



RENDER
FP7-ICT-2009-5
Contract no.: 257790
www.render-project.eu

RENDER

Deliverable D5.1.4

Evaluation of the Tools for Diversity Management in Wikipedia

Editor:	Angelika Mühlbauer, Wikimedia
Author(s):	Angelika Mühlbauer, Wikimedia; Johannes Kroll, Wikimedia; Kai Nissen, Wikimedia
Deliverable Nature:	Report (R)
Dissemination Level: (Confidentiality)	Public (PU)
Contractual Delivery Date:	30 September 2013
Actual Delivery Date:	30 September 2013
Suggested Readers:	Tool developers, Wikipedia editors and readers, researchers
Version:	1.0
Keywords:	Diversity, evaluation, assessment, Wikipedia, assessment tools

Disclaimer

This document contains material, which is the copyright of certain RENDER consortium parties, and may not be reproduced or copied without permission.

In case of Public (PU):

All RENDER consortium parties have agreed to full publication of this document.

In case of Restricted to Programme (PP):

All RENDER consortium parties have agreed to make this document available on request to other framework programme participants.

In case of Restricted to Group (RE):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement. However, all RENDER consortium parties have agreed to make this document available to <group> / <purpose>.

In case of Consortium confidential (CO):

The information contained in this document is the proprietary confidential information of the RENDER consortium and may not be disclosed except in accordance with the consortium agreement.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the RENDER consortium as a whole, nor a certain party of the RENDER consortium warrant that the information contained in this document is capable of use, or that use of the information is free from risk, and accept no liability for loss or damage suffered by any person using this information.

Full Project Title:	RENDER – Reflecting Knowledge Diversity
Short Project Title:	RENDER
Number and Title of Work package:	WP 5 - Diversity Case Studies
Document Title:	D5.1.4 - Evaluation of the Tools for Diversity Management in Wikipedia
Editor (Name, Affiliation)	Angelika Mühlbauer, Wikimedia
Work package Leader (Name, affiliation)	Javier Caminero, Telefónica

Copyright notice

© 2010-2013 Participants in project RENDER

Executive Summary

This deliverable describes the evaluation process of the tools developed in the Wikipedia use case. These supporting tools for Wikipedia aim to help Wikipedia users to understand, to find and to cure articles which contain a lack of diversity. In the deliverable D5.1.2 “Tools for diversity management in Wikipedia” we described both tools – the Article Monitor and the Article List Generator – in detail.

We performed the evaluation process in two phases. During the first tests we collected a list of needs and requirements. Following these findings we improved and adjusted the software and the functionality. In a second testing period we evaluated the usage and the acceptance of the final released versions of the tools. Additionally, we asked for further needs and requirements to involve the participants in an active process of enhancement to broaden the tools in Wikipedia and perspective to all language versions.

The results of the final evaluation presume that both tools can be suitable instruments to support Wikipedia users. For both tools the participants and interested users required further adaptations, in particular to include the tools deeper into Wikipedia’s infrastructure.

Table of Contents

Executive Summary	3
Table of Contents	4
List of Figures	5
List of Tables	6
Abbreviations	7
1 Introduction	8
2 First Evaluation Phase	10
2.1 Methodology.....	10
2.1.1 Participants	10
2.1.2 Methods and Material	10
Thinking Aloud Protocols (TAP)	10
Questionnaires	11
2.1.3 Procedure	11
2.2 Results	11
2.2.1 Article Monitor	11
2.2.2 Task List Generator	12
2.3 Summary of Requirements and Consequences of the First Evaluation.....	13
2.3.1 Requirements for the Article Monitor	13
2.3.2 Requirements for the Article List Generator	13
2.3.3 Moving the Software Components to Wikimedia Labs	14
3 Final Evaluation	15
3.1 Methodology.....	15
3.1.1 Material	15
3.1.2 Participants	15
3.1.3 Procedure	16
3.2 Evaluation Results	16
3.2.1 Article Monitor	16
3.2.2 Article List Generator.....	20
3.3 Comparing the Overall Results of Both Phases.....	22
3.4 Requirements and Needs for the Further Implementation.....	24
4 Summary and Future Work	26
References	27
Annex A Evaluation Materials.....	28
A.1 First Evaluation Phase: Think Aloud Test Tasks and Questionnaires	28
A.1.1 Article Statistics and Quality Monitor	28
A.1.2 Task List Generator	31
A.1.3 General Questions for Statistics (Used in Both Tool Evaluations)	33
A.2 Second Evaluation Phase: Questionnaires.....	34
A.2.1 Article Monitor	35
A.2.2 Article List Generator.....	41
A.2.3 General Questions for Statistics (Used in Both Tool Evaluations)	44
Annex B Evaluation Results	46
B.1 First Evaluation Phase	46
B.1.1 ASQM – Ratings	46
B.1.2 TLG – Ratings	47
B.2 Final Evaluation Phase	47
B.2.1 Article Monitor – Ratings.....	47
B.2.2 Article List Generator – Ratings	51

List of Figures

Figure 1: Screenshot of the Article Monitor.....	8
Figure 2: Screenshot ALG interface	9
Figure 3: Experience with Wikipedia (EP1).....	10
Figure 4: Article Monitor- User Ratings (EP1).....	12
Figure 5: TLG - User Ratings (EP1)	12
Figure 6: Distribution of the User Experience in Wikipedia	16
Figure 7: Overall Estimations of the Article Monitor.....	17
Figure 8: User Ratings for the Benefit of Statistical Parameters	18
Figure 9: User Ratings for the Link Extractor	18
Figure 10: User Ratings for the News Finder	19
Figure 11: User Ratings for the Change Detector.....	19
Figure 12: User Ratings for WikiGini.....	20
Figure 13: ALG - Requests per Filter	21
Figure 14: Overall Estimations of the Article List Generator	22
Figure 15: ALG - User Ratings to Useful Filters	22
Figure 16: Article Monitor User Ratings - EP1 versus EP2	23
Figure 17: ALG User Ratings – EP1 versus EP2	24

List of Tables

Table 1: Transfer from Verbal to Numeric Assessment Scale in EP1	11
Table 2: List of Identified Requirements for the ASQM in EP1	13
Table 3: List of Identified Requirements for the TLG in EP1.....	14
Table 4: Transfer from Verbal to Numeric Assessment Scale in EP2	16
Table 5: Number of Article Monitor User Requests	17
Table 6: ALG - Number of Requests per Search Depth	21
Table 7: Article Monitor (Total) - User Ratings.....	47
Table 8: Statistics - User Ratings for Benefit	48
Table 9: Statistics - User Ratings for the Parameter "Useful for Quality"	48
Table 10: Link Extractor - User Ratings.....	49
Table 11: WikiGini - User Ratings	50
Table 12: Article List Generator (total) - User Ratings.....	51

Abbreviations

ALG	Article List Generator
AM	Article Monitor
ASQM	Article Statistics and Quality Monitor
EP1	Evaluation Period 1
EP2	Evaluation Period 2
TAP	Think Aloud Protocol
TLG	Task List Generator
WMDE	Wikimedia Deutschland
WP	Wikipedia

1 Introduction

The goal of the Wikipedia use case as part of the RENDER project is the support of users in finding, understanding and curing biased articles in Wikipedia. Several analysis approaches concerning fact coverage, currentness and objectivity and editor behaviour have been developed by Wikimedia Deutschland, KIT and JSI. The results are presented in combination with further statistical information and external analysis approaches in two supporting tools for Wikipedia - the Article Monitor and the Article List Generator.

The **Article Monitor** (see Figure 1) aims to help users to get a quick overview about the currently viewed Wikipedia article. The monitor sums up several statistics and provides different analysis approaches like fact coverage and timeliness. Further approaches or assessment results can be included very easily if required. The results by themselves don't provide an immediate quality score for the article. Rather, they have to be interpreted by the user in the context of the specific article or topic. In this way, the reader can use the information to reason about possible quality flaws and derive action items for contributing to the article or the talk page.

A short remark concerning the name of this tool: People had problems to understand the abbreviation ASQM and to find the related tab, especially when the tool has been installed yet. We decided to change the name to simply Article Monitor. So the name of the tab is renamed, too.

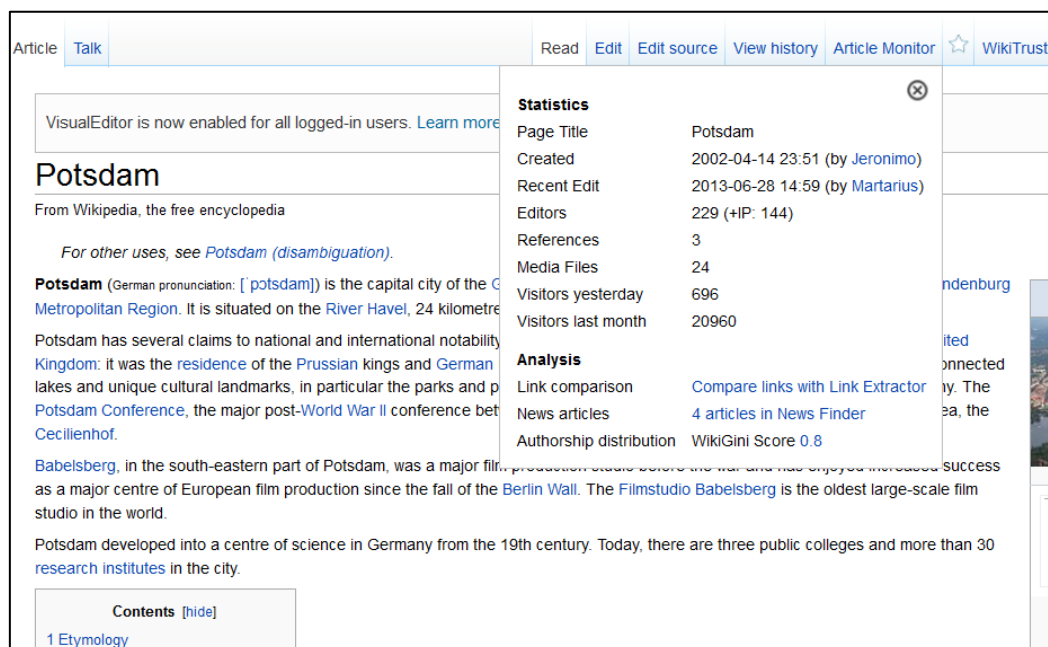



Figure 1: Screenshot of the Article Monitor


The **Article List Generator** (see Figure 2) enables authors to create individual lists of articles. The users can choose certain categories and specify filters to search (e.g. for articles without an image). Currently, the list generator works for the English, German and French Wikipedia. Further languages versions will be added in future.

