



SOFTWARE TOOL ARTICLE

WordCommentsAnalyzer: A windows software tool for qualitative research [version 1; peer review: 1 approved, 1 approved with reservations]

Ehsan Abdekhodaie ¹, Javad Hatami ¹, Hadi Bahrami Ehsan ¹, Reza Kormi-Nouri ²

¹Department of Psychology, University of Tehran, Tehran, Iran

²Center for Health and Medical Psychology, Örebro University, Örebro, Sweden

v1 First published: 03 May 2018, 7:536 (<https://doi.org/10.12688/f1000research.14819.1>)

Latest published: 04 Sep 2018, 7:536 (<https://doi.org/10.12688/f1000research.14819.2>)

Abstract

There is a lack of free software that provides a professional and smooth experience in text editing and markup for qualitative data analysis. Word processing software like Microsoft Word provides a good editing experience, allowing the researcher to effortlessly add comments to text portions. However, organizing the keywords and categories in the comments can become a more difficult task when the amount of data increases. We present *WordCommentsAnalyzer*, a software tool that is written in C# using .NET Framework and OpenXml, which helps a qualitative researcher to organize codes when using Microsoft Word as the primary text markup software. *WordCommentsAnalyzer* provides an effective user interface to count codes, to organize codes in a code hierarchy, and to see various data extracts belonging to each code. We illustrate how to use the software by conducting a preliminary content analysis on Tweets with the #successfulaging hashtag. We hope this open-source software will facilitate qualitative data analysis by researchers who are interested in using Word for this purpose.

Keywords

Computer assisted qualitative data analysis software, Microsoft Word, comments, coding, thematic analysis, code hierarchy tree

Open Peer Review

Reviewer Status

	Invited Reviewers	
	1	2
REVISED		
version 2 published 04 Sep 2018		report
version 1 published 03 May 2018	report	report

- 1 **Ronggui Huang**, Fudan University, Shanghai, China
- 2 **Yazdan Mansourian** , Charles Sturt University, Bathurst, Australia

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Javad Hatami (hatamijm@ut.ac.ir)

Author roles: Abdekhodaie E: Software, Writing – Original Draft Preparation; Hatami J: Methodology, Supervision; Bahrami Ehsan H: Methodology, Supervision; Kormi-Nouri R: Supervision

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2018 Abdekhodaie E *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Abdekhodaie E, Hatami J, Bahrami Ehsan H and Kormi-Nouri R. **WordCommentsAnalyzer: A windows software tool for qualitative research [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2018, 7:536 (<https://doi.org/10.12688/f1000research.14819.1>)

First published: 03 May 2018, 7:536 (<https://doi.org/10.12688/f1000research.14819.1>)

Introduction

Commercial qualitative data analysis (QDA) software tools such as NVivo and Atlas.ti seem to be the most popular in the qualitative research community¹. However, learning to use these complex software tools may be inconvenient for some researchers. Moreover, the purchase of commercial QDA software may not be affordable for some researchers. On the other hand, free or open-source solutions that are available often do not provide a smooth editing and markup experience (e.g., QDA Miner Lite does not support Persian and Arabic languages; CATMA and CAT² are not fast due to their web-based nature). For these reasons, some researchers use professional word processing programs for their qualitative research projects.

The use of Microsoft Word for QDA is commonly documented^{3,4}. Using Word comments provides a straightforward way to annotate specific portions of the text and attach keywords or categories (codes) to them. However, as the amount of data grows, organizing codes in Word comments becomes an exhausting task.

In this article, we present *WordCommentsAnalyzer*, a free, open-source tool that makes it possible for qualitative researchers to automate organization of the qualitative codes through a fast and easy-to-learn user interface while coding the textual material using Microsoft Word as a professional, familiar word processing software.

