

Conducting a literature review: A puzzling task

Jenine Beekhuyzen
Griffith University
Brisbane Australia
jenine@griffith.edu.au

EXTENDED ABSTRACT

A literature review is much like a jigsaw puzzle; piecing together the seemingly endless pieces of published research and other sources, and telling a story with the finished 'picture'. There is a lack of available practical information on how to conduct a literature review, and there is even less available that use qualitative research software to support the process. To address this gap, this paper discusses the journey of an information systems PhD research student using Nvivo for a literature review. In this paper Nvivo8 is proposed as a tool to help any researcher accomplish a rigorous and transparent literature review. Here a practical example of such a process is presented in seven steps, using a well-known qualitative research software that gives the researchers new opportunities to explore and piece together the challenging task of a literature review.

It important to acknowledge the criticisms of qualitative research software however using such software can add benefits and value to the research process. Although there are endless texts on how to conduct research, such text rarely do justice to discussing *how* to do a literature review, especially to do so in any detail. While the purpose and objectives of conducting a literature review are generally clear, the process for actually carrying out the literature review is complex, continuous and can be difficult to explain. Therefore this paper attempts to contribute an account of how to conduct the literature review (although this is just one example), with hints on how to do a literature review using software. It attempts to make explicit as possible the process of conducting a literature review using qualitative research software.

This paper builds on the author's 2007 paper, which discusses using Nvivo7 for a literature review. With the release of Nvivo8 in 2008 and most current users moving to the new version, the paper needs updating. It is important to note that the information presented in the previous paper is by no means redundant and useless; this paper builds on the discussion presented there and is adapted for the Nvivo8 environment. Further, it discusses the use of some of the new features and how they can add value to conducting a literature review. These new features include more flexibility with the casebook (used for demographic information in an empirical project, used to categorise and track papers in a literature review); using video and audio as secondary data sources; and the addition of support for pdf files, which overcome a major challenge to a research project, identified in the previous paper.

I have chosen to use Nvivo8 and other software for quite specific reasons. Firstly, I have been using Nvivo8 (and its predecessors) for eight years across different projects, and it is widely used around the world. Using Nvivo8 to manage most documents relative to my PhD research, I have found it to be a useful tool for coding literature text and for supporting theory generation. I have found it to be a good project management tool for organising nearly all aspects of my PhD research project, and it is important to analogueise how to manage the research project without software (using highlighter pens and a photocopier). I argue that it is important to be able to do this. So far, the software has enabled me to manage my notes and ideas about the large collection of published research papers (electronic documents) and web links I have, as well managing details about the many external (hard copy) sources such as journal papers, books and reports. Combining these abilities as well as being able to code/categorise the summarised details of these sources into themes and being able to search across codes/themes are by far its greatest benefit and justifies its use.

But by no means is Nvivo8 the only option for a research project and it is not used in isolation. It can be used exclusively but its use should also be considered alongside other tools such as Excel and Endnote. In the project presented here, Excel is used to manage the observation data, as Nvivo8 wasn't as useful for this task. In terms of using Endnote, it is a bibliographic reference management tool (as opposed to a data analysis tool) and it is used in compliment to Nvivo8 and Excel. In this research Endnote is used to manage the bibliographic details of all of my literature sources. Even though Endnote enables the researcher to add abstracts and some references to what might be considered codes (i.e. keywords), Nvivo8 takes it a step further by allowing much greater flexibility in assigning any number of pieces of texts (coding) to any number of concepts (nodes). It also enables complex search functions that work on the coded text and nodes, tasks not available in Endnote or Excel. Basically the Nvivo8 activity of coding literature brings together all of your ideas (that you have purposively coded) about a particular concept or theme. By doing this, it is then easy to open just that node (concept/theme) to read filtered information about just that topic, and use this to write up your literature review.

My PhD research provides the basis for the discussion in this paper, as it provides the context in which to

present ideas about ways to use Nvivo8. I am now in the final year of my studies. My research approach is interpretive and my research data is qualitative. My ethnography takes a cultural perspective to understanding the attitudes, motivations and behaviours of online music users. The research papers that make up the literature review for such a project are also qualitative data and they need to be analysed in a way that gives the researcher the ability to deconstruct, analyse and synthesise them in a useful and relevant way. Coding literature is a form of qualitative data analysis and in this paper Nvivo8 is proposed as such a tool to help any researcher accomplish this.

INTRODUCTION

Qualitative research is a somewhat recent inquiry paradigm within the history of research traditions. Interpretive qualitative research is a much debated approach but it is also now an accepted approach for the investigation of information systems and organisations (Hirschheim and Klein, 1989; Walsham, 1992; Klein and Myers, 1999; Myers, 1999). This approach allows an understanding of technological phenomena within a social context, from the perspective of those involved.

Even newer to research traditions is the use of qualitative software to support data analysis. To give an example, in 1984 the first version of the Australian software 'Nud*ist' was released¹ with subsequent versions (of Nud*ist/Nvivo) to follow over the years. Widespread adoption of the software was held up for many years for various reasons; one significant one being the lack of acceptance by university professors of the validity of using software to aid qualitative research, particularly interpretive qualitative research. The argument being that a software tool must surely move the researcher further away from the data they are analysing; this being an opposing objective to the nature of the philosophical assumptions that underlie this approach.