A detailed description of the analysis approaches as well as the supporting tools is available in D5.1.2.

The deliverable is structured as follows. We start by describing the first evaluation phase, its results and requirements to the tools in section 2. After the preparation of the final releases of the tool we performed the final evaluation which is described in section 3. The last section contains a short summary and details about further steps.



Supporting Tools for Wikipedia



Seite auf [Deutsch](#)

Article List Generator

The Article List Generator makes it possible to search categories and compile article lists using different criteria. The query may consist of a one or more categories, the [intersection](#) or the [difference](#) of categories. The user is able to determine the depth of the search and a set of filters shown on the righthand side allows further refinement of the result. The combination of these filters is also possible. The result is displayed in the form of a list.

Language, Categories and Search Depth

Language de ⓘ

Search terms ⓘ

Separate category titles using enter or tab key.

Depth 2 ⓘ

Format HTML ⓘ ☐ by Email ⓘ

Select filters

General

- ☐ All Pages
- ☐ Article Ratings
- ☐ Large Pages
- ☐ No Links to this article
- ☐ No Images
- ☐ Pending Changes (12h)
- ☐ Small Pages
- ☐ Template: Cleanup
- ☐ Template: Technical

☐ Currentness

- ☐ Change Detector
- ☐ Template: Out of date

☐ Completeness

- ☐ Template: Globalize
- ☐ Template: Refimprove

☐ Neutrality

- ☐ Template: Neutrality

Figure 2: Screenshot ALG interface

2 First Evaluation Phase

This section of this document describes the first evaluation period, which was performed in February and March 2013. The goal of the first evaluation period was to test the usage and the functionality of the supporting tools. We wanted to understand at which points the users had problems during the usage. Additionally, we aimed to identify a list of improvement requirements and needs of the Wikipedia community.

2.1 Methodology

For the first test period we used the following methods:

- Think Aloud Protocols: The user had to perform three tasks and to comment their actions, thoughts and feelings loudly.
- Questionnaires: At the end we requested the users to fill out a questionnaire to rate the tools.

2.1.1 Participants

Nine persons participated in the first evaluation period. Figure 3 shows the experience of the testers in working with Wikipedia. Five users tested the Article Monitor; these persons were selected from WMDE and had little experience writing articles in Wikipedia. Four persons attended in the test of the TLG. These Wikipedia editors indicated to have more than three years of experience in Wikipedia.

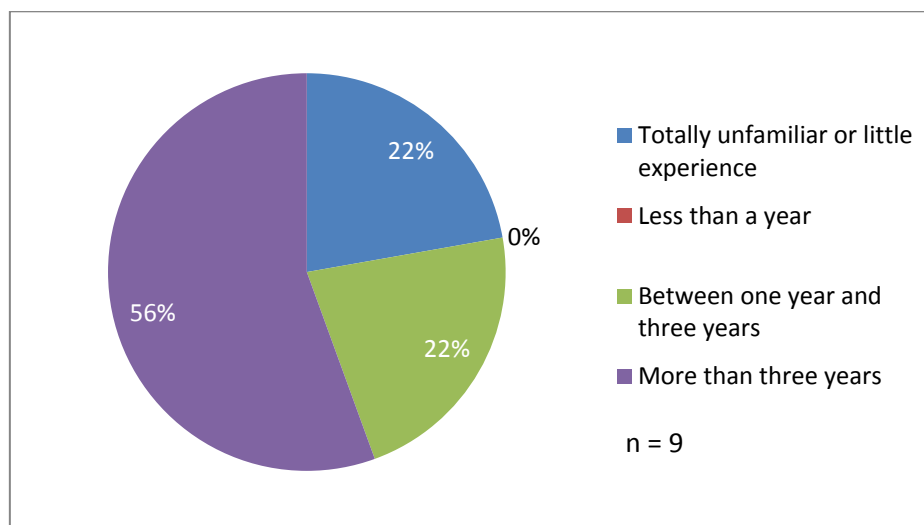


Figure 3: Experience with Wikipedia (EP1)

2.1.2 Methods and Material

Thinking Aloud Protocols (TAP)

We prepared three tasks which the participants had to perform during the test for each tool. The tasks were meant to introduce tools to the users and to focus their attention on all functional features. The list of TAP tasks is available in A.1.1.1 and in A.1.2.1.

Questionnaires

We asked the participants to answer questions concerning the performance, the efficiency and the usability. Efficiency is defined in this context as the cost-benefit ratio. Furthermore, they were asked to comment their needs or suggestions for improvements or expansions. For each supporting tool we prepared a particular questionnaire which contains some further questions concerning age, educational achievements, experiences in Wikipedia and the topics of interest.

The questionnaires are available in Annex A.

2.1.3 Procedure

The first evaluation took place in Wikimedia Deutschland's office rooms. All testers used a laptop with internet connection. For the Article Monitor test, the participants started the evaluation with the browser already opened and logged in to a specifically prepared testing account. So, it was not necessary to install the gadget to the user preferences for the Article Monitor tasks. For the TLG test the internet page containing the online form was already opened. The TAP tasks were read loudly. During the test we encouraged the participants to speak loud what they were doing and thinking. The supervisor of the tests session took minutes. At the end of the tasks we asked the users to fill out the particular questionnaire.

For the analysis of the questionnaires we transformed the verbal assessment scale into a numeric scale as presented in Table 1.

Table 1: Transfer from Verbal to Numeric Assessment Scale in EP1

Verbal scale values	Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
Numeric scale values	1	2	3	4	5	6

2.2 Results

2.2.1 Article Monitor

The participants were able to perform all tasks without further assistance from the supervisor of the test sessions. We identified several improvement requirements. The major problems occurred because of the incomprehensibility of the description text of the analysis approaches of the Link Extractor, the Change Detector and WikiGini. The loading times and performance of Article Monitor were general problems during the test period which were caused by server problems. The complete list of requirements for the tool improvement is listed in subsection 2.3.1.

The analysis of the questionnaire reflected the experiences we collected during the TAP. Figure 4 shows the rating for the Article Monitor. The users assessed 4 out of 5 requested parameters between 2.5 and 3. The best rating was given with the mean value of 2.2 for the efficiency of this tool. The worst rating was given for the parameter "Understand". We observed during the TAP that the users had problems to understand the results of the single analysis approaches in isolation. The detailed overview of all ratings given in the first phase for the Article Monitor is given in B.1.1 (Annex B).

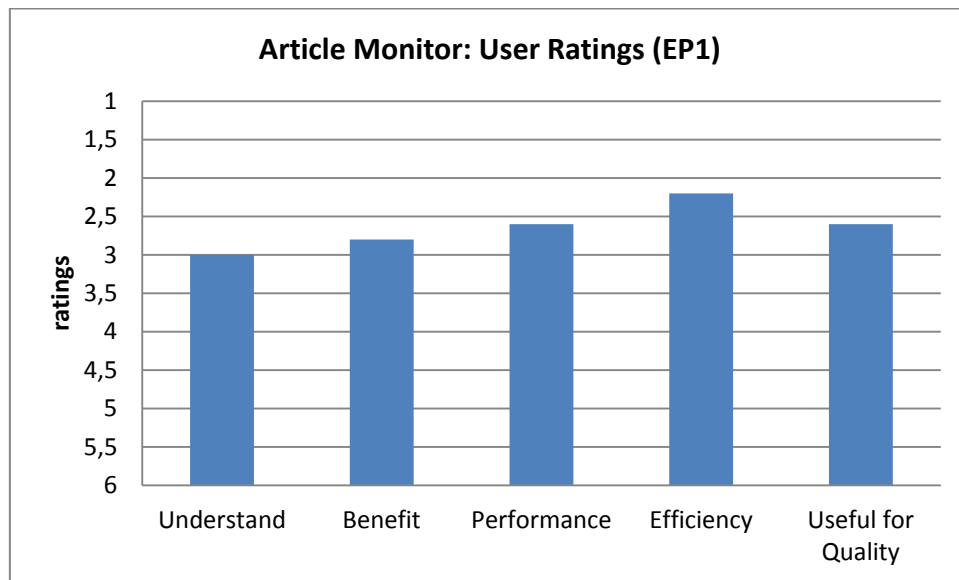


Figure 4: Article Monitor- User Ratings (EP1)

2.2.2 Task List Generator

The participants were able to use the List Generator without further assistance. The identified list of the requirements and further wishes is presented in subsection 2.3.2.

The results from the questionnaire are visualised in Figure 5. The best ratings were given with a mean value of 1.5 for the parameters “Benefit” and “Useful of Quality”. The worst rating was given for the performance with a mean value of 3.75. All testers mentioned the worth of such a tool and formulated the hope of better performance results for the next version. During the testing period major problems occurred with the Wikimedia Toolserver, so these results are not surprising for us.

A detailed overview of all ratings given in the first phase for the TLG is given in B.1.1 (Annex B).