Methods

Implementation

This software is written in C# programming language using .NET Framework 4.5.2. The software also makes use of OpenXml library to extract comments from Word documents. Recent versions of Word store documents in XML format. OpenXml provides an easy way to query comments from a document. To facilitate assigning multiple codes to a piece of text, we assume a simple convention: different codes are entered in a comment with line breaks between them (as the descendant paragraphs of the comment element). The software uses a relational model approach to store the extracted codes and uses language integrated queries to collect different text portions related to each code, to calculate the code frequencies and to sort the codes by frequency. The visual interface of the program consists of three side-by-side panels (Figure 1). The left panel shows the codes in the comments with their counts, the middle one provides a code tree that the user can intuitively organize their codes in and the right panel shows the data extracts pertaining to each code. In the left panel, the code list can be filtered to find specific codes. The user can place codes in the code hierarchy simply by using drag-and-drop. The tree also enables the user to move codes in the hierarchy if needed. The user can introduce a new parent code or a code that is of a higher level of

abstraction. Additionally, the codes are changed or combined by being wrapped in new codes. The code hierarchy tree is saved as a tab-indented text file in the data folder (codehierarchy.txt). The tree is auto-saved every minute and can also be manually saved by clicking a save button in the interface. The previous tree files are backed up in a subfolder of the data folder.

Operation

The requirements for this software are Windows 7 or later and .NET Framework 4.5.2. After installing the .NET Framework, the user can unzip the release package from the [GitHub link](#) and run the “WordCommentsAnalyzer.exe” executable file. The program supports XML Word documents (using the .docx extension). Older Word documents (using the .doc extension) can be easily converted to XML documents by Word 2003 or later (there are also resources available on the web to batch-convert older Word documents). The program allows multiple Word files to be analyzed. This feature can be utilized to separate transcripts of different interview or focus group sessions into different files.

Use case

To illustrate how to use the software, we present a mini-study of Twitter’s Tweets from 17 January 2017 to 10 April 2018. The Tweets with the #successfulaging hashtag were copied into two Word documents based on the year in which the Tweets were posted (Supplementary File 1). We reviewed the Tweets and added comments (line-break-separated codes) to portions of texts containing interesting notions related to successful aging. Two examples of these text portions are reproduced in Figure 2.

After adding comments to Word documents, we run *WordCommentsAnalyzer*, select the folder containing the Word documents and click the *Analyze* button. The program analyzes the comments and shows a list of codes with their counts in the left panel. The middle panel enables us to organize the codes by placing them in a code hierarchy (Figure 3). For example, we can find a number of codes related to health by filtering the code list by the word of “health”. Then we add the code of “Health”, which is a parent code, to the hierarchy by dragging and dropping it onto the root node of “Code Hierarchy”. The codes of “Brain health”, “Physical health”, and “Health care” can then be drag-and-dropped onto the node of “Health”. Likewise, “Oral health” is inserted into “Physical health”. When organizing the codes, we could check the right panel to assure the data extracts support the codes. Also, the codes inserted into the hierarchy will be highlighted in the code list to help keep track of the organized codes.

Figure 4 presents a formatted version of codehierarchy.txt (Supplementary File 2) when we organized the Tweet codes with at least two counts. As shown in this figure, the themes