It is important to remember that Nvivo8 is a software tool, and it should support the way you work. It is necessary to be able to analogise how the literature review would be done without the aid of software, and realise that the software doesn't do the work for us, but rather helps us in achieving our already defined goal.

Traditionally Nvivo has been used mostly for data collected in the field but now it is being used more regularly for reviewing literature; see the recent paper by Bandara (2006) in which she presents "an illustrative demonstration of Nvivo2 for research management" however di Gregorio started the discussion back in 2000 when she demonstrated the tools in Nvivo2's toolkit which she believes "support the various processes and strategies involved in constructing arguments from the literature" (2000). More recently, some other papers on how to use the software for research have been published, and are useful in providing advice to researchers (Dean and Sharp, 2006; Woods and Wickam, 2006).

SEVEN STEPS TO ANALYSING THE LITERATURE

This section outlines the seven steps involved in doing a literature review using qualitative research software. They cover the basics of coding research sources for your literature review. A number of these steps usually are iterative and they are often added to with more coding and 'coding on' phases.

As mentioned earlier, this paper builds on a previous paper (Beekhuyzen, 2007) to accommodate the new version of the software. The steps can still be conducted in the same way as outlined in that paper. A couple of steps have been affected by the upgrade to version 8, and they are discussed here in detail. They are positive steps that provide better tools for doing a literature review.

I have used my own PhD research project to provide context for using Nvivo8. The example given here is to provide exactly that, an example of how I used Nvivo for my literature review. It is hoped you will gain some insights into how the process might work for your own research project.

Step 1: install the software

The software must be installed before you begin, and the latest service pack should be installed (check for updates – help menu). It is important to check for updates every few months if you haven't set the software to do this for you. Many universities now have a site license making access to the software relatively simple.

Once installed, start the software. It is now available in Chinese, Japanese and Spanish in addition to English so choose your preference from the start menu. Then you will be prompted for your name and initials. From the next screen, choose 'new project' or enter the *volunteering* sample project to look around.

¹ by Lyn and Tom Richards of QSR – Qualitative Solutions & Research, Melbourne, Australia

Step 1a: Set up your project

With the addition of new types of sources in Nvivo8 (audio, video, images), the name *Internals* is now given to the top folder in the navigation view (see figure 1). If you prefer, you can create folders underneath the *Internals* folder to separate different sources however this isn't mandatory.