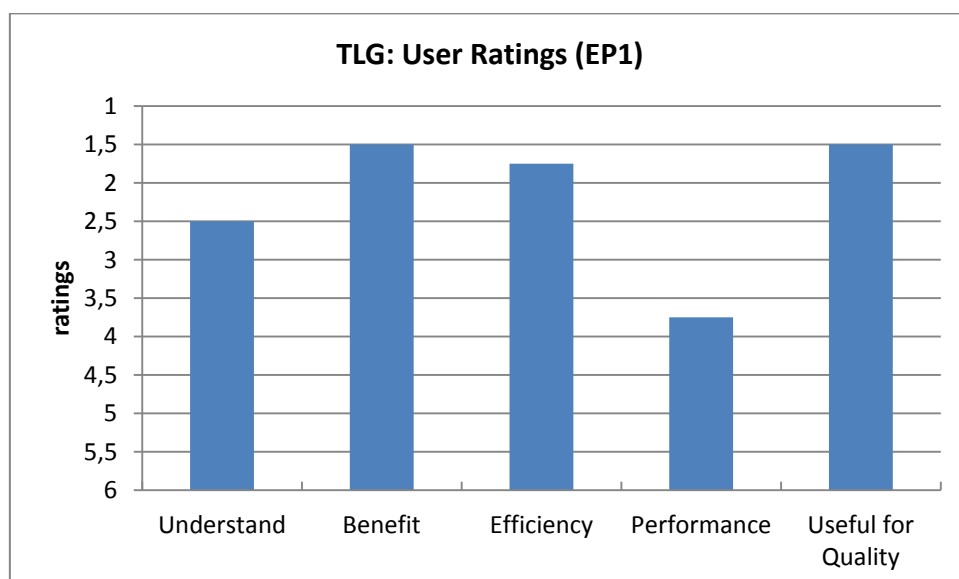


Figure 5: TLG - User Ratings (EP1)

2.3 Summary of Requirements and Consequences of the First Evaluation

During the first testing period we identified several requirements and wishes from the TAP tests and the questionnaires. Additionally, we collected this information during the local events (the RENDER tour see D6.2.5 [2]). Users also commented our RENDER Wikipedia discussion pages and used the feedback forms to contact us.

2.3.1 Requirements for the Article Monitor

In Table 2 we listed the requirements we collected for the Article Monitor. The change of the tool's name was frequently requested as we explained above.

The majority of the needs listed above we realized in preparation of the final release as we highlighted in the table, too.

Table 2: List of Identified Requirements for the ASQM in EP1

Component	Requirements	Priority *	Realisation *
ASQM	Renaming of ASQM to Article Monitor	high	yes
	Restructuring the result box and renaming sections	high	yes
	Parameters are visible if results are available	medium	yes
Statistics	Visualisation of editor numbers: Author (+ IP) and linking to user page	medium	yes
	Visitors today → Visitors yesterday Visitors last 30 days → Visitors last month	low	yes
	Number of authors within a time period	low	no
Analysis Approaches	Improvement of description texts for LinkExtractor, Change Detector and WikiGini	high	yes
	Improve the Link Extractor page: Optical separation of the example section and form section	high	yes
General	Performance problems → Moving to Wikimedia Labs	high	yes

* For the final release

2.3.2 Requirements for the Article List Generator

We decided to change the tool's name "Task List Generator" to "Article List Generator". This tool provides more functionality than to generate task lists. On the other hand we didn't want to stress out that people have to use this tool only as an instrument to work. This might transport the wrong message in the collaboration with voluntarily acting editors.

In Table 3 we listed the requirements we collected during the first evaluation phase. We also marked if the item could be solved in the preparation of the final release. The List Generator could be extended with many further filter functions and many Wikipedia editors would benefit. But we decided to insert only two additional filters which were requested in the majority of feedbacks.

Table 3: List of Identified Requirements for the TLG in EP1

Component	Requirements	Priority *	Realisation *
TLG	Renaming TLG (Task List Generator) → ALG (Article List Generator)	low	yes
	Optimisation for the progress bar	medium	no
	Further language versions	medium	yes
	No underlines in result list visualisation in WikiSyntax	low	yes
	Presentation of number of results	high	yes
Search terms	Simpler search term insertions function	high	yes
	Auto completion	high	yes
	Allow the individual watch list as search terms	medium	yes
Email function	Fixing of bugs: no error message/ results if no email address was inserted	high	yes
Filters	More precise description texts (tool tips)	high	yes
Additional Filters	No links to an article	medium	yes
	Pending article (more than 12 hours)	medium	yes
	Missing sources	medium	no
	Abnormalities in article structure, style of speech and further formal parameters	medium	no
General	Performance problems → Moving to Wikimedia Labs	high	yes

* For the final release

2.3.3 Moving the Software Components to Wikimedia Labs

The major problem of both tools was the performance of the Wikimedia Toolserver. That was noticeable in extremely long loading times and server downs. So we forced the movement to Wikimedia Labs and worked close together with the developer team of the Wikimedia Foundation.

In April and May we improved the software and functionality of the supporting tools and analysis approaches. Additionally, we decided to move the software components from the Wikimedia Toolserver to the Wikimedia Labs. This step was important to make sure that the tools perform in a more efficient, stable and available way.

3 Final Evaluation

The final evaluation period was performed in July and August 2013.

During the first evaluation phase we identified a list of requirements and needs to the supporting tools. In section 3 we listed this information. We improved the tools according to our priority calculation and published the final versions of the Article Monitor and the Article List Generator at the end of June.

In this section of the document we describe the final evaluation processed to assess the acceptance and usability of the final versions of both tools.

3.1 Methodology

3.1.1 Material

As planned and reported in D4.2.2 we evaluated in a quantitative and in a qualitative way.

We logged usage information between the release-date on June the 21 until August the 26. Besides all Article Monitor and Article List Generator requests we logged the availability and results of the analysis approaches. Furthermore, we registered if people used the offered links to further result explanations. The collected data we analysed for a quantitative estimation.

For the qualitative evaluation we used questionnaires again. We expanded the questionnaires of the first evaluation period with questions concerning the activities which users were able to perform with help of the tools. We inserted these points to understand if and how people use the tools for their daily work in Wikipedia. These assessments we used to explore the influences on Wikipedia. The results we described in D5.1.3 [2]. The complete questionnaires are attached in Annex A.2. We included several new questions to find out if the tools enable the users to edit articles and to find flaws. The questionnaires were published as Google forms, so we shared the links via mailing lists and our information pages on Wikipedia and the Meta wiki.

3.1.2 Participants

Altogether, 25 persons took part in the second test phase and filled out a questionnaire - 16 users assessed the Article Monitor and 9 users rated the Article List Generator. The users participated in the second survey did not take part in the first one, but voluntarily answered our questions after seeing the call for participation.

Since our users are located all over the world we decided to carry out the second evaluation completely online. We invited the Wikipedia community to test the tools and to rate them by filling out a questionnaire. For that we used several communication channels like mailing lists, blog entries, and information pages in Wikipedia and on Meta wiki. Furthermore we contacted people who participated in one of our local events directly per e-mail.

As Figure 6 visualizes the majority of the participants are very familiar with Wikipedia and mentioned to have experiences for more than three years.

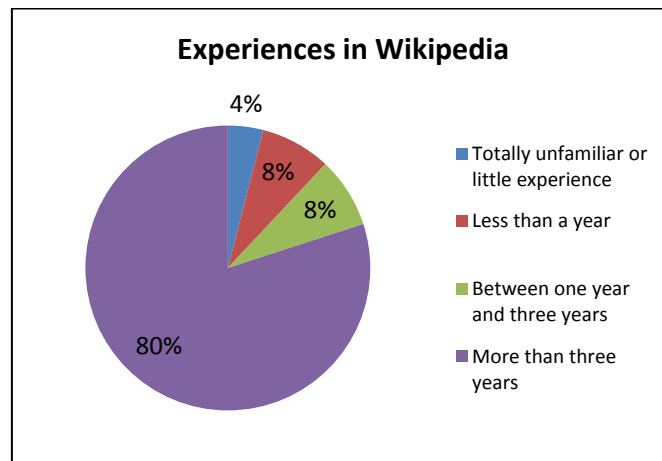


Figure 6: Distribution of the User Experience in Wikipedia

3.1.3 Procedure

The questionnaires were accessible via link to Google forms. The participants could fill out the survey anonymously. It was not necessary to leave the real name or a Wikipedia user name. But we offered a text field to leave the email address in case of being interested in the result of the evaluation.

For the analysis of the questionnaires we transformed the verbal assessment scale into a numeric scale as presented in Table 4.

Table 4: Transfer from Verbal to Numeric Assessment Scale in EP2

Verbal scale values	Very good	Fairly Good	Fairly Bad	Bad
Numeric scale values	1	2	3	4

3.2 Evaluation Results

We analysed all log data and the questionnaires.

3.2.1 Article Monitor

In the following sub-section we describe the quantitative and qualitative results for the Article Monitor.

3.2.1.1 Quantitative Evaluation Results

Installations

The Article Monitor gadget has been installed 64 times. Each user requested the Article Monitor in the mean 11 times during our testing period.

Usage and Requests

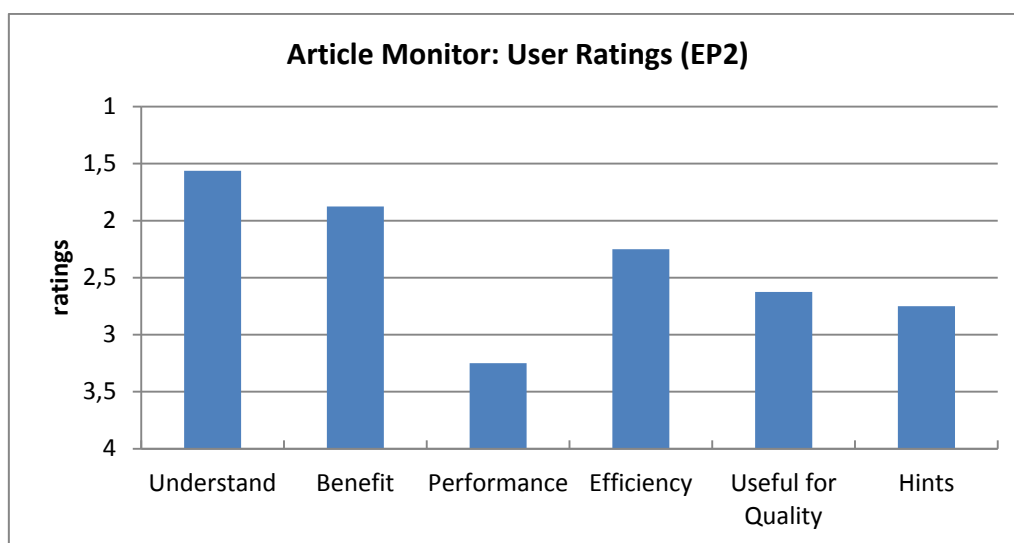
During the testing period the Article Monitor was used for 675 requests. Users requested additional information from the Link Extractor in 136 cases. We observed that 88 times users explored the WikiGini analysis and in 8 cases users clicked on the link to the News Finder result list, as shown in Table 5. Additionally, we found out that in 66 cases users came back to check for the WikiGini calculation results.

