International Journal of Information Studies

Print ISSN : 1911 8408
Online ISSN : 1911 8414

DLINE JOURNALS

IJIS 2026: 18 (1)
https://doi.org/10.6025/ijis/2026/18/1/1-22

A Comparative Evaluation of Professional Book Indexing Software: Capabilities, Limitations, and Future Directions

Devendrappa T M
Research Scholar
Department of Library and Information Science
Kuvempu University, Sharkaraghatta. Shimoga, Karnataka. India
deva.tm@gmail.com

B. S Biradhar
Vice Chancellor-Bidar University
Bidar University, Gnyana Karanji, Halhalli (K), Bhalki (T), Bidar (Dist)
585414, Karnataka. India
bsbiradar53@rediffmail.com

ABSTRACT

The paper provides a comparative analysis of five professional indexing software tools CINDEX, MACREX, SKY Index, TExtract, and Index Manager evaluating their capabilities across 30 features grouped into seven categories: System Functionality, Indexing Process, Structure and References, Editing Tools, Quality Control, Output/Integration, and Automation vs. Manual Indexing. The significant findings reveal that Index Manager is the most well rounded, excelling in quality assurance, backup flexibility, spelling/error checking (using AI), and template support, though slightly limited in machine readable output formats. CINDEX stands out for its superior formatting control and broad compatibility with machine readable output. TExtract offers strong multilingual support and exceptional character support via LaTeX, along with robust backup features. SKY Index performs well in the structured production but is constrained by Windows only compatibility and limited subheading depth. MACREX lags, offering a fully manual workflow with minimal automation suitable only for expert indexers who prefer granular control. The study concludes that while indexing tools have advanced significantly, there remains no universal standard for multilingual or regional language indexing, highlighting a critical gap for future development. We emphasize the ongoing irreplaceability of human indexers, particularly in producing high quality, context aware book indexes, and express skepticism about AI's near term ability to match professional indexing standards.

Keywords: Indexing Software, Book Indexing, Text Retrieval, Automation vs. Manual Indexing, Multilingual Indexing, Information Retrieval Models, Professional Indexers

Received: 17 October 2025, Revised 30 November 2025, Accepted 6 December 2025

1. Introduction

Comprehensive text processing systems are utilized for the storage and retrieval of document collections, including newspaper archives, office automation platforms, and even libraries containing books and article titles. Traditional text processing systems do not adequately meet the demands of full text databases, which require features such as document ranking and text oriented indexing, unlike key based indexing. Text indexing and information retrieval are essential for efficient information searches.

They structure the content for rapid access and employ various methods to connect queries with pertinent information. Familiarity with these models deepens our understanding of how search engines operate. Boolean, vector space, and probabilistic models each have their advantages and drawbacks; choosing the appropriate model and indexing method is vital for ensuring speed, precision, and relevance in information retrieval systems. Text indexing establishes a systematic representation of a set of text documents to enhance efficient searching and retrieval of pertinent information. Indexing identifies critical terms from documents, organises them into an inverted index, and associates them with the documents in which they appear. Text indexing enables quick, precise retrieval of relevant documents in response to user queries, minimising the need to examine the entire collection. [1]

2. Early Studies

Full text systems require an index to enable quick, reliable access to documents based on their content. [2]. Text retrieval has been a prominent area of research in information retrieval, where a system is tasked with delivering relevant information resources in response to users' natural language queries. The evolution of retrieval models has progressed from heuristic based approaches to learning driven ranking functions, adapting to rapid technological advancements.

A significant hurdle in crafting effective retrieval models is the learning of text representations and the modeling of relevance matching. For over three decades, expert indexers have relied on specialized indices to meticulously compile indexes for books, journals, serial publications, databases, and various other documents. Book indexes enable readers to find the information they need quickly. Essential criteria for a book index encompass completeness (it should direct the reader to all relevant information within the book), navigability (it should help the reader locate subtopics and related subjects), and accuracy (it must avoid false or incorrect information and reflect the author's terminology).

Studies assessed whether artificial intelligence (AI) utilizing large language models (LLMs) could produce indexes that satisfy these requirements, and we discovered that such indexes are significantly lacking. Presently, AI is not capable of substituting professional book indexers, and we remain skeptical about its ability to do so in the near future or at any point. [3]

Professional text indexes are tailored to assist indexers working from printed proofs, digital text, the author's manuscript, or a finalized book. The index is produced as a fully independent document; it is not created by tagging or marking the original text, although it is feasible to insert tagged index entries into HTML and XML documents automatically. The purpose of MACREX is to aid indexers in enhancing consistency and boosting productivity by automating routine processes (such as sorting, printing, and repagination), allowing them to

concentrate on the phrasing and structure of index entries. [4]

Advancements in digital publishing have introduced challenges regarding how indexes are formed and displayed. Ongoing research is exploring technological solutions that necessitate collaboration between publishers and indexers.

Content indexing is extensively utilized to search across various applications efficiently. [5] While inverted indexes have been predominantly used for exploring data that can be effectively divided into words, like natural language texts, they may not capture all matching results from data that is challenging to segment into a manageable number of words, such as logs, code, datagrams, sensor signals, bioinformatics samples, and scientific data. [6]. With the ongoing development of suffix index technologies over the last decade, [7, 8]. Indexing systems have recently transitioned to innovative search systems that employ suffix indexes for full text searches across large datasets, rather than relying on inverted indexes.

