

A comparative study of qualitative data mining tools based on fieldwork experience

Huhe Qiao^{1}, Yan Chen²*

¹Northwest Minzu University, Lanzhou, China

²Tianshui Normal University, Tianshui, China

*Corresponding Author. Email: johuhint@163.com

Abstract. Qualitative data mining tools have been widely applied in academic research that involves qualitative data analysis. Based on the authors' experience in analyzing interview materials collected through fieldwork, this study compares several commonly used qualitative data analysis software tools, examining their respective strengths and weaknesses. It evaluates these tools from three dimensions: cost, research needs, and technical competence. The conclusion drawn is that researchers should select qualitative data mining tools that best suit their specific needs and technical skills. Beginners are encouraged to try mature commercial software. Looking ahead, qualitative data mining tools are expected to evolve toward greater automation, connectivity, and mobility.

Keywords: qualitative data, textual data, data mining, NVivo, MAXQDA

1. Introduction

Qualitative Data Analysis (QDA) software has become an essential research tool across numerous academic disciplines. Researchers often rely on common qualitative methods such as oral histories, ethnographic studies, and interviews to collect firsthand data. Any discipline involving textual analysis inevitably requires the use of such tools for in-depth research. A simple search on CNKI or the Web of Science (WoS) database yields a vast number of related academic publications. For example, a search using "NVivo" as the keyword in the CNKI database retrieves over a thousand papers, covering a wide array of fields such as education, healthcare, safety, and sports. As indispensable tools in the research process, the abundance of QDA software has posed a "pleasant dilemma" for many researchers—particularly newcomers—regarding which tool to choose. Previous domestic studies have compared several qualitative analysis tools [1]. However, the distinct feature of this study lies in the authors' hands-on experience, comparing both commercial and open-source software to provide practical insights for the academic community.

This study selects several mature and widely used software tools, analyzes their key functions, advantages, disadvantages, and appropriate use cases based on user experience and tutorials, and ultimately proposes selection strategies tailored to different types of researchers. The study also offers a brief outlook on the future development of qualitative data analysis software.

2. Overview of major qualitative data analysis software

Qualitative data analysis software is widely used in both academia and industry. However, differences in usage environments and user conditions have led to a diversification in how such tools are applied. Commercial software, with its comprehensive features, dedicated support, and timely updates, is highly favored by users. Nevertheless, due to limitations in research methodology and funding, open-source software also maintains a notable presence in the QDA software market.

2.1. Commercial software

Currently, most widely used qualitative data analysis software falls under the category of commercial products. Common features of commercial software include a mature market ecosystem, professional technical support teams, regular updates, and well-developed functionalities—though they often come at a higher cost. Compared to open-source options, commercial tools are more suitable for research teams with sufficient funding and budget flexibility.

2.1.1. NVivo

NVivo is one of the most mainstream professional QDA software programs, developed by QSR International. It is widely used in fields such as social sciences, business research, and public health. Numerous studies in the CNKI database reference the use of NVivo, including English-language research such as Pandey et al.'s study on the impact of artificial intelligence on recruitment and human resources issues [2]. Domestic scholars have also discussed its application in the field of education [3]. The core strengths of NVivo lie in its robust data management, in-depth coding capabilities, and powerful visual analysis functions, making it particularly suitable for handling large-scale, complex qualitative data.

From a functionality perspective, NVivo offers strong support for multimodal data. It supports a wide range of file formats, including text data (Word, PDF, emails, web content), multimedia files (audio, video, images), structured data from Excel, SPSS, and SurveyMonkey, as well as social media data from platforms such as Twitter, Facebook, and YouTube. In terms of intelligent coding and analysis, NVivo features AI-assisted tools for auto-coding themes and sentiment analysis, matrix coding for cross-group comparisons, framework matrices for structured presentation of high-frequency themes, and multilingual text search functions. NVivo also excels in visualization and modeling, offering features such as: Concept maps for dynamically visualizing code relationships, Hierarchical charts for showing the structure of coding schemes, Geospatial analysis for mapping data distributions, Timelines for tracking event sequences over time. When it comes to collaboration and project management, NVivo includes: NVivo Collaboration Cloud for team-based online collaboration, Version control for tracking project edits, Cross-platform compatibility, supporting Windows, macOS, and web-based access. NVivo also supports mixed-methods research. It allows integration of qualitative and quantitative data—for example, linking interview transcripts with statistical findings—and enables data export to SPSS or Excel for further analysis.