Table 5: Number of Article Monitor User Requests

All Requests	Link Extractor	News Finder	Change Detector	WikiGini
675	136	8	0	88
100 %	20,15 %	1,19 %	0	13,04 %

3.2.1.2 Qualitative Evaluation Results

We asked to assess the Article Monitor. As shown in Figure 7 the users rated the understandability and the benefit of the tool very positive. But the testers estimated a need of optimization in the performance of the tool. In Table 7 (Annex B) all user ratings are listed.

**Figure 7: Overall Estimations of the Article Monitor**

Now, we will have a look at the assessment results of the statistics and the single analysis approaches.

Statistics

We asked the users to estimate the benefit of the single statistical values and if these values are useful to assess the quality of an article. As shown in Figure 8 all parameters have a mean value over 2.5 but the parameters “Created” and “Visitors last Month” have been rated best. The users estimated that “References” and the numbers of visitors can help to assess the quality of an article. All ratings are listed in Table 8 and Table 9 (in Annex B).

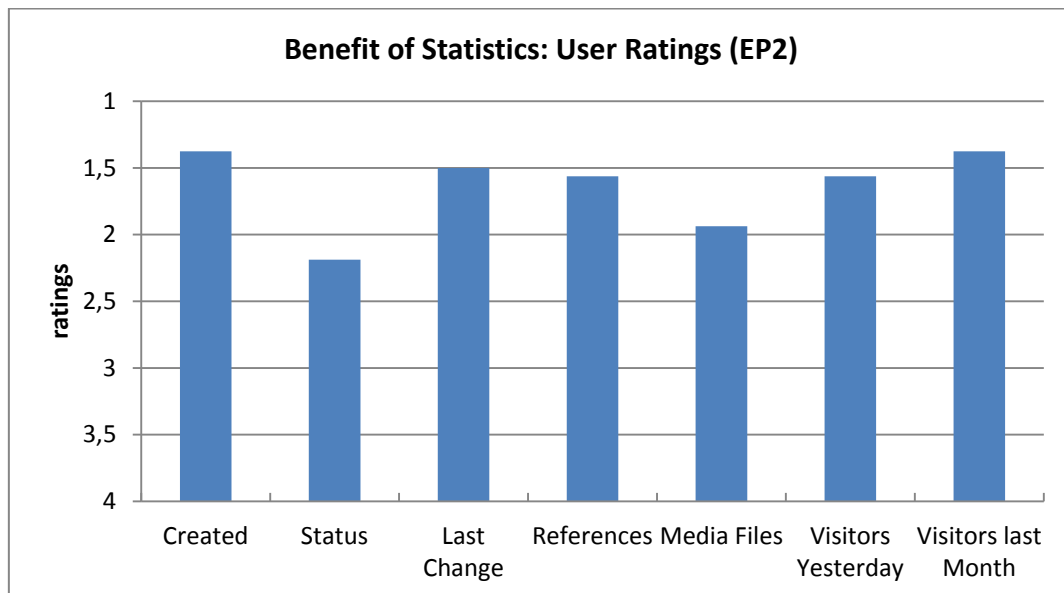


Figure 8: User Ratings for the Benefit of Statistical Parameters

Link Extractor

As visible in Figure 9, the users estimated the understandability, the benefit and the capability of the Link Extractor to give hints to contribute as better than the mean. In Table 10 (Annex B) all user ratings are listed.

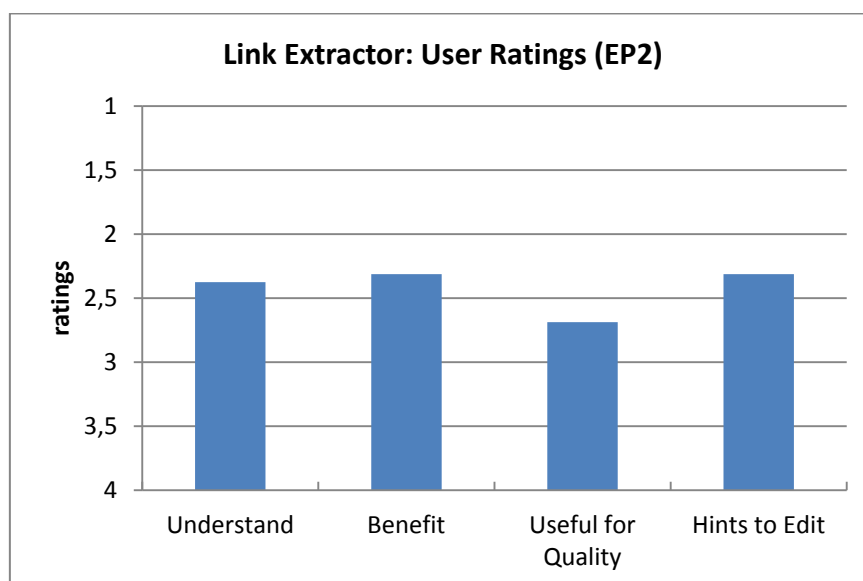


Figure 9: User Ratings for the Link Extractor

News Finder

Three of sixteen testers reported that they saw News Finder results during their usage period. Although the number of ratings is not representative, the ratings very were positive as shown in Figure 10.

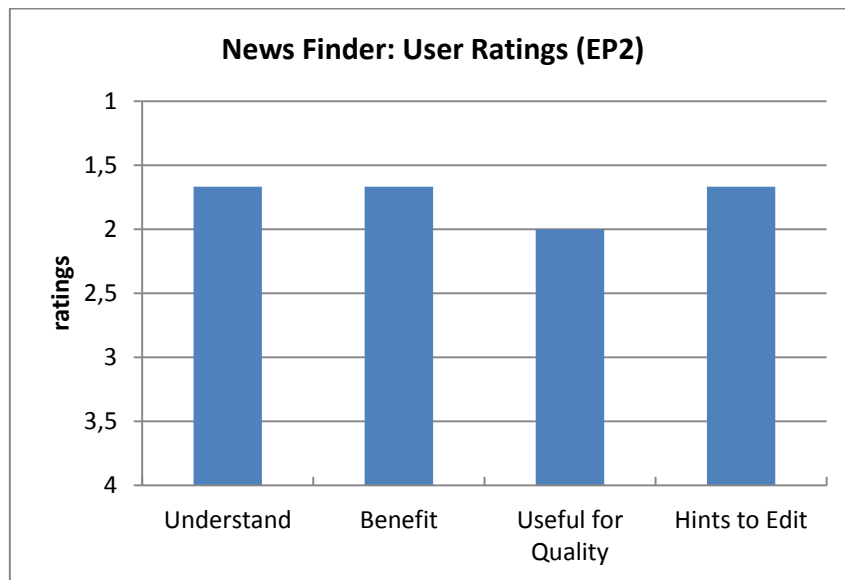


Figure 10: User Ratings for the News Finder

Change Detector

In the questionnaire four people reported the availability of Change Detector results. The ratings are positive but the parameters “Useful for Quality” and “Hints for Edits” were rated better than “Understand” and “Benefit”.

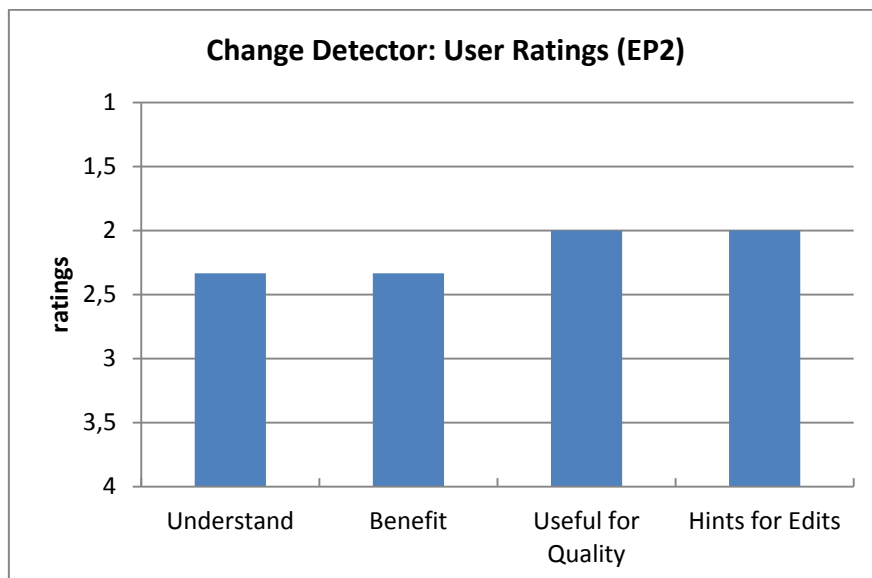


Figure 11: User Ratings for the Change Detector

WikiGini

In Figure 12 the mean ratings for WikiGini are shown. The parameters “Useful for Quality” and “Hints to Edit” got the most negative ratings. That has been expected, because WikiGini is a visualisation tool which helps users to understand the authorship of an article but is not intended to provide additional facts to contribute. The question if WikiGini as a standalone parameter can offer information that helps users to assess the article quality was discussed several times during the testing period. But in combination with other information like number of editors and requests the WikiGini value can be helpful information. All results of the ratings are presented in Table 11 (in Annex B).