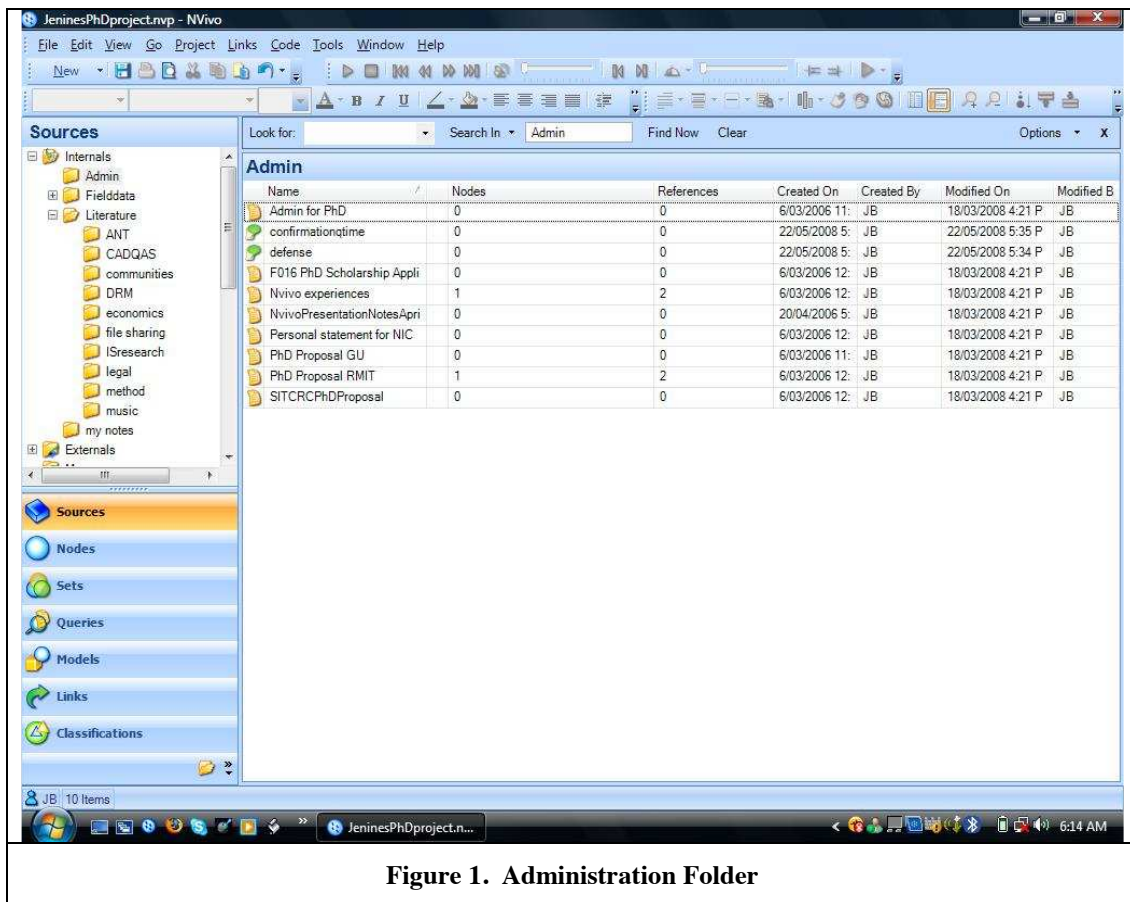


Figure 1. Administration Folder

I recommend beginning your project by creating a new document to record your notes about using Nvivo and working through the research process. Click in the white space of the list view while selecting *sources – internals*. Select *new internal – document*. Call this *project journal* and insert the date/time to begin (and do this each time you write in your journal). It is recommended that you write in this file each time you do any work (or any significant work) in Nvivo. Make comments about the work completed, the work to be completed next time and any other issues. This then becomes an audit log/trail for your work in Nvivo, and also part of your analysis. It also a useful list to show your supervisor or colleagues your progress.

Nvivo can be treated very much like a project management tool, as much as a qualitative data analysis tool. It is an effective way to keep everything together, which is particularly difficult over many years of a research project. I found it useful to import my research proposal, my scholarship applications, my consent forms, as well as the audio recordings from my confirmation (completed after 1 year of PhD study in Australia – it is a 20 000 word report, presentation and defense) see figure 1.

There is abundant help available when using the software. The flash tutorials give overviews of using the software and its various features, and pressing F1 at any time displays the detailed help. The *volunteering* sample project is also a useful way to get help. You can look at the project and see the sample queries, models and node structures to get ideas. The online forum and QSR helpdesk can also be invaluable resources.

Step 2: read and summarise literature

Each person has a different approach to reading and summarising the literature. Whether you print all papers and make notes in the margin, or make notes in a notebook, or read papers on a screen, it is still necessary to read the literature and summarise it in some way.

In addition to literature, other sources may be used in your literature review. Previously Nvivo only dealt with documents, now it accepts audio, video and image files. These may form part of your literature analysis, or they may be better used in the empirical part of your analysis.

Step 3: import literature

One of the most difficult tasks in using software to manage literature is the sheer amount of research papers and books. In a doctoral project (or any research project for that matter), the number of sources might reach 1000 or more. Managing the physical papers in these numbers is challenging, as is managing them in the digital environment. Therefore it is necessary to make this process manageable and to suit the schedule of each individual researcher. The other challenge is managing the combination of electronic and hard copy papers. Ideally it would be useful if all of our research papers were available and accessible in electronic form, however this is not always practical. Some journals and many conferences still do not publish their proceedings online, and books obviously have limitations.

The process for importing sources into Nvivo can vary between people. You may import one paper at a time as you read and summarise it (and maybe even code it at that point), or you may import a number of papers at a time, their summaries, and coding them now or later. There is no right or wrong way; just use the software to support what makes most sense to you.

One very good thing about Nvivo8 which makes this step significantly easier than previous versions of the software, is the introduction of handling pdf files. In my paper about Nvivo7, this was one of the main challenges in importing and coding research papers. Previously it would be necessary to convert pdf files to txt or only include summaries and link to the files as an external source, however this is no longer an issue. It is now a much easier task to import sources which are often in pdf format, and code them like any other document.