Based on this functionality analysis, NVivo is particularly suited to the following four types of research scenarios: Large-scale qualitative research, such as transcription analysis of thousands of interview records. Interdisciplinary team projects requiring collaborative coding. Theory-building studies using grounded theory or discourse analysis methods supported by large data volumes. Business or policy research that prioritizes efficiency and utilizes AI-assisted reporting.

NVivo is available in three versions: NVivo Pro, the full-feature version available via subscription or perpetual license, NVivo Teams, an enterprise-level version for collaborative projects, NVivo Web, a basic browser-based version. Educational discounts are available, reducing student pricing by approximately 60%.

Compared with other commercial QDA tools, NVivo has three key advantages: It offers the most comprehensive feature set and is widely regarded as the industry benchmark. It is the only mainstream QDA tool that supports AI-assisted auto-coding. It provides enterprise-level support for data security and team collaboration. However, NVivo also has some disadvantages: Steep learning curve, often requiring official training to optimize usage; High cost, with team versions priced at over ¥10,000 per year; Cross-platform limitations, as the macOS version lags behind the Windows version in functionality.

In summary, NVivo is the top choice for large-scale, well-funded research projects, particularly those that involve AI integration or international team collaboration. For small- and medium-scale studies, alternatives like MAXQDA or ATLAS.ti may be more suitable. For entirely cost-free solutions, the RQDA + R combination is recommended.

2.1.2. MAXQDA

MAXQDA is a professional Qualitative Data Analysis (QDA) software program that also supports mixed-method research combining qualitative and quantitative approaches. It is widely used in fields such as sociology, psychology, educational research, and market analysis. The software offers robust capabilities in coding, data retrieval, visualization, and team collaboration. Several studies have employed MAXQDA in practice: for instance, Liu et al. [4] used it to analyze the influencing factors of point-based supermarket operations in rural community governance; Abraham et al. [5] applied MAXQDA to explore physicians' perceptions of patients.

Like NVivo, MAXQDA supports a wide variety of data formats. These include: text data (PDF, Word, Excel, plain text), multimedia files (audio, video, images), social media and web content (e.g., Twitter, web pages), Survey data (SPSS, Excel, CSV, etc.). In terms of coding and analysis, MAXQDA offers flexible coding modes, including both open and structured coding. It also features a code system management function that supports hierarchical codes, color tagging, and automatic coding based on keywords or regular expressions. The software allows users to take memos and annotations to document analytical thoughts and track coding decisions. For advanced querying and visualization, MAXQDA provides cross-tab analysis to compare coding patterns across different groups. Users can quickly identify high-frequency terms through word clouds and lexical searches, quantify and visualize code distribution with matrix analysis, and present data geospatially using mapping tools. Additionally, the software supports the graphical modeling of theoretical frameworks, which facilitates conceptual model development. MAXQDA is also well-suited for mixed-methods research. It enables integration of qualitative coding with quantitative data from Excel or SPSS and supports statistical functions such as descriptive statistics and crosstab analysis. These features significantly enhance its applicability in mixed-methods studies. In terms of team collaboration, MAXQDA offers robust support. Its team project sharing function allows multiple users to synchronize coding work, while version control and file merging features prevent data conflicts. It also provides cloud sync capabilities via MAXQDA Cloud or local servers. Regarding cross-platform compatibility, MAXQDA is available as a desktop application for both Windows and MacOS. Additionally, it offers a mobile viewer app, MAXQDA Reader, and a web version for remote collaboration—though the latter includes only basic features.

The following research scenarios are particularly well-suited for MAXQDA: Complex qualitative research, such as grounded theory or content analysis. Mixed-methods research integrating quantitative data. Collaborative team projects involving multi-user coding and data integration. Statistical and visual analyses involving matrix analysis, word frequency statistics, etc.

