

Quality Check with DokuWiki for instant user feedback

Andreas Gohr
CosmoCode GmbH
Prenzlauer Alle 36
10405 Berlin, Germany
+49 (30) 814504083

gohr@cosmocode.de

Detlef Hüttemann
CosmoCode GmbH
Prenzlauer Alle 36
10405 Berlin, Germany
+49 (30) 814504072

huettemann
@cosmocode.de

Daniel Faust
Fraunhofer ISST
Steinplatz 2
10623 Berlin, Germany
+49 (30) 24306-474

daniel.faust
@isst.fraunhofer.de

Frank Fuchs-Kittowski
HTW Berlin
Wilhelmshofstr. 75a
12459 Berlin, Germany
+49 (30) 5019-3372

frank.fuchs-kittowski
@htw-berlin.de

ABSTRACT

The present work describes a DokuWiki based solution to automatically analyze the structural quality of wiki pages by gathering and evaluating predefined metrics. By providing immediate quality feedback, users are educated and the overall quality of the wiki is increased.

Keywords

DokuWiki, Structural Quality, Quality Checks, User Guidance.

1. Quality Problems in Enterprise Wikis

With introducing wikis to the workplace, many users inexperienced in writing and structuring texts are made editors. Additionally with a growing number of pages and users, different ways of organizing and structuring the wiki arise. Both facts lead to various quality problems [1] in basically three categories.

First there are problems with page content itself. Eg. Content can be outdated, wrong, inaccurate or obsolete. Pages can also be too long to comprehend or too short to justify a own page.

Secondly problems with accessing the content can arise. Typical are missing overview pages, dead ends (pages with few or no outgoing links) and orphaned pages with no incoming links. Missing page titles fall into the same category.

The third category is interlinking problems. Link relations can be broken (eg. links to missing or wrong pages) or related content might be scattered over different pages. Content redundancy with the same information on different pages is problematic as well.

This paper describes a DokuWiki based solution to check the wiki for these problems and provide users with quality feedback to support their writing process.

2. "Quality Check" Goals and Requirements

To help users with editing wiki pages, pages should be checked against certain rules automatically. These checks are to be done on a structural and statistical level. Judging the content itself remains to the users.

The automatic check should not enforce certain structures but merely guide users by providing immediate feedback.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WikiSym '10, July 7-9, 2010, Gdańsk, Poland
Copyright © 2010 ACM 978-1-4503-0056-8/10/07... \$10.00

Users shall remain free to decide if and how to fix possible problems, thus keeping the benefits wikis provide over traditional systems. This is also expressed in naming the functionality "Quality Check", not "Quality Assurance".

Beside a general quality benchmark, detailed information and tips on how problems can be solved have to be given. This information shall be available to readers and editors, to let the latter learn from their mistakes and to encourage the former to participate in improving the wiki pages.

Quality assessments should be available on the individual page as well as on namespace overviews to help wiki gardeners to quickly identify problematic areas.

3. Possible Quality Quantifiers

To automatically detect quality problems in a wiki, certain metrics need to be examined and compared to reference criteria.

A large number of possible metrics to analyze the quality of wiki content were already proposed in previous work (i.e. [2], [3], [4]).

The algorithm and the result of the calculation of metrics can have a different complexity. They range from simple numbers and arithmetic operations (e.g. count of page editors) to complex, computed values (e.g. the level of collaboration in the wiki). Created in different contexts, these metrics can be useful for answering different quality questions.

The metrics can generally be computed by measurements in the following contexts:

- a single page itself
- the edit history of a page
- the link network between the various pages
- usage statistics
- data mining and content clustering

As stated in the previous section, our approach focuses on direct user feedback. This influenced our selection of actually analyzed metrics.

Since feedback should be almost instant, we selected metrics that can be acquired without complex computations. We mainly chose metrics available from analyzing a single page, because these are the ones that can be directly influenced by editing the current page at hand. Finally, we had to choose metrics that deliver deterministic quality assessments without a need to comprehend the content of the page.

4. QC plugin for DokuWiki

Our previous deliberations lead to the creation of the "QC DokuWiki plugin". To analyze a page, the plugin implements a renderer. DokuWiki parses wiki syntax into a machine readable

intermediate state called “instructions“. Renderers then can convert these instructions to arbitrary output formats.

