The Wiki Approach to an Online Methods Resource

Frank E Harrell Jr
Vanderbilt University School of Medicine
Department of Biostatistics

March 24, 2009

1   Introduction

The most well-known *wiki* site is `wikipedia.org`, a huge and ever-growing online encyclopedia. Wikipedia defines a *wiki* as "an online content management system designed to facilitate the collaborative editing of documents through an in-browser text interface, which are then immediately presented as publicly visible web pages." *wiki* is a Hawaiian word meaning "quick." The first wiki software was `WikiWikiWeb` by Ward Cunningham in 1994, who described it as "the simplest online database that could possibly work." Many previous failed attempts at implementing the features found in wikis were proprietary. Cunningham's model was open source, which led to rapid improvements and enhancements by the user community who then put their improvements back into later open source implementations. Many companies have replaced their static intranets with wikis.

Content in wikis is created or edited using WYSIWYG editor that renders wiki markup, or by entering the simple markup directly, which is much faster. Here is an example (using the style in `foswiki.org`).

```
---+Multiple Regression
The following provides a brief introduction to *linear models* which
are implemented in a wide variety of statistical software such as =R=,
=Stata=, =SPSS=, and =SAS=.  See
http://biostat.mc.vanderbilt.edu/BioMod for more information.
---++Simple Case: One Predictor
   * Simplest possible case is one predictor that behaves linearly
   * Why should we expect this to fit real data?
      * Unusual for relationships to be linear
---++General Case: p Predictors
   * Can allow for a number of predictors
   * Predictors can be expanded into multiple columns to relax linearity assumption
      * polynomials
```

```
* piecewise polynomials (SplineFunctions)
```

The rendering of this markup can be seen below and at `http://biostat.mc.vanderbilt.edu/twiki/bin/view/Sandbox/WebHome`. You can see that the link to the URL was made automatically, and that the asterisks represent different indent levels for bullet points. `SplineFunctions` is a "WikiWord" that automatically generates a new page underneath the current page if it does not already exist[1]. If you click on the bottom button to view the raw text for the rendered page you will see the markup along with other examples. At the top of the page on the above URL is the directive `%TOC%` which treats all lines beginning with `---+`, `---++`, or `---+++` (representing progressively smaller section titles) as section headings that are hyperlinked.

Multiple Regression

The following provides a brief introduction to **linear models** which are implemented in a wide variety of statistical software such as **R**, **Stata**, **SPSS**, and **SAS**. See http://biostat.mc.vanderbilt.edu/BioMod for more information.

Simple Case: One Predictor

- Simplest possible case is one predictor that behaves linearly
- Why should we expect this to fit real data?
    - Unusual for relationships to be linear

General Case: p Predictors

- Can allow for a number of predictors
- Predictors can be expanded into multiple columns to relax linearity assumption
    - polynomials
    - piecewise polynomials (SplineFunctions[2])

Once pages are created, wikis allow easy navigation between pages and extensive search capabilities.

It is trivial to attach data, images, `pdf`, and other binary files to any wiki page. This is intended for storing content that is in final form not to be edited by users.

2 Advanced Features

Wikis have been extended in a number of ways using plug-ins for such things as inline spreadsheets, slide shows, and action items decided at meetings[2]. At `biostat.mc.vanderbilt.edu` we have implemented plug-ins for the following tasks, several of them demonstrated at `biostat.mc.vanderbilt.edu/ColeBeck`:

1. using LaTeX to render mathematical expressions inside wiki markup

---

[1] If the new topic does not already exist, a question mark appears after the word which the user clicks on to create the new sub-page.

[2] `Foswiki`, for example, will send automatic e-mail reminders once action items are entered into meeting minutes being composed during the meeting.

2. linking to `PubMed` by a simple directive, given the `PMID`, or automatically generating an article's abstract inside the current wiki page

3. using `Graphviz` to draw tree-like diagrams using simple directed graph markup

4. calling out bibliographic citations in a `BibTeX` database

5. conditional appearance of text inside the current page

As an example of the latter feature, a checkbox can be added to a page such that when checked, a whole set of optional text/tables instantly appears, making the current page longer. This is useful for example when writing a generic section with a checkbox labeled `Click here to see common randomization procedures if the study you are designing is a randomized trial`.

3 The Wiki Philosophy

Wikis are designed to trust users with appropriate permissions, making it easy to make and to correct mistakes. There is an audit trail showing who changed what to what when. Inappropriate changes can readily be edited again, or the page can be reverted to a previous version before the changes. The entire version history is saved for each page[3]. A user can elect to be notified via e-mail if a selected page, or any page, is changed by anyone.