The advantages of MAXQDA can be summarized in four points: Comprehensive features with support for mixed-methods research. User-friendly interface, easier to learn than RQDA and more flexible than NVivo. Powerful visualization tools, including model diagrams, word clouds, and maps. Strong team collaboration functions, making it suitable for academic and commercial research teams. However, MAXQDA also has three main drawbacks: As commercial software, it is relatively expensive (though more affordable than NVivo). Some advanced functions (especially for mixed-method analysis) have a steep learning curve. Similar to NVivo, the macOS version lacks some features available in the Windows version.

MAXQDA is available in three editions: MAXQDA Standard, suitable for individual researchers, MAXQDA Plus, which includes additional analytical tools, MAXQDA Analytics Pro, which integrates advanced statistical analysis via R and Python. Educational discounts are available for students and educators, allowing them to purchase at a reduced rate.

In summary, MAXQDA is a well-balanced QDA tool, particularly well-suited for mixed-methods research or qualitative analysis requiring quantitative support. Compared to NVivo and ATLAS.ti, MAXQDA offers a more intuitive interface and is more cost-effective. For researchers on a tighter budget, RQDA may be considered, though MAXQDA offers superior usability and greater functional depth.

2.1.3. *ATLAS.ti*

ATLAS.ti is a powerful commercial Qualitative Data Analysis (QDA) software widely used in social sciences, market research, psychology, and related fields. It supports the coding and analysis of various data types including text, audio, video, and images, and offers a rich suite of visualization tools that help researchers gain deeper insights into complex datasets. For example, Jiao et al. [6] used ATLAS.ti 9 to analyze 32 policy documents related to strengthening the sense of community of the Chinese nation. Yunseon et al. [7] utilized ATLAS.ti to investigate social media issues during the COVID-19 pandemic. However, such applied cases remain relatively limited.

Similar to the other two commercial QDA tools, ATLAS.ti supports multiple data formats including: Textual data (PDF, Word, Excel, plain text), Multimedia data (audio, video, images), Social media data (e.g., Twitter, web content), Survey data (e.g., CSV and SPSS files). ATLAS.ti features a flexible coding system that accommodates both open coding and pre-defined codebooks. It allows for the grouping and hierarchical management of codes, supporting parent and child code structures. The software includes memo functions for recording analytical ideas and annotations. It also supports word frequency analysis and word cloud generation. ATLAS.ti's advanced query and visualization capabilities include Boolean searches (AND, OR, NOT), co-occurrence analysis to examine relationships between codes, network views to visually represent the connection between codes and data, and geocoding tools for mapping data geographically. In terms of team collaboration, ATLAS.ti offers robust support for large research teams through cloud-based project sharing (ATLAS.ti Web) and multi-user collaborative coding. Regarding cross-platform availability, ATLAS.ti provides desktop applications for both Windows and macOS. It also has native mobile applications, with partial support for iOS and Android, and a web version with only basic analytical features suitable for lightweight online collaboration.

ATLAS.ti is well suited for the following research contexts: Complex qualitative studies using grounded theory or discourse analysis, Multimedia data analysis involving large volumes of interview recordings or video content, Collaborative projects requiring coordinated multi-coder efforts, Research involving advanced visualization, such as network graphs or word frequency analysis.

The key strengths of ATLAS.ti include: Comprehensive functionality, supporting diverse data types and advanced analytics; User-friendly graphical interface, more intuitive than RQDA; Powerful visualization tools, including network diagrams, word clouds, and geocoding; Strong support for team collaboration, ideal for academic and commercial research teams. However, it also has several drawbacks: As a commercial product, ATLAS.ti is relatively expensive, making it more suitable for institutions or long-term researchers; Due to its extensive functionality, the learning curve is steep, often requiring formal training; Some advanced features are only available in the desktop version, while the web version remains limited in functionality.

ATLAS.ti is available in two main versions: ATLAS.ti Desktop, the full-featured version available via subscription or one-time purchase; ATLAS.ti Web, which offers only basic features and is intended for lightweight use. The software also provides educational discounts for students and educators, allowing them to access the platform at a reduced price.