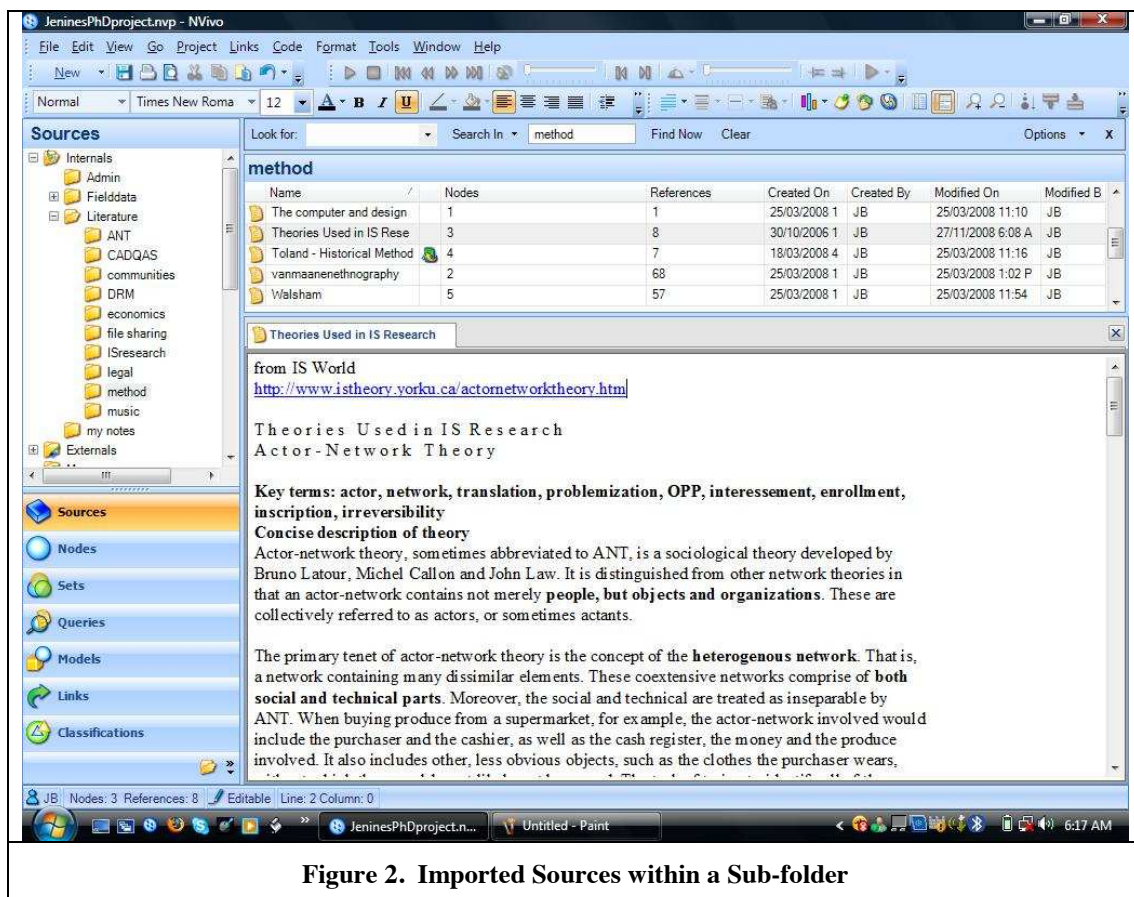


Figure 2. Imported Sources within a Sub-folder

Import each file and give it a description. You may decide to create subfolders for your literature – see figure 2. At this time you have the option to make each document into a case; you then can add attributes to each case (document) such as date imported (date), source type (journal/conf/book/web) and whether it has been coded (yes/no). Additional notes about a source can be added and connected by creating a link to a new memo from a source.

It may be useful at this point to add in any other links. Hyperlinks can be added to any text to link directly to the website (opening the web browser automatically), see figure 3. As you work through your rounds of coding (the next steps), you will add annotations (which are similar to writing in the margins of a paper, or footnotes) and see also links to link from an object (source/node/text) to another object (source/node/text).