As detailed in an excellent book by a team of computer scientists and organizational psychologists, Ebersbach, Glaser, and Heigl [1], the success of wiki comes from its conformance with how work gets done when large groups of volunteers are involved. If a leader makes assignments to group members, a member's assignment may not be in line with their interest, and work does not progress. When a member self-selects an area by contributing somewhat at random to different parts of the wiki, significant multi-author contributions sum to result in extensive content. Wikis are not designed for aesthetics but for content and removing hindrances to multi-authorship.

Perhaps the real secret to the success of wikis is that the majority of humans put their energy into things they know something about. When someone composes content that is not very good, or changes someone else's content, other knowledgeable persons come along and fix the content. Eventually the less knowledgeable person "loses steam". A case in point is the `WikiPedia` entry for the Wilcoxon two-sample test at `http://en.wikipedia.org/wiki/Mann-Whitney_U`. When I first viewed this page, it contained much misinformation including a claim that a bird scientist invented the test in 1970. I fixed the page. Someone else came along and undid half of my fixes. Others repaired most of that half, and then I fixed the rest and this has stood for many months. The persons who contributed the misinformation have lost energy.

---

[3]For efficiency, only text differences are stored, not the entire current text.

A major component of the wiki approach is *refactoring* of text. Unlike an ever-lengthening e-mail that is sent from colleague to colleague for comments, the wiki approach is to have each contributor refactor the text on a page when new contributions that are made that obsolete old text. A wiki does not accumulate all thoughts on a subject, but keeps the good and unique ones. Old thoughts are saved in the version history.

4   Strengths and Weakness of the Wiki Model

Like the successful open source movement typified by the R, Linux, Emacs, Apache, and LaTeX projects, wikis thrive on controlled chaos. At the expense of removing barriers to contributions for a large user community, bad content may appear from time to time. More harmful though is the possibility for lack of contributions in certain important areas.

When topics needing to be developed are still rapidly evolving, the wiki approach has its greatest benefits. Any journal article can quickly become obsolete or there may be difficulties putting forward links in articles to new articles, software, etc. It is extremely easy to update any wiki content using a standard web browser.

Contributors to a wiki are rewarded primarily by satisfaction from assisting a large number of researchers and from using the current content in teaching. Contributors' names can appear at the bottom of each page, but more commonly only appear when a user looks at the revision log for the page. More formal recognition could be given by listing significant contributors on a consortium's home page. Prizes for most contributions could be given as an incentive, but I am not familiar with any such example.

While page views are one reflection of the value of a page, wiki contributions do not go into a citation index, although `Google Scholar` will find wiki content (even content inside `pdf` attachments to a wiki page). A curriculum vitae could list content that is primarily due to a faculty member and could include the number of page views, but I do not know how promotion committees view this.

5   Resources Required to Maintain the Wiki

Open source wiki management software such as `Foswiki`[4] are easy to maintain. If IT personnel are already managing a server, managing wiki software is not a very large additional task. A large `Foswiki` system can be maintained by about one-tenth of a full time IT staff member.

---

[4] `Foswiki` started as `Twiki` which was an open source project that was attempted to be commercialized by its founder. The attempt led to community developers abandoning the project which is now virtually dead. Fortunately, `Twiki` was open source so the user community quickly re-formed around the excellent `Foswiki` project to which our department is currently converting. A test version may be seen at `biostat3.mc.vanderbilt.edu` which in May 2009 is expected to become `biostat.mc.vanderbilt.edu`.

Many wikis exist without an editor or webmaster, but for the online resources we are envisioning, an editor would be of help especially in soliciting content, watching that newly created pages are not redundant, creating formatting templates for new pages, monitoring overall quality, adding new links to other sites, and publicizing the site. This would take less time than editing a traditional journal.

6 Types of Content for Biostatistics in Biomedical Research

1. Good statistical practice (not specific to any biomedical technology)

2. Recommended statistical methods for each technology (assays, flow cytometry, mass spectrometry, gene microarray, confocal microscopy, fMRI and other image analysis, etc.)

3. Links to reporting guidelines for each technology (e.g., a recent article on reporting of flow cytometry experiments)

4. Recommended software for specific technologies (e.g., flow cytometry package in R)

References

[1] A. Ebersbach, M. Glaser, and R. Heigl. *Wiki Web Collaboration.* Springer, New York, 2005. ISBN-10: 3540259953 ISBN-13: 978-3540259954.