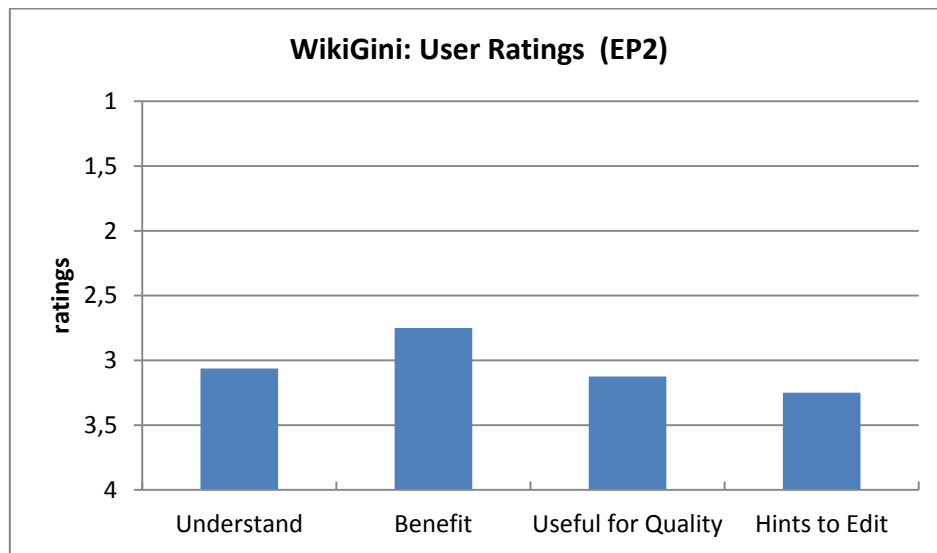


Figure 12: User Ratings for WikiGini

3.2.2 Article List Generator

This section describes the quantitative and the qualitative results we found for the Article List Generator.

3.2.2.1 Quantitative Evaluation Results

Number of Requests

Within the testing period the ALG was requested 447 times. In the majority (380) the German Wikipedia was requested. 67 times the ALG was started for the English Wikipedia.

Used Filters

Figure 13 visualizes the number of ALG requests per filter. All filters were chosen during the usage. The filter ALL was requested with the highest frequency of 139. The filter “No Images”, “Pending Changes”, and “No links to this article” are requested very often with a frequency of 121, 95 and 67. We included the filters “Pending Changes” and “No links to this article” according to the results of the first evaluation phase in the final version of the ALG. The findings confirm that the users are looking for articles which contain this kind of flaws.

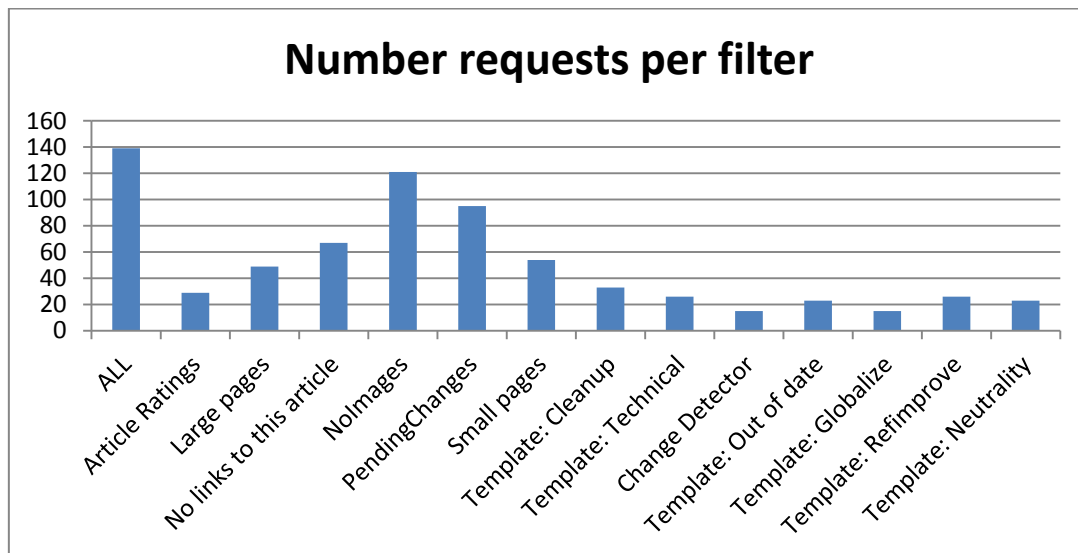


Figure 13: ALG - Requests per Filter

Search Depth

The majority of requests was performed with the search depth 2, which is the given default value. Some users tried requests up to a depth of 1,000,000,000. These values don't fit with real Wikipedia category tree layers and might be used to test the resistance and performance of the tool. In Table 6 the numbers of requests per filters is presented.

Table 6: ALG - Number of Requests per Search Depth

Search depth	Number of requests
1	51
2	274
3	26
4	22
5	23
6	3
8	14
9	1
10	6
15	1
10000	3
100000	9
1000000	2
100000000	1
1000000000	11

3.2.2.2 Qualitative Evaluation Results

The users rated the ALG positive. All parameter assessments are over 2.0. The best rating with a mean of 1.44 was giving for the efficiency of the tool. These results encourage the assumption that the ALG is accepted by the users.

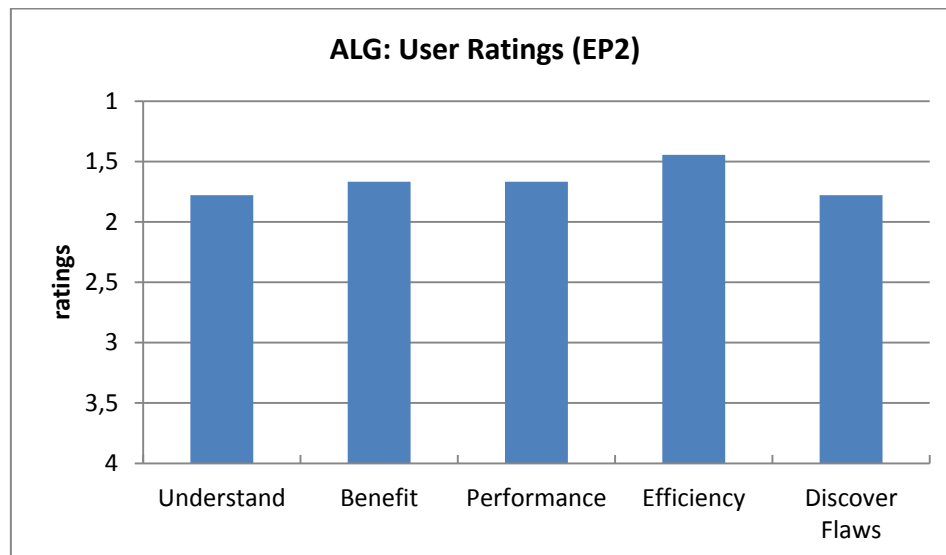


Figure 14: Overall Estimations of the Article List Generator

The most users assessed the filter “No images” as helpful. Also, the new filters “No link to this article” and “Pending changes” we included after the first evaluation phase were appraised as helpful by the majority of the participants. In Figure 15 the results are visualised.

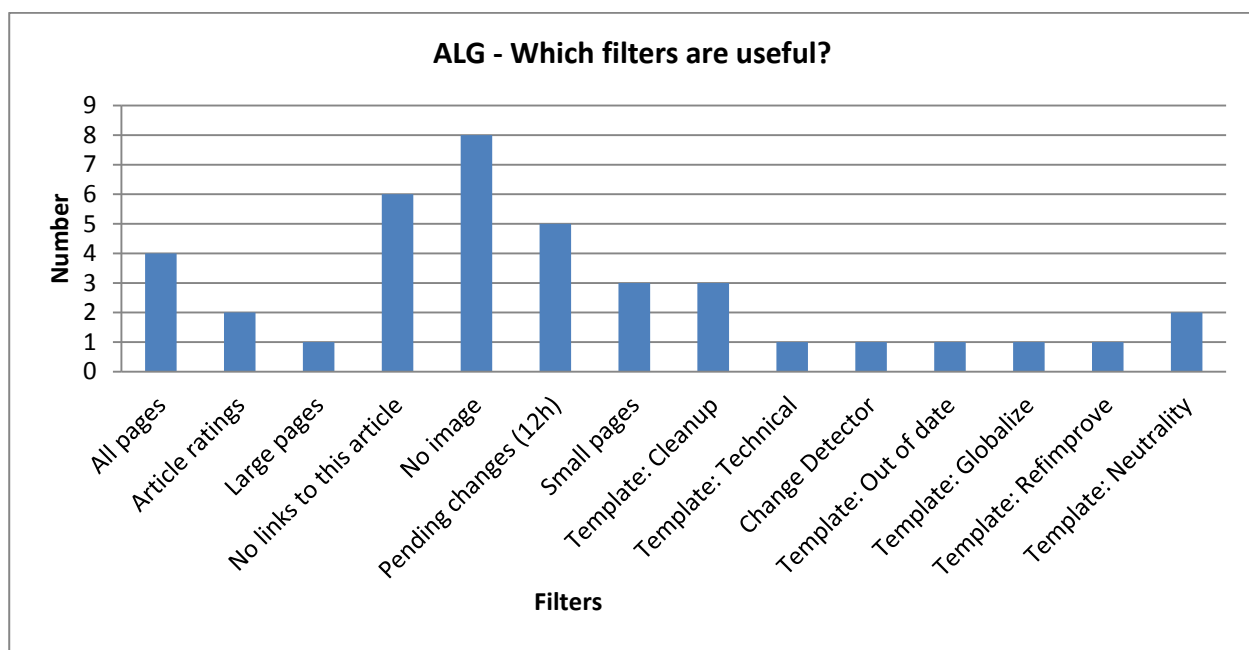


Figure 15: ALG - User Ratings to Useful Filters

3.3 Comparing the Overall Results of Both Phases

We were interested how and if the acceptance and the assessment of usability increased after the improvement and the release of the final versions of the supporting tools. We therefore compared the ratings of both testing periods. For this we transformed the scale of the first evaluation into the scale of the second round.

In Figure 16 the results of this comparison is presented. The ratings for the understandability and the benefit increased to a value of 1.56 and 1.88, but assessment of the efficiency and the performance was

worse than in the first testing phase. The efficiency decreased from 1.72 to 2.25. That means the rating was worse but is still a positive tendency.

We assume that these results are caused by the different test arrangements. While the first users sat next to the supervisor and could use loading times to ask questions about some functionalities the testers of the second test phase used the tool alone. In this situation the perception might be different. Especially for worse results for efficiency might be caused by a misunderstanding of this question.