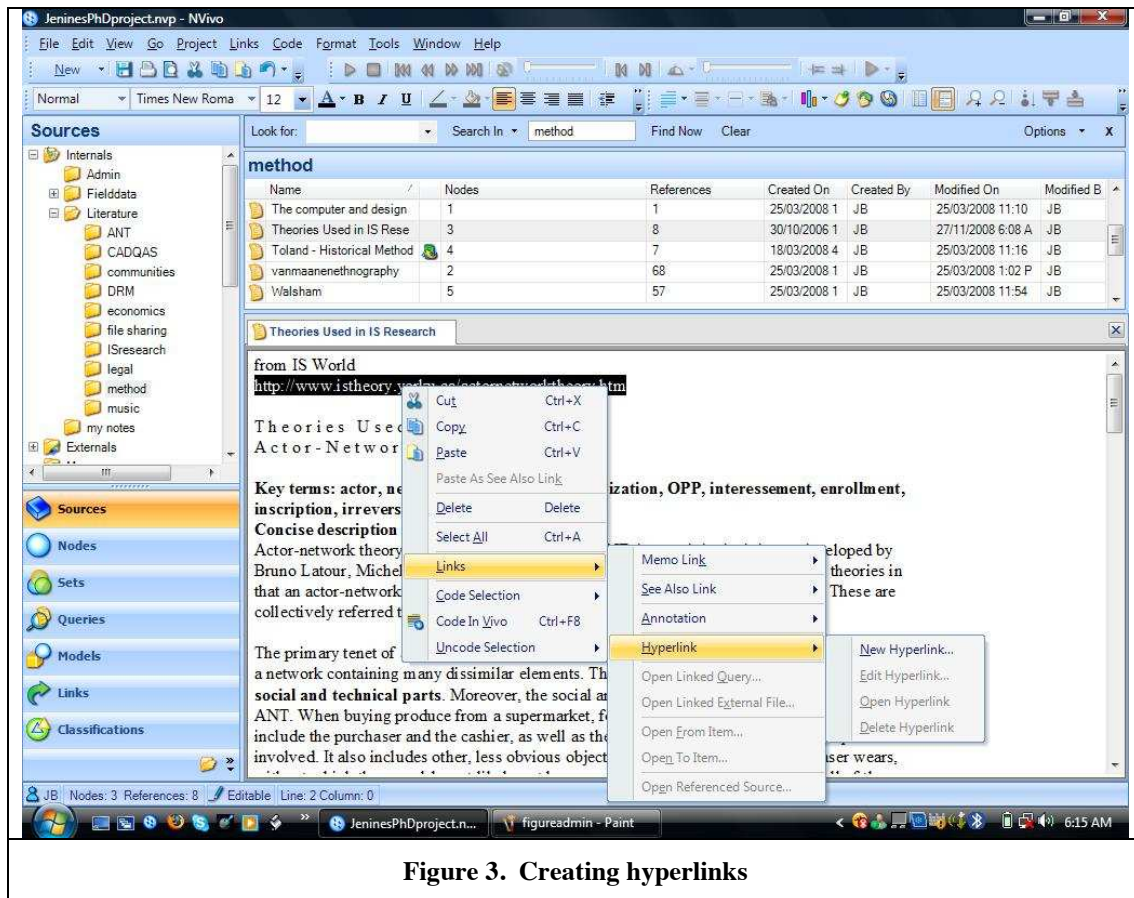


Figure 3. Creating hyperlinks

3b Using the Casebook

The casebook can be a useful feature when conducting a literature review. With empirical data, it is usually used for demographic data, but any data can be used if you think broadly enough. Anything that can be represented as a multiple-choice list or yes/no can be included in the casebook.

But what is the value of this? It enables us to run queries and filter the responses by fields in the casebook. For instance, we can record the attributes of empirical papers including the method used (interviews/observations/survey), the participant focus (students/management etc), the year of study, usefulness of the paper (yes/no), qualitative/quantitative, interpretivist/positivist philosophical approach etc. It is possible to then query the coding of the papers according to different criteria. Doing this helped me to classify the existing empirical research in my and related fields, and on related topics, with different foci according to stakeholder group (students, musicians, record industry). This was useful to help me to position my own research and the contribution it makes (Beekhuizen, 2008).

Step 4: coding 1st round

Coding, like summarising and importing, can take numerous forms. Once the sources are imported into Nvivo8, coding, or organising (or even deconstructing), the literature can begin. Firstly, open the document you want to start with. Then from the navigation menu (bottom left of screen), choose nodes, free nodes and the window for these is displayed (although it will be empty at this starting point). Having the document detail window and the node window open at the same time allows us to create nodes and code while reading through the source; the node window is displayed at the top of the screen and the document is displayed below. One difficulty is that once we start creating more than 30 or so nodes, seeing them all and coding to them with the detail view showing on the bottom can get difficult. One way to overcome this is to move the detail view to the right of the screen ('view – detail – right'). This allows you to see more than 50 nodes at one time and is a logical way for

many people to work at coding – selecting text and placing it in concepts (i.e. nodes - dragging from right to left, rather than bottom to top).

Once you start reading your first document, you can start coding. Nodes should be considered as ‘containers for ideas’ and are created to store ideas or text about related concepts together. When you see concepts in the literature that you want to code, highlight the text, right click and create a new free node from the selection. Another option is to click in the white space of the free node window (on top), right click and create new free node. Once created, highlight the relevant text, and simply drag it over the free node you want to code it to. This will code the text to that node. Continue this process of creating new nodes and coding at them until you reach the end of your document.