The QC plugin uses the render mechanism to analyze the page structure and stores the results for later utilization. Implementing a general renderer makes it possible for DokuWiki syntax plugins to feed additional Quality Check tests to the QC plugin. This makes the plugin easily adjustable to the special DokuWiki set-ups in different companies.

The various checks result in “problem scores“. Depending on how serious we consider a found problem to be, a different score is assigned. Some checks simply assign fixed scores, e.g. a missing first headline adds 5 points to the score. Other checks add a small number of problem points to the score on each occurrence. E.g. each invalidly nested headline will add an additional point.

The score of a check is the result of a comparison of a metric with a predefined value. For example, the “Too many Headlines“ check compares the “headlines / text length ratio“ with the chosen lower limit of 200 and assigns one problem point when the limit is not reached.

Comparison values and assigned scores were chosen based on personal experience and literature review. Values will be adjusted over time based on user feedback, evaluation and further research.

From the previously identified possible quality quantifiers, currently 18 different checks are implemented. Namely we check for the number of FIXMEs, if a main headline is present, for too many main headlines, incorrectly nested sections, too many horizontal rules, too many forced line breaks, deeply nested quotes, if the document has a single author only, for very small or very large documents, too many or too few headlines, missing wiki links, many links to non-existing pages, too much or long text formatting and missing backlinks.

A special case is FIXME. In DokuWiki, the string 'FIXME' creates a graphical symbol which can easily be spotted in the text. This mechanism is commonly used to indicate missing or wrong content at a certain point in a wiki page. This makes it an important mechanism to actually judge the content of a page. Each occurrence of a FIXME increases the problem score of the document.

As mentioned above, the problem score is indicated in an image on the assessed page. Additional to the score, the number of FIXMEs is displayed separately.



Figure 1. Aggregated quality score

Clicking the image opens an info window. This window is populated through an AJAX request and shows some statistical data on the page and detailed explanations on what quality problems were identified and how they could be fixed.

Statistical data includes the following information:

- Creation date
- Last modification date
- The top five authors
- The number of edits the page received
- The number of characters in the page
- The number of words in the page

Figure 2. Quality problems

To give wiki gardeners an easy way to find problematic areas in the wiki, the problem score and number of FIXMEs for each page are also shown in DokuWiki's index view.

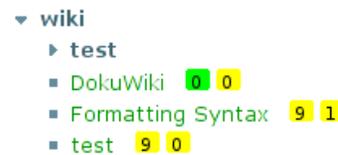


Figure 3. Index view with quality score

Since quality checks may not be reasonable on certain pages, e.g. pages that create their content mainly through plugins, it can be disabled on these pages through the simple syntax `~~NOQC~~`.

5. Future Work

The Quality Check plugin was developed as part of the ICKE 2.0 research project. Part of that project is a roll-out of the new DokuWiki-based collaboration platform, including the QC plugin to our three pilot user companies in late summer 2010. Parallel to the project roll-out, the plugin was released to the general public at the WikiCamp at CeBIT 2010.

For the prototype implementation, we selected relevant metrics from different areas (link check, text structure, text style etc.).

Feedback provided by pilot users and interested public users will also be used to enhance and improve implemented tests and comparison values.

6. References

- [1] Treitz, M. 2007. Characterizing Proliferation in Enterprise Wikis using Anti-Patterns. Bachelor Thesis, International University in Germany (IU).
- [2] Sonovis 2009. Sonovis Metric Portal. http://www.sonovis.org/wiki/index.php/Metric_Portal.
- [3] Ebersbach, A. Krimmel, K. and Warta, A. 2008. Auswahl und Aussage von Kenngrößen innerbetrieblicher Wiki-Arbeit. In: Alpar, P. and Blaschke, S. (eds.): Web 2.0 - Eine empirische Bestandsaufnahme, Vieweg+Teubner, Wiesbaden, 131-156.
- [4] Blumenstock, Joshua E. 2008. Size matters: word count as a measure of quality on Wikipedia. Proceeding of the 17th international conference on World Wide Web, 1095-1096..