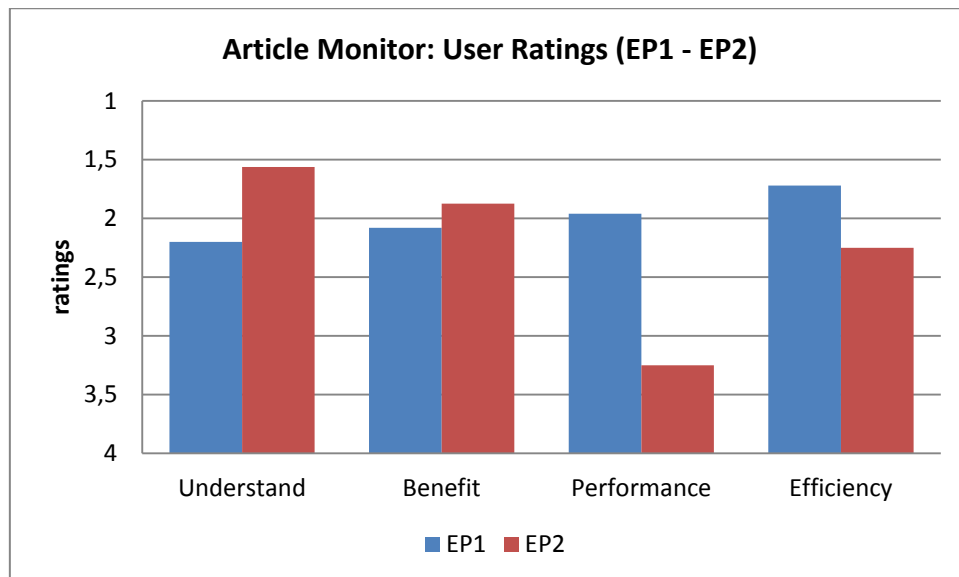


Figure 16: Article Monitor User Ratings - EP1 versus EP2

Figure 17 presents the comparison results for the Article List Generator in EP 1 and EP2. After the move of the software components to Wikimedia Labs, the performance of the ALG significantly increased. The users rated the performance of the tool much better than in the first tests. Also the understandability and the efficiency were rated slightly better than in the first evaluation phase. All parameters were assessed between 1.44 and 1.78 which means a good overall rating.

In the second test phase the benefit was assessed worse but still positive compared to EP1. This observation might be caused by the different test arrangements. While the first evaluation was carried out in a personal atmosphere the second one was held completely online. Probably people dared to be more honest in an anonymous test situation.

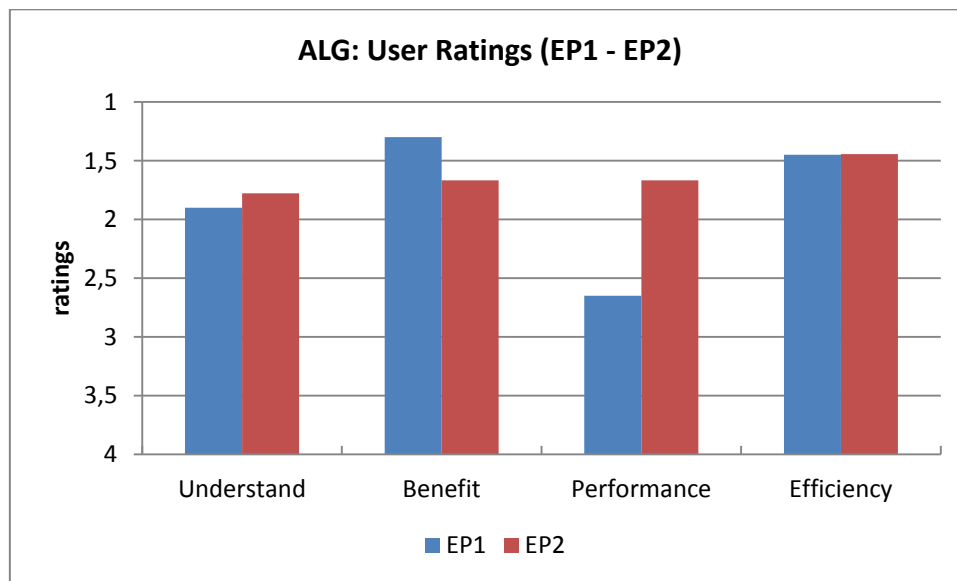


Figure 17: ALG User Ratings – EP1 versus EP2

3.4 Requirements and Needs for the Further Implementation

The participants and other users which are closely connected to the project recommended several further functionalities for both tools:

Article Monitor

- Direct inclusion as gadget
- Change “Status” to “Award” or something like this; Highlight if an article had a specific status in the past
- Better explanation for the WikiGini Score and better understandable GUI
- Many very good are written from one or a few authors
- Better explanation of the Link Extractor
- AM Very slow
- Further information where the link is coming from (article content, Navigation box, Infobox, ...)
- Fast functionality – results in the AM result window no new page
- No external tools
- Information about the amount of contributions of each author
- Number of used literature
- Number of sources

Article List Generator

The testers and users commented on the discussion pages requested some more filters but also features:

Filters:

- Articles larger/smaller than ____ byte

- Articles created ____ years ago
- Articles not edited longer than ____ hours/days/month/years
- Missing articles
- Articles for deletion in a category

Features:

- Possibility to choose and/or- relation in filter list
- More detailed usage of maintenance templates
- Expansion for further Wikipedia language versions

General:

- Inclusion in Wikipedia, no external pages
- Establishment of a centralized method which allows the automatic inclusion of the result lists in Wikipedia pages

4 Summary and Future Work

In this deliverable we described the evaluation process of the supporting tools which was performed in two phases.

In the initial evaluation phase we used Thinking Aloud Protocols and questionnaires to understand how the users work with the tools and at which stages problems or misunderstandings occurred.

After the analysis of these test results we identified a list of requirements and improvement needs.

Following these facts we expanded the tools, improved the description texts, moved the software components to Wikimedia Labs and released the final versions.

The second and final evaluation period was performed in July/August. We logged the user requests and number of installations in an anonymized way. Additionally, we invited Wikipedia users to test the tools and to answer a questionnaire.

The results of the final evaluation indicate that both tools already provide valuable information for the users for improving the articles have the potential to become highly productive instruments to support Wikipedia users if certain adaptations are made. We identified several points and requirements for the future. For both tools the participants and interested users required further adaptations (see 3.4). In particular the deeper inclusion of the tools into Wikipedia's infrastructure was requested as highly necessary.

We plan to adapt both tools for more up to all language versions. During the Wikimania in Hong Kong we discussed a pilot project with the Indian Chapter to test and adapt the tools to Indian languages.

All software components are under free licence and online available. So, we are going to improve and expand the functionalities together with the Wikipedia and the Wikimedia developer community.

References

- [1] Angelika Adam, Felix Keppmann, Delia Rusu. RENDER Deliverable D 5.1.2 – Tools for diversity management in Wikipedia. 2012
- [2] Angelika Mühlbauer, Kai Nissen, Johannes Kroll. D5.1.3 – Understanding the feedback-effects of metrics on Wikipedia. 2013
- [3] Felix Leif Keppmann, Angelika Mühlbauer, Maurice Grinberg. D6.2.5 – Report on community building activities Y3. 2013

Annex A Evaluation Materials

A.1 First Evaluation Phase: Think Aloud Test Tasks and Questionnaires

All documents of the first evaluation period were in German. Here are the translated versions.

A.1.1 Article Statistics and Quality Monitor

A.1.1.1 Tasks of the Thinking Aloud Tests

Task 1: Analyse a random article with ASQM

ASQM has been installed and provides you with more information about a Wikipedia article. Open a random Wikipedia article and start the ASQM tool. Look at the results and describe your thoughts.

Task 2: Use further analysis tools – NewsFinder and ChangeDetector

Please request the article "Benedikt XVI" in the German Wikipedia. Look at the results of the NewsFinder and the Change Detector (if available) in more detail. Click on the relevant links. Does this information help you to understand the article and to assess the quality of the article?

Task 3: Use further analysis tools – Link Extractor and WikiGini

Please request the article "Barack Obama" in the German Wikipedia. Look at the results of the Link Extractor and WikiGini. Click on the relevant links. . Does this information help you to understand the article and to assess the quality of the article?

A.1.1.2 Questionnaire

Please help us with the evaluation of the Article Statistics and Quality Monitor (ASQM). The ASQM was developed during the RENDER project by Wikimedia Deutschland. It enables Wikipedia users to quickly gain an overview about the state and quality of a Wikipedia article.

Thanks a lot for your help.

Question 1: Which information presented in ASQM is new to you?

<input type="checkbox"/>	Created ...
<input type="checkbox"/>	Last Change...
<input type="checkbox"/>	Authors...
<input type="checkbox"/>	References...
<input type="checkbox"/>	Media files...
<input type="checkbox"/>	Visitors today...
<input type="checkbox"/>	Visitors last 30 days...
<input type="checkbox"/>	LEA
<input type="checkbox"/>	NewsFinder
<input type="checkbox"/>	Wikibu.ch
<input type="checkbox"/>	ChangeDetector
<input type="checkbox"/>	WikiGini

Question 2: How do you rate the understandability of the ASQM?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 3: How do you rate the benefit of the ASQM?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 4: How do you rate the speed of the ASQM?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 5: How do you rate the efficiency (effort versus benefit) of the ASQM?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 6: How do you rate the benefit of the statistics?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 7: Does the ASQM provide suitable hints for editing Wikipedia articles?

Yes

No

<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------

Question 8: Does the ASQM help you to evaluate the quality of an article?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 9: How do you rate the understandability of the Link Extractor analysis?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 10: How do you rate the benefit of the Link Extractor analysis?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 11: Do the results of the Link Extractor help you to assess an article?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 12: How do you rate the understandability of the NewsFinder?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 13: How do you rate the benefit of the NewsFinder?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 14: Do the results of the NewsFinder help you to assess an article?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 15: How do you rate the understandability of the ChangeDetector?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 16: How do you rate the benefit of the ChangeDetector?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 17: Do the results of the ChangeDetector help you to assess an article?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 18: How do you rate the understandability of the WikiGini?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 19: How do you rate the benefit of the WikiGini?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 20: Do the results of the WikiGini help you to assess an article?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 21: Your suggestions for further functionalities**A.1.2 Task List Generator****A.1.2.1 Tasks of the Thinking Aloud Tests****Task 1: Get an overview of the Task List Generator (TLG)**

Look at the surface of the TLG. Try to explain its functionalities. There are no “right” or “wrong” answers. Simply describe your thoughts. We do not test your performance, but the tool.