3. Web-based Indexes

The indexing system not only indexes web pages but also performs information extraction, filtering, and classification, providing metadata, metrics, and feedback such as link analysis and query independent ranking features to both the crawling and query processing systems. By processing the crawler's output and supplying input to the query processor, it serves as a crucial bridge between these two components. [9]

Only a small number of search engines crawl the web at scale, and third party developers are left waiting for crumbs from those few vendors to build downstream applications based on search. [10] While the massive Common Crawl corpus is publicly accessible, it remains more affordable to crawl and index a smaller, application specific dataset than to operate a full sized general purpose index like Common Crawl motivating our construction of the Open Web Index.

It is intended to be a publicly funded common infrastructure on top of which anyone can easily and transparently build custom search indexes for their use case. We aim to place it and its related data as a new, open intermediary for web information. Many studies describe the inception prototype, discuss planned future development, and outline ways in which efforts in the information retrieval community could leverage or assist the effort, such as through resources, preprocessing tools/pipelines, or the development of specialised search engines/ test collections. [11-26]

4. Dataset and Methodology

4.1 Dataset: We selected five indexing models and identified 30 features for study and comparison. The data primarily consists of five indexes: Cindex, Macrex, SKY Index, Extract book indexing software, and Index Manager. These are the primary indexing tools used across disciplines. These five indices are outlined as follows.

4.11 Cindex: is an indexing program for preparing an index of books, newspapers, and periodical publications. It provides the professionals with an indexing feature which can be used to prepare glossaries or subject authority lists based on its index [6].

4.12 Macrex: is computer based indexing software designed to assist indexing professionals in creating the indexes from printed proofs, text available in digital format, the author's manuscript, and already completed books. The index created using this software has become a completely independent document. It is not constructed by tagging, although it provides the ability to automatically insert tagged index entries into HTML and XML files [6].

4.13 SKY Index: is a Windows based computer software package designed to assist professional indexers in generating or compiling back of book style indexes. This software is considered the "look and feel" of each indexing document. It enables indexers to perform data entry and editing with relative ease, sort data, manage indexing features, provide cross references, support formatting, define authority structures, and produce printed and machine readable output for online and printed documentation [6].

4.14 TExtract book indexing software: TExtract, also known as Texyz, specialises in automatic and semi-automatic indexing software and services for analysing and accessing text information from books and text data from survey sources. This software can index content in different languages. This software is used in universities, publishing companies and editing services in over 60 countries worldwide [6].

4.15 Index Manager: is indexing software used to generate indexes for book, catalogue, technical documents and search applications in digital products [6].

4.2 Methods

The five indices provide a lead on how indexing is carried out.

We classified the thirty features into seven convenient groups to extract and document the data. The seven groups include: System and Basic Functionality, Indexing Process and Data Entry, Structure and References, Tools and Editing options, Quality Control and Backup, Output and Integration, and Automated vs. Manual Indexing Features. These seven features perform as the baseline for a meaningful systematic study.

5. Data Analysis

We used the five indices to understand and study how they function as a tool. The features that we used are Operating System, Types of Materials indexed, Number of indexes open and size of the records. The resulting data is presented in Table 1.

System and Basic Functionality

Feature	Cindex	Macrex	SKY Index	TExtract	Index Manager
Operating System	Windows or Mac	Windows	Windows	Windows or Mac	Windows or Mac
Type of Material Indexed	Books, Newspapers, Periodicals, etc.	Printed proofs, digital text, manuscripts, books.	Back of book style indexes.	All electronic documents create a back of book index.	Book, catalogue, technical documents, and search applications.

Number of Indexes Open	Many (limited by memory)	Multiple	Only One	Multiple	Multiple
Record Size	Default 100 characters (can be increased)	Up to 100,000 entries (instant checking/editing)	Limit is 255 characters per field	NA	Limit is 250 characters per field

Table 1. System and Basic Functionality

Table 1 compares key functional and technical characteristics of five indexing software tools: CINDEX, MACREX, SKY Index, TExtract, and Index Manager. The comparison focuses on four core features relevant to indexing workflows:

1. Operating System Compatibility – Indicates whether the software runs on Windows, Mac, or both.

2. Type of Material Indexed – Describes the range of document types each tool is designed to handle (e.g., books, manuscripts, electronic documents).

3. Number of Indexes Open Simultaneously – Reflects the software's multitasking capability, from handling only one index at a time to supporting multiple or many indexes (limited by system memory).

4. Record Size Limitations – Specifies constraints on entry or field length, which may affect the detail and structure of index entries.

This table provides a concise, side by side overview to help evaluate and select indexing software based on the technical requirements and project scope.