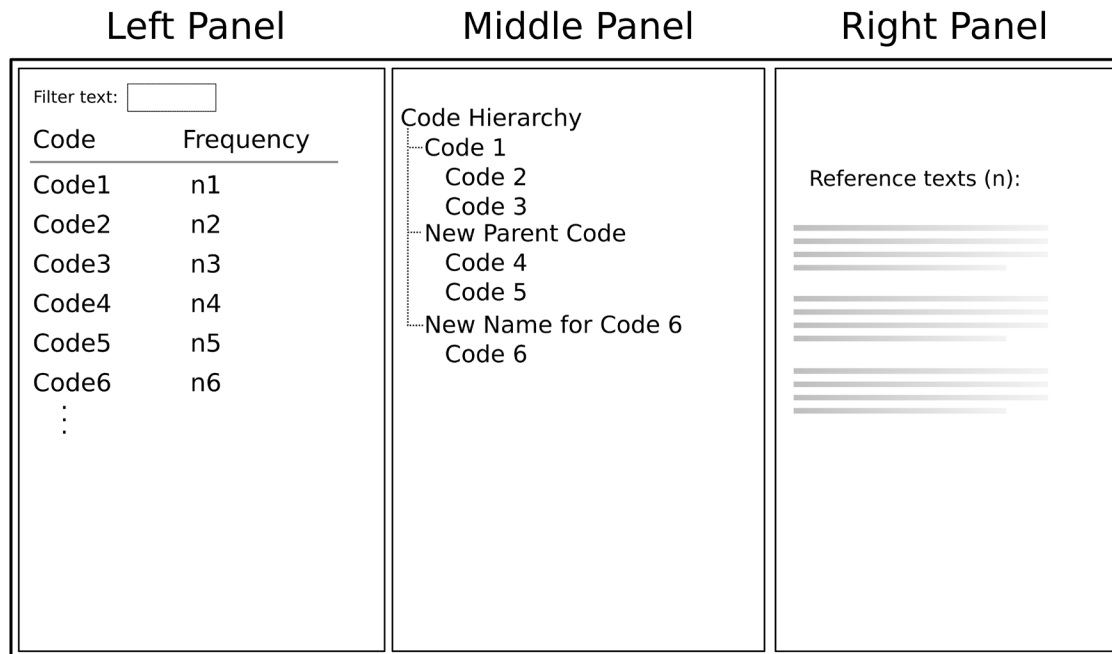


Figure 1. An illustration of the three side-by-side panels of the *WordCommentsAnalyzer* user interface. The left panel shows the codes in the comments with their counts, the middle panel provides a code tree for intuitive organization of the codes and the right panel shows the data extracts pertaining to each code (or to children of a parent code). The code list in the left panel can be filtered to find specific codes. The user can place codes in the code hierarchy simply by using drag-and-drop. The tree also enables the user to move codes in the hierarchy if needed. The user can introduce a new parent code. The codes are changed or combined by being wrapped in new codes.



Figure 2. Two text samples of #successfulaging Tweets, which are commented using line-break-separated codes. The codes describe notable topics concerning the text samples.