Task 2: Ask a simple request to the Task List Generator (TLG)

Please insert a request to the TLG. Browse the German-language Wikipedia in category "Politics" with depth 3 for outdated articles. Choose as output format HTML. What do you notice during this performance?

Task 3: Use the output for the orientation in Wikipedia

Please, formulate a query of your choice. Choose a topic and to any filter. Look at the result of your request. Can you do something with the list of items? How would you continue to work with these issues? Can you understand the results of the TLG?

A.1.2.2 Questionnaire

Please help us with the evaluation of the Task List Generator (TLG). The TLG was developed during the RENDER project by Wikimedia Deutschland. It helps authors to create lists of articles according to their individual search criteria and certain filters to choose among.

Thanks a lot for your help.

Question 1: How do you rate the understandability of the TLG?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 2: How do you rate the the on-screen usage hints of the TLG?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 3: How do you rate the benefit of the TLG?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 4: Does the TLG provide suitable hints for editing Wikipedia articles?

Yes	No
<input type="checkbox"/>	<input type="checkbox"/>

Question 5: How do you rate the efficiency (effort versus benefit) of the TLG?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 6: How do you rate the speed of the TLG?

Very good	Good	Satisfactory	Sufficient	Unsatisfactory	Insufficient
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Question 7: Which filters are useful to create result lists?

<input type="checkbox"/>	All pages
<input type="checkbox"/>	Article Feedback Tool
<input type="checkbox"/>	Large pages
<input type="checkbox"/>	No images
<input type="checkbox"/>	Small pages
<input type="checkbox"/>	Template: Cleanup
<input type="checkbox"/>	Template: Technical
<input type="checkbox"/>	Change Detector
<input type="checkbox"/>	Template: Out of date
<input type="checkbox"/>	Template: Globalize
<input type="checkbox"/>	Template: Refimprove
<input type="checkbox"/>	Template: Neutrality

Question 8: Which filters do you need additionally?

--

Question 9: Does the TLG helps you to find articles which need to be improved in Wikipedia?

Very good

Good

Satisfactory

Sufficient

Unsatisfactory

Insufficient

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Question 10: Do you have suggestions for further functionalities?

--

A.1.3 General Questions for Statistics (Used in Both Tool Evaluations)**Question 1: How old are you?**

<input type="checkbox"/>	under 18 years
<input type="checkbox"/>	18 to 25 years
<input type="checkbox"/>	26 to 35 years
<input type="checkbox"/>	36 to 45 years
<input type="checkbox"/>	46 to 55 years
<input type="checkbox"/>	56 to 65 years

<input type="checkbox"/>	over 65 years
<input type="checkbox"/>	Prefer not to answer

Question 2: What's your highest level of educational achievement?

<input type="checkbox"/>	Not (yet) graduated
<input type="checkbox"/>	Lower secondary level
<input type="checkbox"/>	Higher secondary level (high school graduation)
<input type="checkbox"/>	University degree
<input type="checkbox"/>	Prefer not to answer

Question 3: How much experience have you gained while working with Wikipedia?

<input type="checkbox"/>	Totally unfamiliar or little experience
<input type="checkbox"/>	Less than a year
<input type="checkbox"/>	Between one year and three years
<input type="checkbox"/>	More than three years
<input type="checkbox"/>	Prefer not to answer

A.2 Second Evaluation Phase: Questionnaires

Find below the screenshots of the Google forms we used for both tool questionnaires.

We listed here only the English versions of these questionnaires.

A.2.1

Article Monitor

[RENDER] Evaluation "Article Monitor"

Please help us with the evaluation of the Article Monitor. The Article Monitor was developed during the RENDER project by Wikimedia Deutschland. It enables Wikipedia users to quickly gain an overview about the state and quality of a Wikipedia article.

If you haven't used the tool yes, install the Article Monitor and try it on a few articles. Here, you can find more information about the installation, the usage and the functionality:

https://meta.wikimedia.org/wiki/RENDER/Supporting_Tools_for_Wikipedia/Article_Monitor

At the same time or afterwards you can fill in the following questionnaire. Answering the questions takes approximately 10 – 15 minutes.

Thanks a lot for your help.

Please write to render@wikimedia.de if any questions occur.

Evaluation of the Article Monitor

First, please evaluate the Article Monitor as a whole.

You will find other separate questions in the following paragraphs.

Page 1/8

How often did you use the Article Monitor? *

- ☐ Several times a day
- ☐ Once a day
- ☐ Several times a week
- ☐ Once a week
- ☐ Several times a month
- ☐ Once a month
- ☐ Less than once a month
- ☐ I don't know

Please evaluate the Article Monitor altogether *

	very good	fairly good	rather bad
How do you rate the understandability of the Article Monitor?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the understandability of the Article Monitor?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the Article Monitor?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the speed of the Article Monitor?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the Article Monitor help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the Article Monitor help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Statistics

Please evaluate the statistics shown in the results window of the Article Monitor.

Page 2/8

Benefit of the statistics *

According to you, are the following parameters useful?

	very good	fairly good	rather bad	bad
Created	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Status	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Recent edit	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
References	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Media files	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visitors yesterday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visitors last month	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Evaluating quality with the help of the parameters *

Do the following parameters help you to evaluate the quality of an article?

	very good	fairly good	rather bad	bad
Created	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Status	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Last change	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
References	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Media files	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visitors yesterday	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Visitors last month	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Link Extractor

Please evaluate the Link Extractor - a tool that compares internal links of articles in different languages.

Page 3/8

How do you evaluate the Link Extractor Analysis *

	very good	fairly good	rather bad	bad
How do you rate the understandability of the results of the Link Extractor?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the Link Extractor	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do the results of the Link Extractor help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the Link Extractor provide suitable hints for editing Wikipedia articles?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which changes could you perform on articles due to the results of the Link Extractor? *

Multiple answers possible

- ☐ The results didn't encourage to edit an article.
- ☐ Starting a new article
- ☐ Completion of missing information
- ☐ Inserting missing links into an article
- ☐ Other:

News Finder

Please evaluate the News Finder that shows news that fit to a recently called article.

Page 4/8

Were results of the News Finder displayed while using the Article Monitor? *

If no, skip the next two questions and choose "continue"

- ☐ yes
☐ no

How do you evaluate the News Finder?

	very good	fairly good	rather bad	bad
How do you rate the understandability of the News Finder	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the News Finder?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do the results of the News Finder help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the News Finder provide suitable hints for editing Wikipedia articles?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which changes could you perform on articles due to the results of News Finder?

Multiple answers possible

- ☐ The results didn't encourage to edit an article.
☐ Starting a new article
☐ Completion of missing information
☐ Inserting missing links into an article
☐ Other:

Change Detector

Please evaluate the Change Detector - a tool that reviews if the article is up to date by checking how often it was edited in other languages.

Page 5/8

Were results of the Change Detector displayed while using the Article Monitor? *

If no, skip the next two questions and choose "continue"

- ☐ yes
☐ no

How do you evaluate the Change Detector?

	very good	fairly good	rather bad	bad
How do you rate the understandability of the Change Detector?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the Change Detector?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do the results of the Change Detector help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the Change Detector provide suitable hints for editing Wikipedia articles?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which changes could you perform on articles due to the results of Change Detector?

Multiple answers possible

- ☐ The results didn't encourage to edit an article
- ☐ Starting a new article
- ☐ Completion of missing information
- ☐ Inserting missing links into an article
- ☐ Other:

WikiGini

Please evaluate WikiGini - a tool for analysing the distribution of authors within an article.

Page 6/8

How do you evaluate the WikiGini analysis *

	very good	fairly good	rather bad	bad
How do you rate the understandability of the WikiGini result?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the WikiGini value?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do the WikiGini results help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the WikiGini analysis provide suitable hints for editing Wikipedia articles?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How do you evaluate the WikiGini analysis *

	very good	fairly good	rather bad	bad
How do you rate the understandability of the WikiGini result?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the WikiGini value?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Do the WikiGini results help you to evaluate the quality of an article?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the WikiGini analysis provide suitable hints for editing Wikipedia articles?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Which changes could you perform on articles due to the results of the WikiGini analysis? *

Multiple answers possible

- ☐ The result didn't encourage to edit an article
- ☐ Starting a new article
- ☐ Completion of missing information
- ☐ Inserting missing links into an article
- ☐ Other:

Further analyses and wishes

Page 7/8

Integration of internal tools of analysis *

How do you evaluate the integration of other tools, also from external projects like Wikibu.ch?

- ☐ very good
- ☐ fairly good
- ☐ rather bad
- ☐ bad

Which additional information would you like Article Monitor to have?

You can also leave remarks here

A.2.2

Article List Generator

[RENDER] Evaluation "Article List Generator"

Please help us with the evaluation of the article list generator (ALG). The ALG was developed during the RENDER project by Wikimedia Deutschland. It helps authors to create lists of articles according to their individual search criteria and certain filters to choose among.

You can find the ALG here: <http://tools.wmflabs.org/render/stools/alg>

Here you can find more details on the functionality: https://meta.wikimedia.org/wiki/RENDER/Supporting_Tools_for_Wikipedia/Article_List_Generator

At the same time or afterwards you can fill in the following questionnaire. Answering the questions takes approximately 5 – 10 minutes.

Thanks a lot for your help.

Please write to render@wikimedia.de if any questions occur.

Evaluation of the Article List Generator

First, please evaluate the Article List Generator as a whole.

Page 1/3

How often did you use the Article List Generator? *

- ☐ Several times a day
- ☐ Once a day
- ☐ Several times a week
- ☐ Once a week
- ☐ Several times a month
- ☐ Once a month
- ☐ Less than once a month
- ☐ I don't know

Please rate the Article List Generator as a whole. *

	very good	fairly good	rather bad	bad
How do you rate the understandability of the Article List Generator?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the benefit of the Article List Generator?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the speed of the Article List Generator?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
How do you rate the effort/benefit ratio?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Does the Article List Generator help you to find articles that need to be edited?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Filters

Please evaluate the following filters.