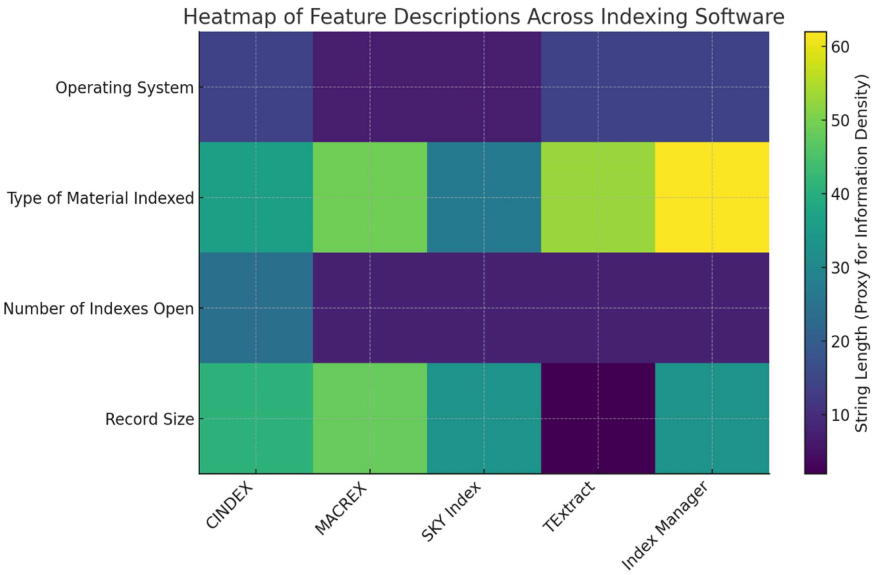

Figure 1. Heat Map of the record size for the five tools

This heat map (Figure 1) visualises the Record Size comparison across the five indexing software tools:

- Higher intensity corresponds to larger record size capacity.

- "NA" is shown where information was not available (TExtract).

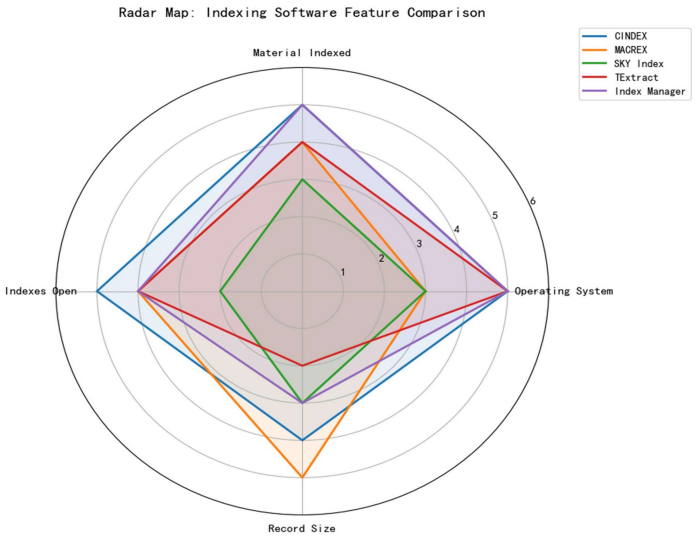

Figure 2. Radar Map of the studied features of the system functionality

The radar map (spider chart) in Figure 2 compares the five indexing tools across four features: Operating System, Type of Material Indexed, Number of Indexes Open, and Record Size.

Each axis represents one feature (scored 1–5 based on capability/flexibility), and each colored polygon corresponds to a software tool. The larger the area, the more capable the tool across these dimensions. This visualisation highlights the following facts.

- CINDEX is the most balanced and capable overall (especially in OS support and multitasking).
- SKY Index is the most limited (Windows-only, single-index, narrow material scope).
- Index Manager and TExtract offering good OS flexibility but moderate performance on other metrics.
- MACREX excels in record/entry handling (scored highest on "Record Size" due to large entry capacity), despite being Windows only.

Next, we studied how the indexing model operates and how data should be entered. The data entry process is significant because it determines the time required.

Indexing Process and Data Entry

Feature	CINDEX	MACREX	SKY Index	TExtract	Index Manager
Process of Indexing	Fully automated indexing in virtually any format.	Indexing creating tool (similar to a word processor, not automatic).	Fully automated indexing process with editing options.	Automated index of the initial Index.	Automated indexing and manual process option.

Data Entry	The data entry metaphor is a prominent distinguishing feature.	Manual entry feature.	The data entry metaphor is a prominent distinguishing feature.	Other than the initial index, the user can add manually.	Automatically generates the index from the document.
Type-a head Feature	Auto entry type-ahead feature available.	Manual Entry of index terms.	Auto Completion type-ahead feature available.	Auto entry type-ahead feature available.	Auto entry type-ahead feature available.

Table 2. Indexing Process and Data Entry

Table 2 highlights the features such as the Indexing process, data entry, and type-ahead across five source indexes, which is also visualised below.