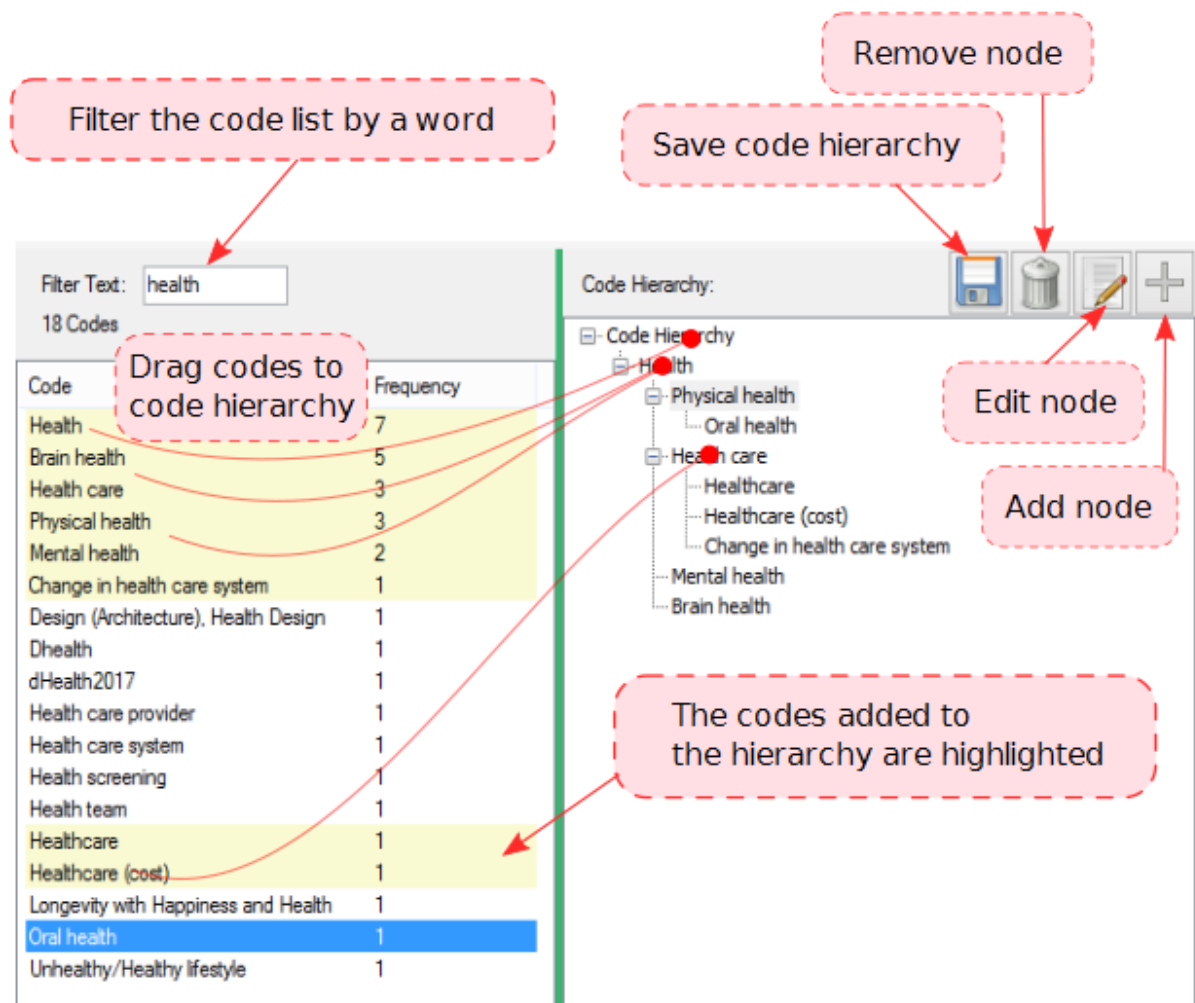


Figure 3. Features in the left and middle panels of the WordCommentsAnalyzer user interface. The user can find specific codes by filtering the code list (e.g., by the word of "health") and organize the codes (from the left panel) by dragging and dropping them into the code hierarchy tree (the right panel).

1	Health	51	Goals
2	Physical health	52	Planning ahead
3	Eating	53	Goals and dreams
4	Nutrition	54	Perception of aging
5	Fitness	55	Aging population
6	Health care	56	Role models
7	Mental health	57	Figures
8	Mental exercise	58	Models
9	Brain health	59	Active
10	Enhanced cognition	60	Physical activity
11	Mental exercise	61	Exercise
12	Homecare	62	Athletic
13	Longevity	63	Running
14	Longevity with Happiness and Health	64	Marathon
15	How to	65	Social engagement
16	Marketing	66	Social relationships
17	marketing	67	Friends
18	Marketing strategies	68	Social relationships (absence)
19	Technology	69	Social isolation
20	Critical	70	Loneliness
21	Contrarian	71	Social network
22	Biomedical model	72	Staying active
23	Social model of successful aging	73	Volunteering
24	Definition	74	Working longer
25	Retirement	75	Acceptance
26	Enjoy Retirement	76	Change
27	Filling retirement	77	Belief
28	Life after retirement	78	Habits
29	Retirement communities	79	Place for aging
30	Retirement community	80	Best place for aging
31	Retirement village	81	Aging in place
32	Award	82	Cities
33	Event	83	High functioning portrays
34	Expo	84	Education (HE) when aged
35	Challenge	85	Enhanced cognition
36	New experiences	86	Family
37	Happiness	87	Couple
38	Meaning	88	Communities
39	Optimism	89	Retirement communities
40	Never too late	90	Retirement community
41	Athletic	91	Retirement village
42	Record	92	Website
43	Worth	93	Website, blog
44	Life after being aged	94	Website partnership
45	Wisdom	95	Experts opinions
46	Growth	96	Scientific research
47	Aging diseases	97	Medical expert
48	Dementia	98	Scientific article
49	Caregiving	99	Dance
50	Learning	100	Pet

Figure 4. A formatted version of tab-indented text output file of the code hierarchy tree. When we organized the Tweet codes with at least two counts. The large branches of the code tree can help the researcher identify the richest themes in the data. Thus, themes of health, retirement, happiness, and being active are probably the major themes in the Tweets with the hashtag #successfulaging.

of health, retirement, happiness and being active represent the richest themes in the Tweets of #successful aging.

