowledgements

This work is based on the results of a practical course. It would not have been possible without the work of the following people: Moritz Bartl, Daniel Schroeder, Claas Wilke, Toni Dietze, Stefan Illgen and Thomas Gerhardt.

References

[Fra] FragFinn.de. http://www.fragfinn.de.

[Glu] Glubble. http://www.glubble.com/.

[Hag97] A.G.: Hagel, J.; Armstrong. *Net Gain: Expanding Markets Through Virtual Communities*. Harvard Business School Press, 1997.

[ICE] ICEfaces. http://www.icefaces.org/main/home/index.jsp.

[Imb] Imbee.com. http://www.imbee.com/.

[Isk05] A.; Kutscher N.: Iske, S.; Klein. Differences in Internet Usage. *Social Work and Society*, pages 215–223, 2005.

[Kuh] Blinde Kuh. http://www.blinde-kuh.de/informationen.html.

[Mun99] D.: Munro, A.; Benyon. Footprints in the Snow, 1999.

[Raz] C. Razor, K.; Renfro. Youth Internet Usage Statistics.

[Rel] Whitehouse Press Release. http://www.whitehouse.gov/news/releases/2002/12/20021204-1.html.

[Saf] SafeKids.Com. http://www.safekids.com/kidsrules.htm.

[UoSC] Center for the Digital Future University of Southern California. Online World As Important to Internet Users as Real World?

[(W3a] World Wide Web Consortium (W3C). WebServices.

[(W3b] World Wide Web Consortium (W3C). WSDL Specification.

[Was05] E.: Wass. The Small, But Limitless World of .kids.us. *http://www.circleid.com/posts/the_small_but_limitless_world_of_kidsus*, pages 215–223, 2005.

[Wut01] M.: Wutka. *J2EE Developer's Guide*. Markt+Technik, 2001.