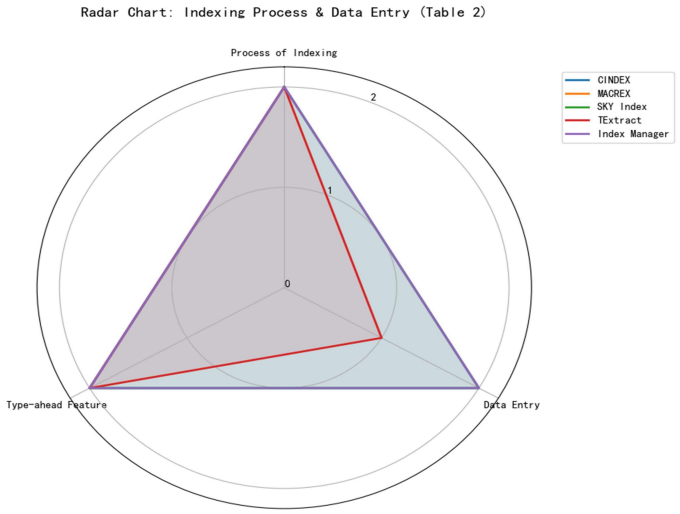

Figure 3. Radar Map for the Indexing Process and Data Entry

The radar map (Figure 3) for Indexing Process and Data Entry is evaluated using a qualitative scoring (0–2) for three key features: Indexing Process, Data Entry, and Type-ahead Feature.

We derived the essential insights from the data as follows.

- MACREX scores lowest (0 across all), reflecting its manual, non-automated approach.
- CINDEX, SKY Index, and Index Manager all score maximum (2), indicating strong automation, adequate data entry support, and type-ahead functionality.
- TExtract scores 2, 1, 2—fully automated initial indexing and type-ahead, but limited manual entry afterwards.

The chart visually highlights the contrast between fully automated tools and manual or semi-automated alternatives.

Feature	CINDEX	MACREX	SKY Index	TExtract	Index Manager
Provision for Subheading	Up to 16 levels of heading/ subheading.	Allows for adding many subheadings.	Allows only five levels of subheading.	Allows for adding many subheadings.	Allows for adding any subheading.
Page reference	Set Automatically	An index can be presented in page-number order; separate indexes can be created by extracting entries from a page range.	Set Automatically	Set Automatically	Set Automatically
Cross reference	Programs let you specify the terms you intend to use for cross-references.	Programs let you specify the terms you intend to use for cross-references.	Programs let you specify the terms you intend to use for cross-references.	Programs let you specify the terms you intend to use for cross-references.	Programs let you specify the terms you intend to use for cross-references.
Authority Control	Creating multiple authority files is possible.	Creating multiple authority files is possible.	Creating multiple authority files is possible.	Creating multiple authority files is possible.	Creating multiple authority files is possible.

Table 3. Structure and references in the five indices

In Table 3, we examine features such as content structure and reference inclusion alongside a textual description.

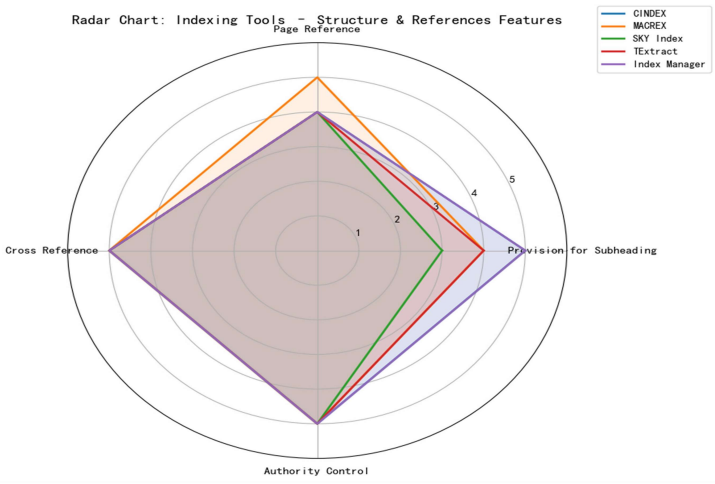

Figure 4. Visualisation of structure and references

Figure 4 evaluates each tool across four key features, scored on a 1–5 scale (higher = stronger/more flexible):

1. Provision for Subheading

- CINDEX and Index Manager score highest (5) due to support for 16 levels and unlimited subheadings, respectively.
- SKY Index is the most limited (3) with only five levels.
- MACREX and TExtract (4) allow "many" subheadings but lack explicit highlevel support.

2. Page Reference

- All tools automatically set page references.
- MACREX gets a slight edge (5) because it allows indexes by page range a unique advanced feature.

3. Cross Reference

- All tools score equally (5): each allows users to define cross reference terms.

4. Authority Control

- All tools score equally (5): all support multiple authority files.

We derived a few key insights from this table and visualisation.

- CINDEX and Index Manager lead in structural flexibility, especially for complex, hierarchical indexes.
- SKY Index is the most restrictive in subheading depth, which may limit use in significant scholarly works.
- MACREX stands out in page-reference customisation, despite average subheading support.
- All tools are equivalent in cross-referencing and authority control core professional indexing functions.

Feature	**CINDEX**	**MACREX**	**SKY Index**	**TExtract**	**Index Manager**
Editing	Editing feature available.	Editing feature available.	Editing feature available.	Editing feature available.	Editing feature available.
Find and Replace	The word processor find and replace function is used.	The word processor find and replace function is used.	The word processor find and replace function is used.	The word processor find and replace function is used.	The word processor find and replace function is used.
Splitting and Joining	Splitting of heading and subheading is possible.	Splitting of heading and subheading is possible.	Splitting of heading and subheading is possible.	Splitting of heading and subheading is possible.	Splitting into levels is possible.