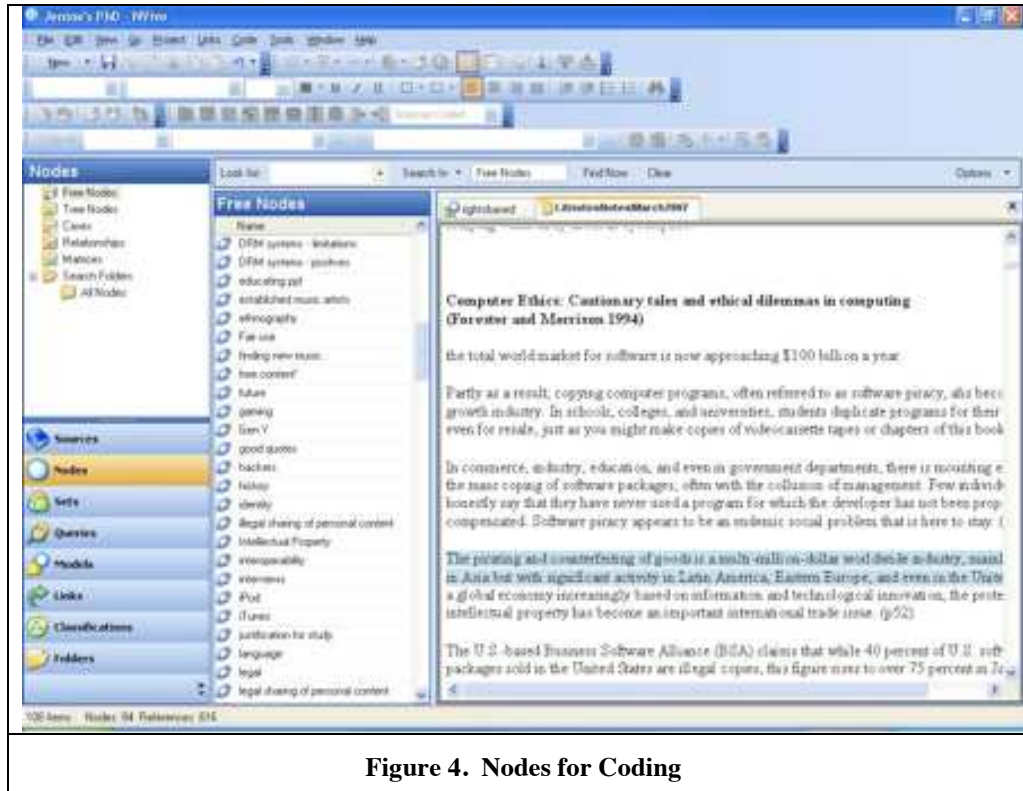


Figure 4. Nodes for Coding

To give an example, at the start I created a nodes for concepts such as music use, technology, online stores and others (see figure 4). I coded text from my sources that relate to these themes as I have identified them. If a node category didn't exist yet, I simply created it as needed.

A further option for coding is to highlight some text that you want to be the name of a new node, choose *code invivo* and a new node will be created with that highlighted text coded there and the name of the node is the text you highlighted. This is probably not so useful having only a few words coded here so go to the 'coding – spread selection' option to get the context of the coded text. This is automatically saved.

When going through this process, there are some important things to remember. First is that any piece of text can be coded infinite times. However just because this is technically possible, it doesn't mean that it is recommended practice. Generally you will find that a piece of text is rarely coded at more than 10 nodes at this point in the coding process (of course there are exceptions). So try to keep your codes rather broad and high level at this stage. You will break them down into detail soon enough. But if you find you are coding it at more nodes than this in this instance of 1st round coding, then you might be guilty of coding at too low a level. This is an important consideration because if you try to code at too low a level in this first instance at this early stage, you will find you will feel as though you will never reach the end of the coding, and that it is redundant and somewhat useless.

The other point to note is that not every part of the document needs to be coded. There is no problem in leaving parts of the document uncoded; these can be coded in a later phase, or left uncoded indefinitely. Not all parts of every paper or summary notes are relevant however you will find that a larger percentage of the summary notes will be coded compared to the amount coded if the individual full papers that are imported. This is because the summaries have already been filtered so that they include what you have decided is the most relevant information, whereas the full papers are just that, complete papers will all details. Also useful when creating

nodes is to add a description to explain what the node means. You may have a clear idea today what you mean but this is almost certain to change in six months time.

The various steps outlined so far are important to follow, and jumping ahead can cause anxiety and frustration early on. On this note, my personal approach is that it is important to disregard any thoughts of structure during this step, however I acknowledge that some other researchers and trainers approach this differently by starting with structure. I believe that creating structure and organising nodes at this time can cause disruption to your coding and thought processes. Coding a document is a very different process to structuring and should be kept as separate phases in the research process.

This process of coding 1st round (or broad coding) should be repeated until all imported documents are coded once. Once done, then structuring the nodes can begin before you start on more detailed coding.

Step 5: structuring nodes

One you have some free nodes and they have been coded to, and you have reached something of a stop point in the literature review (maybe finished the 1st round of coding on all documents), it is then time to think about structure. You might find that you don't want to include this phase in your research process, or you might want to include it at a later time (or even earlier), of which all are plausible and valid choices.

Structure at this point refers to tree nodes. Free nodes can be moved around to form tree nodes i.e. tree nodes consist of siblings (nodes on the same level) and children (sub-nodes). Thus by creating tree nodes you infer some form of hierarchy or grouping. The process of structuring your free nodes into tree nodes is done by copying a free node, and pasting it in the appropriate place in the tree node structure. Copying is used in case a mistake is made in moving it; this is a safety measure to ensure we don't lose any of our coding (and possibly even a fallback because of the omission of an undo button in previous versions of the software).

In choosing which free nodes would be appropriate as tree nodes, one way to decide at this point is to find the nodes you want to break down further (i.e. into sub-categories). By putting them as top-level tree nodes, it is then easy to open the top-level tree node (which was previously a free node) and 'code on' from here. 'Coding on' refers to coding from the already coded node (which was done in the 1st round of coding). This is a useful approach as it allows you to code from one concept (i.e. culture) and break this down into sub categories (i.e. norms, values, beliefs, rituals), without being distracted by irrelevant data/literature (it has been filtered out). This next phase of coding is discussed in step 6 below – 2nd round coding.

In my project I moved the 'DRM definition' free node to a tree node and broke it down into 'tech-based' with 18 references and 'rights-based' with 7 references from the literature indicating the different ways of defining digital rights management (see figure 5). I also found I had other definitions coded at this high-level node so I created a new node for 'general definitions'.