Conclusion

This article presents a Windows software tool for organizing comments in Word documents. *WordCommentsAnalyzer* facilitates organizing codes in a code hierarchy for qualitative researchers who are interested in using Word documents to annotate their data.

Software availability

Source code available from: <https://github.com/ehsabd/word-comments-analyzer>.

Archived source code at time of publication: <https://doi.org/10.5281/zenodo.1228604>⁵.

License: GNU General Public License 3.0.

Competing interests

No competing interests were disclosed.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Supplementary material

Supplementary File 1. Tweets hashtagged with #successfulaging from 17 January 2017 to 10 April 2018.

[Click here to access the data.](#)

Supplementary File 2. The tab-indented text file of code hierarchy.

[Click here to access the data.](#)

References

- Lewis RB: NVivo 2.0 and ATLAS.ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs. *Field Methods*. 2004; 16(4): 439–64. [Publisher Full Text](#)
- Lu C-J, Shulman SW: Rigor and flexibility in computer-based qualitative research: Introducing the Coding Analysis Toolkit. *Int J Mult Res Approaches*. 2008; 2(1): 105–17. [Publisher Full Text](#)
- Chenail RJ, Duffy M: Utilizing Microsoft® Office to produce and present recursive frame analysis findings. *Qual Rep*. 2011; 16(1): 292. [Reference Source](#)
- La Pelle N: Simplifying qualitative data analysis using general purpose software tools. *Field Methods*. 2004; 16(1): 85–108. [Publisher Full Text](#)
- Abdekhodaie E: WordCommentsAnalyzer: A windows software tool for qualitative research (Version 2.0.2.1). *Zenodo*. 2018. [Data Source](#)

Open Peer Review

Current Peer Review Status:



Version 1

Reviewer Report 02 August 2018

<https://doi.org/10.5256/f1000research.16128.r36504>

© 2018 Mansourian Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Yazdan Mansourian 

Charles Sturt University, Bathurst, NSW, Australia

This article reports an overall description about developing a free and open source software designed for qualitative data analysis. This software assists researchers to organise and categorise the initial codes they create through data analysis process in various qualitative methods such as thematic analysis, ethnography and grounded theory. As this software is free and easily accessible, it will be useful for many researchers who do not have access to more sophisticated tools.

The authors highlighted the ease of use and free access of this tool as its main benefits and I do agree with them. Nonetheless, they did not compare the capabilities and performance of this software with one or two well-known tools in this field to explain its strengths and weaknesses. Obviously, the existing tools have several options for deeper level of analysis and also some facilities for analysis of non-text data such as image, audio and video. As a result, I think the authors should also remind the reader about the inevitable limitations of this product. I also recommend the following revisions in the article:

1. The article should be more informative in terms of distinctive features of this software comparing to similar tools in this context.
2. The first citation in the introduction section (Lewis, 2004) has been published 14 years ago while since then the area of qualitative data analysis (QDA) has been developed considerably. As a result, I recommend to include a few recent citations in this section to provide the reader with a more accurate image of the current trends and issues in this area.
3. The case presented in this article is based on a relatively small collection of data. Accordingly, we still do not know how effective the software will be in larger data sets. As sometimes, the huge volume of the data in large scale studies may reduce the performance of a software.
4. The authors mentioned popular qualitative data analysis tools such as QDA Miner Lite does not support Persian and Arabic languages. Nonetheless, the case presented in the article is English and we still do not know about the real performance of this new tool in other languages such as Persian and Arabic.

5. The conclusion is too brief at this stage and it should be more detailed and insightful. A good conclusion summarizes the main points of the article and invites the reader to think further about the focal point presented in the discussions.

In general, this article is based on a creative idea but still requires some revisions and I hope the software presented here will be useful for researchers who use qualitative approach in their studies.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Partly

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Partly

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Aug 2018

Ehsan Abdekhodaie, University of Tehran, Tehran, Iran

We thank Dr. Mansourian for his valuable comments about our manuscript/software. The comments surely helped us improve both the software and the manuscript.