Sorting	Letter by letter, word by word in ASCII format is available.	The program allows a record sorting feature.	Letter by letter, word by word, in ASCII format is available.	The program allows a record sorting feature.	The program enables a record sorting feature.
Speech recognition	Supports Dragon Naturally Speaking.	Supports Dragon Naturally Speaking language.	Supports Dragon Naturally Speaking.	Supports Dragon Naturally Speaking.	Supports Dragon Naturally Speaking language.
Macros	Up to 10 Macros can be created and stored.	Up to 48 Macros can be created and stored.	Up to 20 Macros can be created and stored.	Allows the creation of Macros.	Enables the creation of Macros.
Grouping	Grouping facility is available.	Grouping facility is possible.	Grouping facility is available.	Grouping facility is available.	Grouping facility is available.
Selecting records	Selecting keywords for creating a group facility is available.	Users can choose the records for creating an indexing group.	Selecting keywords for creating a group facility is available.	Users can select records to develop an indexing group.	Selecting keywords for creating a group facility is available.
Duplicating Records	Duplicating single records and bulk records is possible.	Allows the creation of multiple index records.	Duplicating single records and bulk records is possible.	Enables the creation of numerous index records.	Allows the creation of multiple index records.

Table 4. Tools and Editing Options

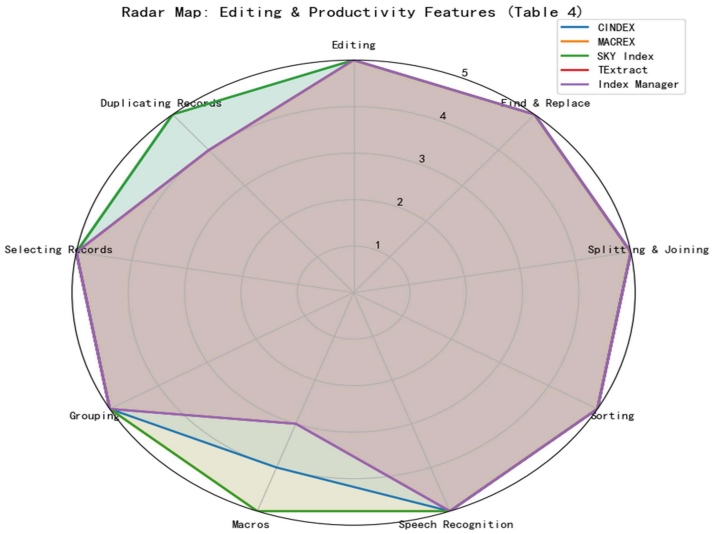

Figure 5. Tools and Editing Options Visualisation

All five tools (Table 4 and Figure 5) perform very similarly across most editing features reflecting industry-standard capabilities in professional indexing software. However, subtle differences emerge in two key areas:

1. Macros Support

- MACREX leads with 48 macros → highest score (5)
- SKY Index (20 macros) also scores 5
- CINDEX (10 macros) scores 4
- TExtract and Index Manager only state "allows macros" with no limit → conservatively scored 3

Implication: Users who rely heavily on automation (e.g., repetitive formatting, batch edits) may prefer MACREX or SKY Index.

2. Duplicating Records

- CINDEX and SKY Index explicitly support bulk duplication → score 5
- The other three state "allows multiple records" but lack explicit bulk functionality → scored 4

Implication: For largescale index restructuring, CINDEX and SKY Index offer more efficient workflows.

3. Uniform Strengths

All tools score 5 in:

- Editing
- Find & Replace
- Splitting/Joining headings
- Sorting
- Speech recognition
  (via Dragon Naturally Speaking)
- Grouping
- Selecting records

Feature	Cindex	Macrex	SKY Index	TExtract	Index Manager
Checking spelling	Comes with an English spelling dictionary.	Any corrections that need to be made manually.	Need to use a word processor for checking, and reload it after checking.	Comes with an English spelling dictionary.	Software has an intelligent algorithm system to find different spellings.
Error Trapping	Checks the index for errors on the fly or in a batch process.	Errors in the placement of punctuation and spaces are automatically corrected.	Checks the index for errors on the fly or in a batch process.	Checks the index for mistakes on the fly or in a batch process.	Checks the index for errors on the fly or in a batch process.
Handling of Special Characters	Uses its own management system.	Index accuracy is checked, and spelling, punctuation, and spaces are automatically corrected.	Uses the character map that comes with Windows.	Handled by LaTeX software.	Uses its own management system.
Backup	Allows the creation of a copy of a compact "archive" file.	Allows saving a backup file as a text file.	Supports a backup facility as an archive file format.	Allows saving backups as text files and archives.	Provides source documents and work files backup.
Statistics Generating	Provides different types of statistical information.	Provides different kinds of statistics.	Offers different types of statistical information.	Provides different types of statistics.	Provides different kinds of statistics.