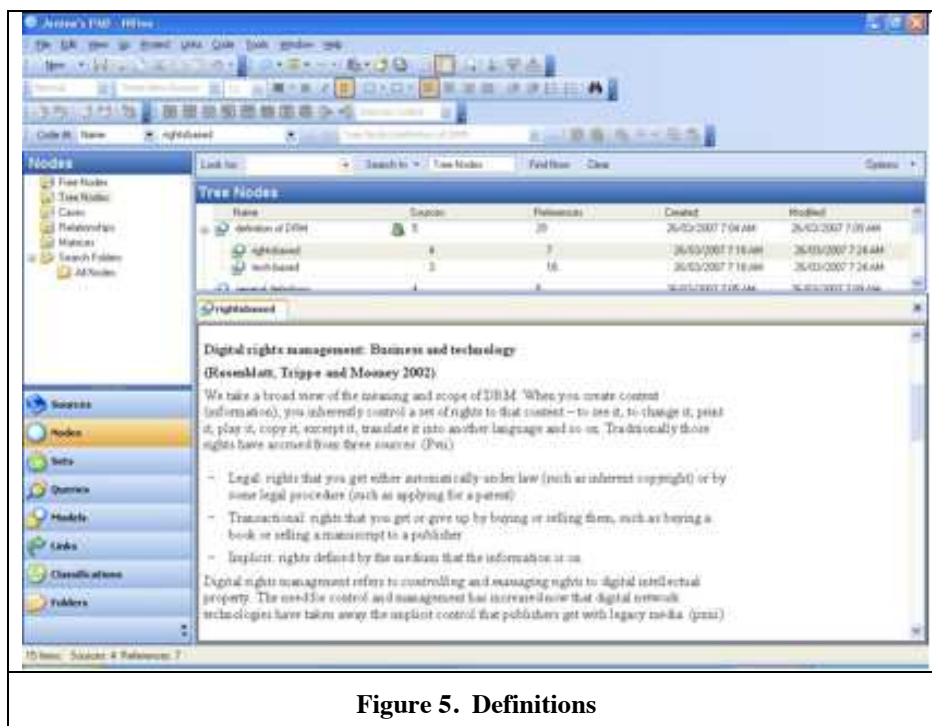


Figure 5. Definitions

The other part to the node structuring that can be carried out at this time is to include any concepts from models, theoretical or conceptual frameworks from the literature that you are using to guide your research. These can be set up as tree nodes, as they usually are hierarchical or at least presented in a connected way. Tree nodes allow representation of this structure and connections, and now is a good time to set them up for coding. Even if you decide not to code to them in step 6, you can do it at any other time.

Structuring may also refer to free nodes. After the first round of coding I found that some free nodes were doubled up or there were more than one with a similar description. These were merged together. In my project, I created 94 free nodes in the 1st round of coding. The structuring I did within the free nodes included:

- o 'bittorrent' was expanded to include 'P2P' (peer-to-peer) - merged
- o 'watermarking' and 'fingerprinting' were merged into TPM (technological protection measures)
- o 'record labels' were merged into 'corporate' – a node describing all corporate interests in online music
- o 'INDICARE' (a European online music project) was merged into 'DRM studies/projects'
- o 'DMCA' (the US digital copyright law) was merged into 'legal'

The other nodes to set up here are the relationship nodes. At this point, you may have some ideas about relationships you have found or intend to look for in the literature (hunches). These can be coded at the relationship node rather than at individual nodes; the relationship is a connection of two nodes and you code evidence of that relationship existing from the data.

It is also suggested that you get very familiar with your nodes and their structure before coding further. You might find it useful to print a copy of your node listing and keep it beside you or on the wall in front of you when coding. This helps to overcome the earlier discussed problem of viewing all of your nodes at one time on the screen.

Step 6: coding 2nd round

This round of coding is focused on breaking down the already created nodes into sub-nodes, or 'coding on'. As mentioned previously, 'coding on' refers to coding from an already coded node. For example, a free node called 'culture' is coded in step 4. Once relevant literature is coded there, and all imported documents are coded, the free node 'culture' is moved to be a top-level tree node. Added to this top-level node are sub-nodes (or children), which are categories or concepts generated as being connected and or/deriving from the top-level node.

To continue the example, concepts of culture found in the literature may include norms, values, beliefs and rituals; these are created as sub-nodes or child nodes, and then from within the 'culture node', continue to code into the lower-level nodes just created. So in essence, you are coding data that has already been coded in Step 4. This approach has the bonus of having a focus on the filtered literature for further coding. These sub-nodes (or children) can then be further broken down into more child nodes if required until you are happy that your coding is complete (or at least somewhat complete for now).

Step 7: analysis

Once you have coded to a level you are happy with (and it need only be happy for now), you need to make some sense of what you have coded. Putting a literature review together is much like a jigsaw puzzle; arranging all of the pieces so that they firstly make logical sense, and then telling a story with the final picture. It is not an easy task to write a good literature review, regardless of all of the advice on how to do so (Webster and Watson, 2002).

Unfortunately it is out of the scope of this paper to discuss the literature analysis in detail however in the following section I have highlighted some useful tactics for making the most of your coded literature. Another reason for discussing this only briefly is that analysis is shaped by your chosen research methodology and underlying research assumptions and is greatly determined by individual project needs and characteristics, so thus it is difficult to discuss in general terms. But using the process discussed thus far brings literature on your defined concepts and topics together, allowing for a focus on related literature that forms your literature review.