Some aspects of the software have been updated at the time we present this revised manuscript. Indeed, apart from the tweaks done in the software performance so that it can handle larger datasets, comparison of the software features to the features of other well-known QDA software compelled us to add some useful features to our software including search in code hierarchy feature and visualization tools. We also specifically respond to the reviewers comments in the following:

1. We compared the features of this software to the similar tools in a new Table in the manuscript and added an accompanying paragraph which highlights the main points regarding this comparison and describes the specific limitations of our software. We also mentioned the limitations of WordCommentsAnalyzer regarding analysis of non-textual materials, doing complex

queries, or presenting sophisticated visualizations.

2. We added a few citations. Specifically, one to provide more recent data on the usage of popular QDA software (Woods, Paulus, Atkins and Macklin, 2016) and others to give the readers a glance on current challenges in learning complex QDA software (Silver and Rivers, 2014; Woods, Macklin, and Lewis, 2016).

3, 4. The software indeed had issues with large dataset as the reviewer suspected. Consequently, we have updated the software so that it can handle larger datasets more efficiently. Then, we added another use case that is based on a collection of Iranian journal abstracts to test the software performance on a large dataset. This way we could also address the reviewer's comment about performance of the tool in the presence of Persian/Arabic texts. It is worthy to note that, we developed the feature of search in the hierarchy and move in the hierarchy since we realized that it is not feasible for large datasets to find the codes by scrolling the hierarchy tree or to move the codes by merely dragging and dropping.

5. We tried to make the conclusion more informative and more linked to the rationale of the study. Also, we revised it so as to present a better summary of the points made in the article.

Competing Interests: No competing interests were disclosed.

Reviewer Report 11 June 2018

<https://doi.org/10.5256/f1000research.16128.r34052>

© 2018 Huang R. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Ronggui Huang

Department of Sociology, Fudan University, Shanghai, China

Various QDA software provide similar functionalities in terms of coding operations and organization of codes, for instance, RQDA (<http://rqda.r-forge.r-project.org/>), WeftQDA (<https://www.pressure.to/qda/>), Py3QDA (<https://github.com/Ronggui/PyQDA/>), among others. A more systematic comparison of existing tools and *WordCommentsAnalyzer* will provide a clear picture on the relative advantages and disadvantages of the latter.

It seems that *WordCommentsAnalyzer* mainly organizes the codes and shows the related coded text segments, but does not support coding operations on-the-fly. The operations of coding, remove coding, and re-coding have to be conducted on the Word Processor side. It would be helpful to potential users to describe this point clearly.

Since the operation of coding must be done with Word Processor, it seems that coders have to remember and type the names of codes directly or via the copy-and-paste method. It would be valuable if coders can do the coding via *WordCommentsAnalyzer*.

The most obvious advantage of *WordCommentsAnalyzer* is its easy use, especially for Windows users.

Overall, *WordCommentsAnalyzer* is a valuable new tool for organizing codes based on Word Processor.

Is the rationale for developing the new software tool clearly explained?

Partly

Is the description of the software tool technically sound?

Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?

Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?

Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Sociology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 29 Aug 2018

Ehsan Abdekhodaie, University of Tehran, Tehran, Iran

We would like to thank Dr. Huang for reviewing our software/manuscript and for his kind comments about the potential value this software provides. Here we respond to the reviewer's comments as follows:

We added a 'more systematic comparison' of existing tools and WordCommentsAnalyzer through a table which included three commercial QDA tools and RQDA along with WordCommentsAnalyzer. We also discussed the relative pros and cons of the program.

We asserted that our program is not meant to do coding operations on-the-fly and these operations should be done on the word-processing side.

The reviewer mentioned the concern that it may be difficult for the researcher to memorize and type the names of codes or copy-and-paste them into Word. We added a feature that enables the user to drag and drop codes (from either the code list or the hierarchy tree) into the Word comment. So the user can both remain on the word-processing side for editing the files and reuse

the developed code list/hierarchy effectively.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research