Table 5. Quality Control and Backup

The quality control and backup features are described in Table 5, and the radar map (Figure 6)

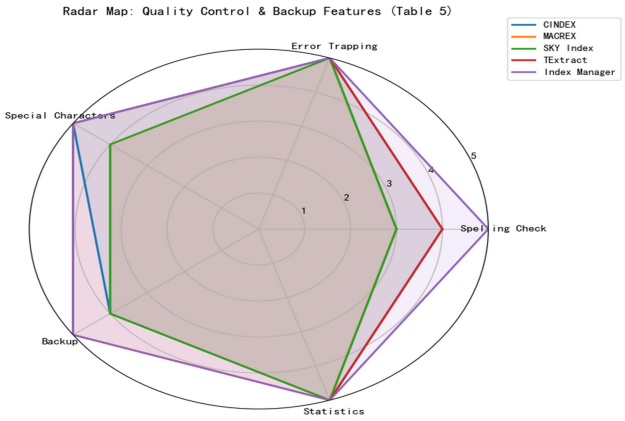

Figure 6. Quality Control and Backup

Table 5 and Figure 6 evaluate five indexing tools across five quality and reliability features, scored on a 1–5 scale (higher = more robust):

1. Spelling Check

- Index Manager (5): Uses an intelligent algorithm to detect multiple spelling variants—most advanced.
- CINDEX & TExtract (4): Include a built in English dictionary.
- MACREX & SKY Index (3): Rely on manual correction or external word processors, making them less efficient.

2. Error Trapping

- All tools score 5: Each offers real-time or batch error checking.
- MACREX uniquely auto-corrects punctuation and spacing, though this didn't raise its score since all meet the core requirement.

3. Handling of Special Characters

- CINDEX, TExtract, Index Manager (5):
- CINDEX & Index Manager: own character management systems
- TExtract: leverages LaTeX, which is highly capable for technical/special characters
- MACREX & SKY Index (4): Use Windows Character Map or basic auto-correction—functional but less flexible.

4. Backup

- Index Manager & TExtract (5): Support multiple backup formats, including source documents, text files, and archives.
- CINDEX, MACREX, SKY Index (4): Offer archive or text backups, but not both plus source files.

5. Statistics Generation

- All tools score 5: Each provides various index statistics (e.g., entry counts, term frequency)a standard professional feature.

Key Derivations

- Index Manager leads overall, scoring 5 in every category, reflecting comprehensive quality control and backup.
- TExtract is a close second, matching Index Manager in all but spelling (4 vs. 5), likely due to reliance on a standard dictionary rather than AI-driven matching.
- CINDEX performs well but lacks advanced spelling intelligence and multi format backup.
- MACREX and SKY Index are weaker in spelling and character handling, indicating greater reliance on external tools for quality assurance.

Feature	Cindex	Macrex	SKY Index	TExtract	Index Manager
Template	Allows to import the indexing template from other software.	No such template available; can import.	Allows to import the indexing template from other software.	Default Template is available.	Default Template is available.
Importing/ Exporting	Can import backup files made by Macrex.	Import/Export provides index in RTF / TXT format.	Can import backup files made by Macrex.	Can import backup files made by other indexing software.	Possible in various file formats like XML, IDML, RTF / TXT.
Formatting	Formatting options are infinite.	Formatting is possible.	Can set out formatting for all levels of subheading.	Formatting is available for end users.	Formatting is available for end users.
Out Put	Provide print and machine readable output with different file formats.	Text file format.	Provide print and machine readable output with different file formats.	Provides EPBS and HTML based output format.	Output is available in XML and IDML format.
Machine Readable Index	RTF, ASCII, Text file, Quark express, AAP, Chicago and HTML, etc.	RTF / TXT – ASCII format.	RTF, ANSI/ NIOS, Text ASCII file, Quark express, AAP, Chicago and HTML, UCP tag, etc.	RTF / TXT – ASCII format.	RTF / TXT – ASCII format.

Table 6. Output and Integration

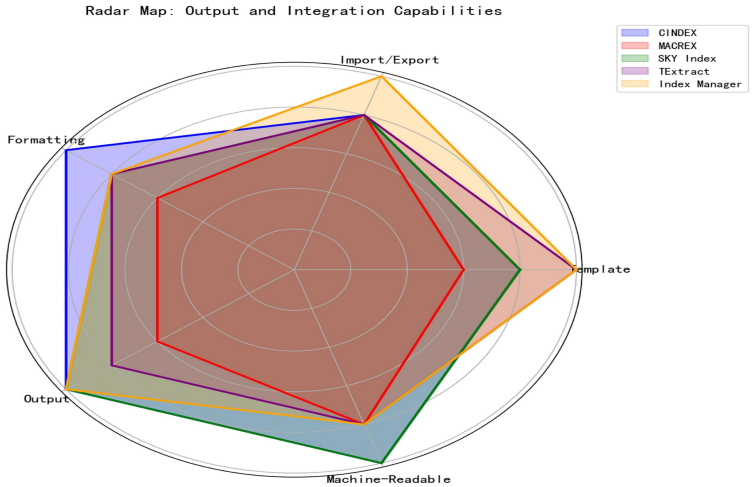

Figure 7. Output and Integration

The radar chart 7 compares five indexing tools across five critical output and interoperability dimensions. Here's what the data reveals:

1. Template Support

- Index Manager & TExtract (5): Offer default templates, enabling faster setup.

- CINDEX & SKY Index (4): Support importing templates from other software—flexible but requires external files.

- MACREX (3): Lacks native templates and offers only basic import—least supportive for standardized workflows.