In conclusion, ATLAS.ti is one of the most functionally comprehensive QDA tools available, ideal for complex research projects and team collaboration. For researchers with limited budgets, alternatives like RQDA or Taguette may be considered, although these offer fewer features and more limited data processing capabilities.

When viewed from a broader perspective across all commercial QDA software, NVivo remains the most widely used. According to data from CNKI, 1,276 publications reference the use of NVivo, while the Web of Science (WOS) lists 12,964 publications. In comparison, MAXQDA appears in only 24 CNKI articles and 3,043 WOS records. Interestingly, ATLAS.ti is cited in just 11 CNKI papers but appears in 5,497 WOS publications. This suggests that NVivo is the dominant choice in both domestic and international academic and industry settings. Between the other two, MAXQDA seems to be more accepted among Chinese researchers, whereas ATLAS.ti appears to be more recognized internationally.

2.2. Open-source software

Open-source software serves as an alternative to the aforementioned commercial software, offering the advantages of being free, open-source, and lightweight. However, compared to commercial software, it has notable shortcomings such as limited maintenance, simplified functionalities, and a requirement for some technical expertise. Due to the constraints in functionality and update frequency, open-source software has not become mainstream in the field of qualitative data analysis. It is more suitable for beginners and small-scale research projects.

2.2.1. *RQDA*

R Qualitative Data Analysis (RQDA) is a tool developed under the GPL license and is part of the R programming language package. As an open-source qualitative data analysis software based on R, it is primarily used for coding, categorizing, and visualizing text data. As an extension package of R, RQDA can directly call R's statistical and plotting capabilities and supports scripted operations, making it suitable for reproducible research. It also provides a user-friendly GUI interface that allows non-programmers to perform coding, annotation, and retrieval operations, significantly lowering the entry barrier.

RQDA supports open coding and axial coding and allows for the creation of multi-level coding systems. It also offers retrieval and query functions, enabling users to quickly locate text segments by keywords or code combinations. The built-in memo function allows users to add analytical notes to text or codes. RQDA includes some visualization tools, such as generating code frequency charts, word clouds, or network diagrams.

It is important to note that RQDA has limited data compatibility, supporting only formats such as TXT, PDF, and RTF for import, and exporting coding results in CSV or HTML reports. As an open-source tool, RQDA's interface and stability may not match those of commercial software like NVivo, and its updates are relatively infrequent. Since RQDA follows the GPL license, it is better suited for researchers with limited budgets but requires configuring the R environment beforehand and the additional installation of the GTK+ library.

Overall, RQDA is suitable for researchers familiar with R or those seeking flexible analysis, and it is especially recommended for social science users who value the integration of open-source tools with quantitative analysis. It is advised to use it alongside R packages such as *quantda* to enhance its functionality.

2.2.2. *Weft QDA*

Weft QDA is a free and open-source Qualitative Data Analysis (QDA) software designed for the coding, categorization, and retrieval of textual data. It features a simple design, making it suitable for small-scale research projects or beginners.

Developed in Ruby, Weft QDA includes core functionalities such as text coding, category management, retrieval capabilities, and basic visualization. It supports open and selective coding, allows for hierarchical categorization (Categories), and enables text segment searches by code or keyword. The software provides basic statistical summaries of code frequencies and generates simple reports. In terms of data formats, Weft QDA supports only plain text (.txt) import and does not support more complex formats such as Word or PDF. It is cross-platform and can run on Windows, Linux, and macOS (based on Java).

The strengths of Weft QDA lie in its small file size, easy download, and free, open-source nature, making it suitable for budget-conscious researchers. However, its functionalities are quite basic compared to other tools: it lacks advanced analysis features, cannot generate network diagrams or perform automated text mining, and its interface is relatively outdated, with the last stable version released around 2009.

In summary, Weft QDA is best suited for small-scale qualitative research, classroom demonstrations, or for beginners to become familiar with QDA workflows. For more powerful features, RQDA or Taguette is recommended.

2.2.3. *Taguette*

Taguette is a free and open-source qualitative data analysis (QDA) tool that focuses on text coding and annotation, making it ideal for academic research, social sciences, and user studies. Developed using modern technologies, Taguette supports both local and cloud-based use, and features a clean and intuitive user interface.