2/3

Which filters help you? *

According to you, which filters are very useful for the creation of a result list?

- ☐ All pages
- ☐ Article ratings
- ☐ Large pages
- ☐ No links to this article
- ☐ No images
- ☐ Pending changes (12 hrs)
- ☐ Small pages
- ☐ Template: Cleanup
- ☐ Template: Technical
- ☐ Change Detector
- ☐ Template: Out of date
- ☐ Template: Globalize
- ☐ Template: Refimprove
- ☐ Template: Neutrality

Which tasks could you carry out because you used the article list generator? *

Multiple answers possible

- ☐ Create links
- ☐ Create images
- ☐ Create own maintenance manuals
- ☐ Conduct sightings
- ☐ Review content
- ☐ Expansion of articles
- ☐ Place maintenance templates
- ☐ Delete maintenance templates
- ☐ Add citations
- ☐ Other:

Which additional filters would you like to have?

You can also leave remarks here

A.2.3 General Questions for Statistics (Used in Both Tool Evaluations)**General questions for statistics**

The following questions help to analyse the results of the questionnaire. With them, it is possible to conduct analyses according to groups (gender, age, academic career, experience with Wikipedia). The answers can and will not be used for reasoning on an individual basis. Every question gives you the possibility not to answer.

3/3

How old are you? *

- ☐ under 18 years
- ☐ 18 to 25 years
- ☐ 26 to 35 years
- ☐ 36 to 45 years
- ☐ 46 to 55 years
- ☐ 56 to 65 years
- ☐ over 65 Jahre
- ☐ Prefer not to answer

Please indicate your gender *

- ☐ Female
- ☐ Male
- ☐ Prefer not to answer

What's your highest level of educational achievement? *

- ☐ Not (yet) graduated
- ☐ Lower secondary level
- ☐ Higher secondary level (high school graduation)
- ☐ University degree
- ☐ Prefer not to answer

How much experience have you gained while working with Wikipedia? *

Indicate since when you have been editing in Wikipedia.

- ☐ Totally unfamiliar or little experience
- ☐ Less than a year
- ☐ Between one year and three years
- ☐ More than three years
- ☐ Prefer not to answer

Your field of work and interests in Wikipedia

Which main focus does your work on Wikipedia have?

I would like to be informed about the results

Please provide your e-mail address (optional), then we can inform you about the results.

Annex B Evaluation Results

B.1 First Evaluation Phase

B.1.1 ASQM – Ratings

	Q 2	Q 3	Q 4	Q 5	Q 6	Q 7	Q 8
User	Understand ASQM	Benefit ASQM	Performance ASQM	Efficiency ASQM	Benefit of the statistics	Suitable hints	ASQM Useful for Quality
user1	2	3	3	3	2	yes	3
user2	4	2	5	2	3	yes	2
user3	2	2	1	2	3	no	3
user4	4	3	2	1	4	no	2
user5	3	4	2	3	4	no	3
Mean	3	2,8	2,6	2,2	3,2	-	2,6
Standard deviation	0,89	0,75	1,36	0,75	0,75	-	0,49
Min	2	2	1	1	2	-	2
Max	4	4	5	3	4	-	3

	Q 9	Q 10	Q 11	Q 12	Q 13	Q 14
User	Understand LEA	Benefit LEA	LEA Useful for Quality	Understand NewsFinder	Benefit NewsFinder	NewsFinder Useful for Quality
user1	4	2	4	1	2	2
user2	4	2	4	1	1	3
user3	4	3	4	-	-	-
user4	5	4	5	-	-	-
user5	4	4	4	-	-	-
Mean	4,2	3	4,2	1	1,5	2,5
Standard deviation	0,40	0,89	0,40	0,00	0,50	0,50
Min	4	2	4	1	1	2
Max	5	4	5	1	1,5	3

	Q 15	Q 16	Q 17	Q 18	Q 19	Q 20
User	Understand CD	Benefit CD	CD Useful for Quality	Understand WikiGini	Benefit WikiGini	WikiGini Useful for Quality
user1	3	2	3	5	5	5
user2	-	-	-	5	3	4
user3	-	-	-	4	3	3
user4	-	-	-	4	2	2
user5	-	-	-	5	5	4
Mean	3	2	3	4,6	3,6	3,6
Standard deviation	0,00	0,00	0,00	0,49	1,20	1,02
Min	3	2	3	4	2	2
Max	3	2	3	5	5	4

B.1.2 TLG – Ratings

	Q 1	Q 3	Q 5	Q 6	Q 9
User	Understand	Benefit	Efficiency	Performance	Useful for Quality
user1	2	2	2	3	1
user2	3	2	2	4	2
user3	2	1	1	3	2
user4	3	1	2	5	1
Mean	2,5	1,5	1,75	3,75	1,5
Standard deviation	0,58	0,58	0,50	0,96	0,58
Min	2	1	1	3	1
Max	3	2	2	5	2

B.2 Final Evaluation Phase**B.2.1 Article Monitor – Ratings****Table 7: Article Monitor (Total) - User Ratings**

	Understand	Benefit	Performance	Efficiency	Useful for Quality	Hints
user1	2	2	3	2	3	2
user2	2	2	4	3	2	2
user3	2	2	4	3	3	3
user4	1	2	3	1	2	3
user5	2	1	3	1	1	1
user6	1	1	2	1	2	2
user7	2	1	4	3	3	2
user8	2	2	3	2	3	3
user9	1	2	4	3	2	2
user10	1	1	2	3	4	3
user11	2	2	4	2	2	3
user12	2	4	4	4	4	4
user13	1	1	3	1	2	3
user14	1	2	3	2	4	4
user15	2	3	4	3	3	4
user16	1	2	2	2	2	3
Mean	1,5625	1,875	3,25	2,25	2,625	2,75
Standard deviation	0,50	0,78	0,75	0,90	0,86	0,83
Min	1	1	2	1	1	1
Max	2	4	4	4	4	4

Table 8: Statistics - User Ratings for Benefit

Benefit of Statistics							
	Created	Status	Last Change	References	Media Files	Visitors Yesterday	Visitors last Month
user1	2	2	2	2	2	1	1
user2	1	2	1	3	2	2	2
user3	1	4	1	2	2	1	1
user4	2	4	2	2	3	1	1
user5	1	2	2	2	2	1	1
user6	1	1	1	1	1	2	1
user7	2	2	2	2	3	2	2
user8	1	4	3	1	1	3	2
user9	1	1	1	1	3	2	2
user10	1	1	1	1	1	1	1
user11	1	2	1	1	3	1	1
user12	3	2	2	2	2	2	2
user13	2	2	2	2	2	1	1
user14	1	1	1	1	1	1	1
user15	1	3	1	1	1	1	1
user16	1	2	1	1	2	3	2
Mean	1,38	2,19	1,50	1,56	1,94	1,56	1,38
Standard deviation	0,62	1,05	0,63	0,63	0,77	0,73	0,50
Min	1	1	1	1	1	1	1
Max	3	4	3	3	3	3	2

Table 9: Statistics - User Ratings for the Parameter "Useful for Quality"

Useful for quality							
	Created	Status	Last Change	References	Media Files	Visitors Yesterday	Visitors last Month
user1	3	3	3	2	2	3	3
user2	2	3	2	3	2	3	3
user3	2	1	1	2	1	1	1
user4	2	4	2	2	3	1	1
user5	2	2	2	1	1	1	1
user6	1	1	1	1	1	2	1
user7	3	2	3	2	3	3	3
user8	1	4	1	1	1	1	1
user9	4	1	4	2	4	2	2
user10	3	3	3	3	3	2	2
user11	1	1	1	3	3	2	2
user12	4	2	4	4	4	4	4
user13	2	2	2	2	2	2	2
user14	4	4	4	4	4	4	4
user15	1	3	1	1	3	1	1
user16	2	1	2	1	2	3	3
Mean	2,31	2,31	2,25	2,13	2,44	2,19	2,13
Standard deviation	1,08	1,14	1,13	1,02	1,09	1,05	1,09
Min	1	1	1	1	1	1	1
Max	4	4	4	4	4	4	4

Table 10: Link Extractor - User Ratings

	Understand	Benefit	Useful for Quality	Hints to Edit
user1	2	2	3	2
user2	2	3	3	2
user3	2	2	2	2
user4	3	3	3	3
user5	2	1	1	1
user6	1	1	1	1
user7	3	3	3	2
user8	2	3	4	3
user9	3	2	2	2
user10	2	2	3	2
user11	2	2	3	3
user12	3	2	3	2
user13	2	2	2	2
user14	4	4	4	4
user15	3	3	3	3
user16	2	2	3	3
Mean	2,38	2,31	2,69	2,31
Standard deviation	0,70	0,77	0,85	0,77
Min	1	1	1	1
Max	4	4	4	4

Table 11: WikiGini - User Ratings

	Understand	Benefit	Useful for Quality	Hints to Edit
user1	4	4	4	4
user2	2	3	4	4
user3	4	4	4	4
user4	2	2	2	3
user5	2	1	2	2
user6	2	2	2	2
user7	4	4	3	2
user8	2	2	2	2
user9	4	1	4	4
user10	2	2	3	3
user11	4	4	4	4
user12	4	2	2	4
user13	2	2	3	3
user14	4	4	4	4
user15	4	4	4	4
user16	3	3	3	3
Mean	3,06	2,75	3,13	3,25
Standard deviation	0,97	1,09	0,86	0,83
Min	2	1	2	2
Max	4	4	4	4

B.2.2 Article List Generator – Ratings**Table 12: Article List Generator (total) - User Ratings**

	Understand	Benefit	Performance	Efficiency	Discover Flaws
user1	2	2	2	2	2
user2	2	1	2	1	1
user3	2	3	1	2	3
user4	1	2	2	1	2
user5	3	2	2	1	2
user6	1	1	1	1	1
user7	2	2	2	3	2
user8	1	1	1	1	1
user9	2	1	2	1	2
Mean	1,78	1,67	1,67	1,44	1,78
Standard deviation	0,67	0,71	0,50	0,73	0,67
Min	1	1	1	1	1
Max	3	3	2	3	3