2. Importing/Exporting

- Index Manager (5): Stands out with XML, IDML, RTF, TXT—ideal for publishing pipelines (e.g., Adobe InDesign via IDML).

- All others (4): Limited to RTF/TXT or basic cross-software imports—functional but less versatile.

3. Formatting Flexibility

- CINDEX (5): "Infinite formatting options" make it ideal for custom typographic control.

- SKY Index (4): Allows level-specific formatting—good for structured indexes.

- MACREX (3): Only basic formatting—may require post-processing.

4. Output Variety

- CINDEX, SKY Index, Index Manager (5): Deliver both print-ready and machine-readable outputs in multiple formats.

- TExtract (4): Supports EPUB and HTML—strong for digital publishing, but fewer print options.

- MACREX (3): Text-only output—least flexible for professional publishing.

5. Machine-Readable Index Support

- CINDEX & SKY Index (5): Support QuarkXPress, HTML, AAP, Chicago, UCP tags, etc.—excellent for legacy and modern systems.

- Others (4): Limited to RTF/TXT/ASCII, sufficient for basic exchange but not advanced integration.

Key Derivations

- Index Manager excels in modern interoperability (XML/IDML) and template usability, making it ideal for digital publishing workflows.

- CINDEX leads in formatting control and broad machine-readable support, best for custom print indexes.

- SKY Index is a strong all-rounder in output but weaker in templates and macros (from prior tables).

• MACREX is the least capable in output flexibility, relying heavily on plain text may suit simple projects only.

• TExtract bridges digital formats (EPUB/HTML) but lacks CINDEX's depth in formatting or SKY's legacy compatibility.

Program	Primary Indexing Process	Key Data Entry Feature	Page Reference Setting	Spelling/Error Checking
CINDEX	Fully automated indexing in virtually any format.	The data entry metaphor is a prominent distinguishing feature.	Set Automatically	Comes with an English spelling dictionary.
MACREX	Indexing creating tool, not automatic (similar to a word processor).	Manual Making entries feature.	Manual (Index presented in page number order).	Manual corrections needed; errors in punctuation/ spaces are auto-corrected.
SKY Index	Fully automated indexing process with editing options.	The data entry metaphor is a prominent distinguishing feature.	Set Automatically	Requires a word processor for checking, then reloads.
TExtract	Automated index of the initial Index.	Other than the initial index, the user can add manually.	Set Automatically	Comes with an English spelling dictionary.
Index Manager	Provides an option for Automated indexing and manual process of indexing.	Automatically generates the index from the document.	Set Automatically	Software has an intelligent algorithm system to find different spellings.

Table 7. Automated vs. Manual Indexing Features

Table 7 and Figure 8 outline how the indexing models respond to the automation and how they move from manual indexing processes.

Automated vs. Manual Indexing Features, accurately reflecting the capabilities of each indexing tool across four key dimensions:

• Indexing Automation

- Data Entry Flexibility
- Page Reference Handling
- Spelling & Error Checking

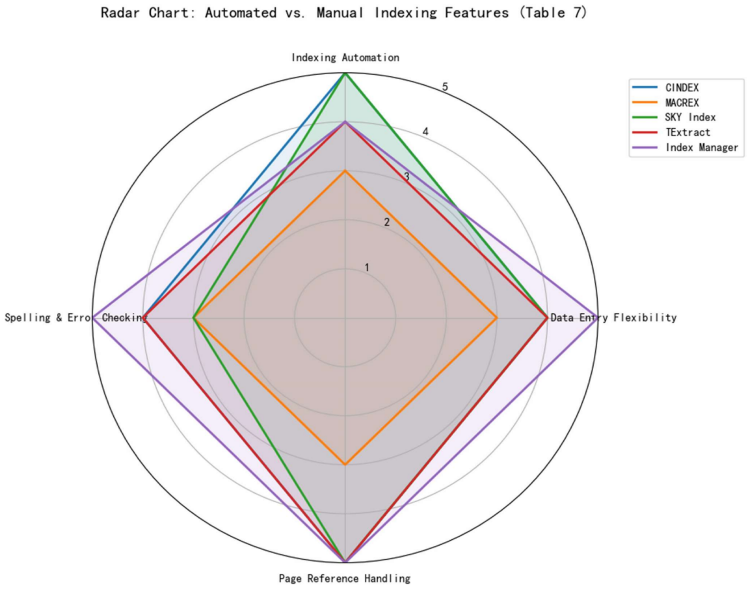

Figure 8. Automated vs. Manual Indexing Features visualisation

Key Observations from the Chart

- Index Manager (purple) leads in Spelling & Error Checking and Data Entry, with strong performance across all areas.
- CINDEX (blue) and SKY Index (green) are fully automated and handle page references automatically, but SKY Index lacks built in spelling support.
- TExtract (red) offers a hybrid approach: automated base index + manual additions.
- MACREX (orange) is the only truly manual tool, scoring lowest in all categories - suitable for expert users who prefer complete control.

Table 8 provides a comprehensive summary of the features and their respective scores obtained in the evaluation.

The above table is consolidated below with the overall scores.