In terms of data compatibility, Taguette supports the import of TXT, PDF, DOCX, and ODT formats, although PDF processing relies on system-level OCR. It allows exporting of coded results in HTML, CSV, or JSON formats for further analysis. Because Taguette is web-based, it can run in browsers on Windows, macOS, and Linux, and also offers desktop and server versions. The server version supports collaborative team coding. Key features include text coding, project management, full-text search, and collaborative analysis. It supports highlighting, multi-level tagging, and managing multiple research projects. It also enables keyword or code snippet searches and allows for simultaneous multi-user coding (when deployed on a server). Since data is stored locally by default, it is especially suitable for research involving sensitive data.

Taguette's limitations lie in two main aspects: it lacks advanced analysis features and does not provide visual tools such as word clouds or network graphs, requiring integration with other tools (e.g., RQDA) for these purposes. Additionally, its OCR capabilities rely on external tools, making PDF text extraction potentially unstable.

Thus, Taguette is well-suited for individual researchers or small teams conducting qualitative data analysis. It is recommended for users who prioritize privacy or prefer open-source solutions. However, it is not ideal for complex statistical analysis or large-scale text mining. For more advanced analytical needs, it can be used in combination with RQDA or NVivo.

In general, due to limitations such as simplistic functions and infrequent maintenance, open-source software has seen relatively low adoption in Chinese academia. Searches using the names of the three open-source tools as keywords in the CNKI database yielded no results. Therefore, I hope that more researchers with programming expertise will participate in the development of these tools and contribute to their improvement. In the WOS database, the number of foreign publications using these three open-source tools is also relatively small: only 47 for RQDA, 28 for Weft QDA, and 26 for Taguette, respectively. Compared with the extensive use of commercial software, the usage volume of open-source software remains significantly lower. Consequently, no related literature is cited in this section.

3. Pathways for choosing qualitative data mining tools

The previous sections have provided a general overview of the currently commonly used Qualitative Data Analysis (QDA) tools, including their advantages, limitations, and appropriate application scenarios. The author has experimented with all of the aforementioned software during the analysis of fieldwork data, and due to budget constraints, ultimately adopted a combined solution using RQDA and Weft QDA. In addition, the author has taken elective courses in C# and PHP during university studies, and has also been exposed to other programming languages such as Ruby, Python, JavaScript, and R, thus possessing a certain level of programming competence. Therefore, future research will primarily rely on RQDA. Regarding the choice of software tools, the author believes that it is determined by three main factors: research needs, budget, and technical capabilities.

First, based on research needs, qualitative analysis can be divided into four categories: text-based research, multimedia data analysis, team collaboration projects, and mixed-methods research. For text-based research, when budget allows, user-friendly MAXQDA or feature-rich NVivo is recommended; when budget is limited, Taguette or RQDA can be considered. In the case of multimedia data analysis, due to the limited format support of open-source software, commercial tools such as NVivo or ATLAS.ti should be prioritized, with MAXQDA as an alternative. For team collaboration, both local and cloud-based options are available: Dedoose is preferred for cloud-based collaboration, while NVivo remains the primary choice for local teamwork. For mixed-methods research, MAXQDA or Dedoose is most suitable; for those with strong technical backgrounds, RQDA combined with R's quantitative analysis packages is also a viable option.

Second, researchers may select suitable tools based on team budget. For teams with over USD 1,000 allocated for data analysis tools, a comprehensive solution would be NVivo. If the focus is on theory building, ATLAS.ti may be more appropriate; for mixed-methods research, MAXQDA is better suited. For teams with a software budget between USD 100 and 1,000, a subscription-based tool like Dedoose is worth considering, as most commercial software offers around 50% academic discounts. However, many individual researchers or small teams may have no budget at all. In such cases, technically proficient researchers can use RQDA, while those with basic needs may find Taguette or Weft QDA sufficient.