Coding tips

There are a number of features of Nvivo8 that are particularly useful during your literature analysis. These include searching for common terms, and the use of coding stripes, links and relationships. There are also some generic nodes that are useful for a range of projects, and these are often not considered as an option.

Searching 'common' terms

It is a useful task to search for common terms that have appeared in the literature that you want to be sure to code. You may already have these terms coded, depending on the details of your previous coding, but you may not have them all covered. To be sure, in my project I searched for terms such as Australia, security, privacy, Apple/Sony/Microsoft, iTunes, iPod, definition and many more.

Coding Stripes

Coding stripes are useful indicators. During coding they can be used to get a quick view as to whether and where a piece of text has been coded – answering the question 'did I code that?' Sometimes it is difficult to tell if a piece of text was dragged to the intended place, so moving the mouse over the displayed stripes gives a clear and quick answer. Coding stripes may also have benefit as they show other literature coded at a particular node. This may indicate links in concepts, establishing a thread between the literature, which could then be made explicit in the writing of your literature review.

Useful Nodes

In my review of the literature I have found that creating some other more generic nodes is useful. For instance, I created the following nodes and coded relevant text to them:

- Statistics
- Good quotes
- Justification for my study
- History
- Definitions
- Empirical studies
- Method
- To follow up
- Questionable statements
- Stakeholders: artists, corporate, users
- Ideas for interviews

I also found that creating a number of memos (like a small document – think of a sticky note) can also be useful. The memos I created for recording thoughts were:

- Nvivo experiences
- Methodological problems

If you have something that you want to code and don't know where to code it, simply create a free node or a memo. Try not to spend time worrying about where it should go, even if you create a node called 'to do later' and code it there then at least it is stored for later access. These are just a number of extra things that you can do in your project. You will find the more you use the software, the more confident and thus explorative you will become.

CONCLUSION

This paper discusses using Nvivo8 for a literature review, using a current PhD research project as an example. The practical steps outlined can be applied to a variety of research projects. As discussed early in this paper, qualitative research software is now becoming more mainstream in its use, particularly by research higher degree students.

The process of literature analysis and synthesis is rarely discussed in detail, apart from discussions in university research courses and some research textbooks. Because of this, as researchers we need to make more effort to share this information. Part of the challenge of going through the process of a literature review is the lack of detailed and practical help compared to other parts of the research process. By explaining our processes, we can help to inform those just starting out by giving them examples of how it has been done successfully (or not so successfully is often still useful).

In the most part, Nvivo8 is used to analyse empirical data however this paper has attempted to show its applicability for conducting a large-scale literature review. The various necessary steps for setting up and coding the literature are the intermediate outcome of one particular PhD project and give some guidance on how another may go about doing similar tasks. This paper presents a practical account of conducting a literature review in a qualitative research project.

REFERENCES

- Bandara, W. (2006) Using Nvivo as a research management tool: A case narrative. Proceedings of QualIT2006: Quality and Impact of Qualitative Research, Brisbane, Australia, Griffith University, 6-19, 27-29 November.
- Beekhuizen, J. (2007) Putting the pieces of the puzzle together: Using Nvivo for a literature review. Proceedings of QualIT2007: Qualitative Research, From the Margins to the Mainstream, Wellington, New Zealand, Victoria University of Wellington, 18-20 November, <http://www.qualityconference.com/>.
- Beekhuizen, J. (2008) How does Technology Influence Online Music Access and Use? A Taxonomy of Empirical Studies 19th Australasian Conference on Information Systems Christchurch, New Zealand, University of Canterbury, 3-5 Dec 2008.
- Dean, A. and J. Sharp (2006). "Getting the most from NUD*IST/NVivo." The Electronic Journal of Business Research Methods 4(1): 11-22.
- di Gregorio, S. (2000) Using Nvivo for your literature review. Strategies in Qualitative Research: Issues and Results from Analysis using QSR Nvivo and Nud*ist, London, Institute of Education, 29-30 September.
- Hirschheim, R. and H. K. Klein (1989). "Four paradigms of information systems development." Communications of the ACM 32(10): 1199-1216.
- Klein, H. K. and M. D. Myers (1999). "A set of principles for conducting and evaluating interpretive field studies in information systems." MIS Quarterly 23(1): 67-94.
- Myers, M. D. (1999). Qualitative research in information systems. IS World. 19.06.06: <http://trip.itss.auckland.ac.nz/msis/isworld/>.
- Walsham, G. (1992). Interpreting information systems in organizations. New York, John Wiley & Sons.
- Webster, J. and R. T. Watson (2002). "Analyzing the past to prepare for the future: Writing a literature review." MIS Quarterly 26(2): 13.
- Woods, M. and M. Wickam (2006) Methodological implications of software use: an empirical investigation of software programs on literature analysis using n6 and nvivo. Proceedings of QualIT2006: Quality and Impact of Qualitative Research, Brisbane, Australia, Griffith University, 163-172, 27-29 November.