Finally, the selection may also depend on the researcher's technical proficiency. For those lacking technical skills, commercial software such as MAXQDA—offering a good balance between usability and functionality—or NVivo—with comprehensive yet complex features—should be considered. For researchers with some technical background, more flexible theory-building tools like ATLAS.ti or the cloud-friendly Dedoose are recommended. Users with strong programming skills may fully customize their analysis workflows using RQDA or build pipelines with Python and NLTK.

Currently, qualitative data analysis software is showing three emerging trends: AI-assisted analysis, cloud-based collaboration, and mobile support. Regarding AI-assisted analysis, NVivo has begun integrating AI-based coding suggestions and new features, while MAXQDA has introduced automatic theme detection, and some commercial software can integrate with third-party tools like MonkeyLearn. For cloud-based collaboration, Dedoose offers a fully online solution, NVivo provides cloud collaboration plugins, and RQDA users can build their own solutions through Git version control. As for mobile support, MAXQDA offers mobile data collection applications, and ATLAS.ti now supports usage on iPads. As for how QDA tools will evolve in the future, we shall wait and see.

4. Conclusion

This paper compares six major QDA tools commonly used in academia and industry, clarifying their respective strengths, weaknesses, and key features. It proposes different pathways for selecting qualitative data mining tools based on research needs, budget, and technical capabilities. Researchers should choose appropriate tools that align with their specific requirements and technical skill levels. Beginners are advised to start with mature commercial software or simple open-source tools. In the future, qualitative data mining tools will increasingly evolve toward automation, cloud integration, and mobile support. Researchers must continuously enhance their skills to adapt to the progress of research methods and technological tools.

In conclusion, the author recommends that most novice researchers conducting text-based analysis begin with a feature-rich and user-friendly tool like MAXQDA. For those with programming skills, RQDA is a suitable option. Large-scale team projects should consider NVivo or Dedoose. Advanced users may use Taguette for initial coding, followed by in-depth analysis with NVivo,

and conclude with statistical and visualization work using R or Python. Regardless of the tool used—especially for commercial tools—it is advisable to start with free or demo versions. Most vendors offer a 30-day trial period for new users.

References

- [1] Fu, G., Chen, C., Pauli, C. (2005). Qualitative analysis techniques: A comparison of three research tools. In Chinese Psychological Society (Ed.), *Abstracts of the 10th National Academic Conference on Psychology* (p. 160). Jilin University Department of Social Psychology; South China Normal University Department of Psychology; University of Zurich Department of Educational Psychology.
- [2] Pandey, P., Sharma, A., Jeswani, S. (2025). Spotlighting recruitments: Is AI dominating human resource practices? Qualitative research using NVIVO. *Quality & Quantity, Advance online publication*, 1–20.
- [3] Meng, Y., & Shen, W. (2024). The scientific nature of qualitative data analysis: With a discussion on the application of NVivo in the field of education. *Journal of Soochow University (Educational Science Edition)*, 12(1), 24–34.
<https://doi.org/10.19563/j.cnki.sdj.2024.01.003>
- [4] Liu, H., Li, X., & Chen, X. (2025). A study on the influencing factors of the operation of points-based supermarkets in rural community governance: A qualitative text analysis based on MAXQDA software. *Journal of Shandong University of Technology (Social Sciences Edition)*, 41(1), 33–46.
- [5] Abraham, J. A., Krishnappa, P., & Federico, F. (2025). Perception of healthcare professionals on patient safety culture and associated factors: A qualitative study using MAXQDA software. *Frontiers of Nursing*, 12(1), 123–131.
- [6] Jiao, R., & Shilamaocao, S. (2025). Policy content analysis of forging a strong sense of community for the Chinese nation from the perspective of policy tools: A qualitative analysis of 32 national policy texts using ATLAS.ti9. *Qinghai Ethnic Studies*, 36(1), 79–89.
<https://doi.org/10.15899/j.cnki.1005-5681.2025.01.004>
- [7] Yunseon, C., Jiyeon, L., & Gyehee, L. (2022). Exploring values via the innovative application of social media with parks amid COVID-19: A qualitative content analysis of text and images using ATLAS.ti. *Sustainability*, 14(20), 13026.
<https://doi.org/10.3390/su142013026>