Feature	Cindex	Macrex	SKY Index	TExtract	Index Manager
Spelling & Error Check	4 (dict)	3 (manual)	3 (WP-dependent)	4 (dict)	5 (AI spelling)
Special Characters	5 (own system)	4 (auto-correct)	4 (Win char map)	5 (LaTeX)	5 (own system)
Backup Flexibility	4 (archive)	4 (text)	4 (archive)	5 (text + archive)	5 (full source + work files)
Template Support	4 (import)	3 (no native)	4 (import)	5 (default)	5 (default)
Import/ Export	3 (limited compat.)	4 (RTF/TXT)	3 (limited compat.)	4 (some compat.)	5 (XML, IDML, RTF, etc.)
Formatting Control	5 (“infinite”)	3 (basic)	4 (per-level)	4 (user-available)	4 (user-available)
Machine-Readable Output	5 (HTML, Quark, AAP, etc.)	4 (RTF/TXT)	5 (HTML, Quark, UCP,	4 (RTF/TXT)	4(RTF/TXT)

Table 8. Scores of the studied indexes using seven features

Tool	Total Score
Index Manager	31
CINDEX	30
SKY Index	28
TExtract	29
MACREX	25

Table 9. Total Scores Compared (out of 35)

Table 9 presents the final evaluation scores for each index and provides an inference. The differences between the five indices in terms of evaluation scores are marginal; however, in terms of functionality, the scores reveal a significant impact.

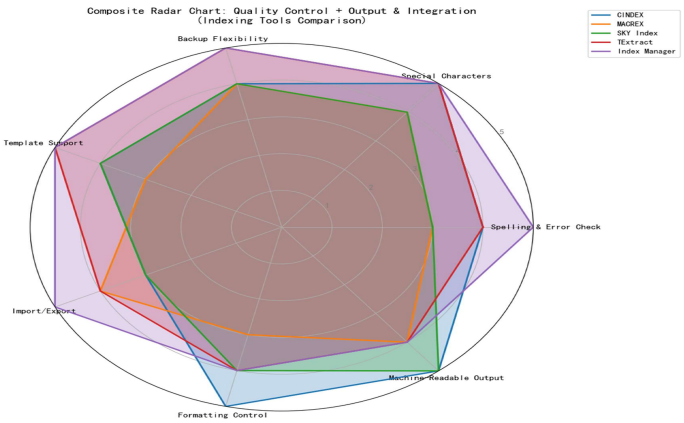

Figure 9. An interactive-quality radar chart visualising the seven dimensions

The interpretation and derivation of the features identified are described below.

6. Final Assessment

• Index Manager stands out as the most well rounded option: it shines in quality assurance, backup capabilities, and template functionality, although it is somewhat less proficient in machine readable formats.

• CINDEX is at the forefront when it comes to output flexibility and formatting, boasting robust quality features.

• TExtract is a solid all-around choice, particularly excelling in backup and handling special characters (through LaTeX).

• SKY Index is strong in output but has limitations in integration. • MACREX consistently trails behind ideal for users who favour manual control over automated processes. Index Manager (purple) is the most comprehensive option, earning a 5 out of 7 across five categories, particularly excelling in automation, backup, and template support. CINDEX (blue) excels in formatting management and machine readable output, although it falls slightly short in import/export options. TExtract (red) is particularly strong in backup and special character handling (via LaTeX), offering a well rounded performance overall. SKY Index (green) is comparable to CINDEX in machine-readable formats but is less effective in import/export and spelling features. MACREX (orange) consistently receives the lowest scores most suitable for users who prefer manual, controlled processes that do not rely on automation.

Macrex and SKY Index are Windows only, while CINDEX, TExtract, and Index Manager are compatible with both Windows and Mac. Our findings indicate that all five indexing systems can handle various types of materials; there is no overlap in this capability. All models, except Macrex, can perform a fully automated indexing process with editing options. Each indexing system defaults to different templates, but CINDEX and SKY Index allow importing templates from other platforms. Regarding the import and export of indexes, the study shows that any indexing software can import backups into compatible software via the import function, and all systems can export indexing files in various formats that are not unique. Except for Macrex, all indexing software can check for index errors in real time or in batch. Each software can create and save varying numbers of macros, as illustrated in the table. All examined systems can produce different index file formats.

All systems can generate various machine-readable index formats, with text, RTF, and ASCII being the most common.

7. Conclusion

Cataloguing literary pieces in an online setting, whether in English or other languages, poses considerable difficulties. Moreover, reliably cataloguing literature in regional languages online is particularly challenging. As the IT sector continues to evolve and indexing software progresses, various tools and methods currently on the market are being assessed. Among the leading five indexing software choices, Cindex and Textract book indexing software stand out as the most efficient and share a similar methodology for indexing English literature. The research indicates that no all-encompassing standard for indexing multilingual content exists, underscoring the need for a universal standard bibliography to encompass literary works in multiple languages. To establish consistency in the indexing database for regional language literature, developers of indexing software should prioritize creating solutions to tackle these persistent issues.

References

[1] Five able. (2024, August 14). Text indexing and retrieval models Natural Language Processing. https://five able.me/natural-language-processing/unit-11/text-indexing-retrieval-models/study-guide/CjNim3yQfiwwDQoR

[2] An Efficient Indexing Technique for Full Text Database Systems Justin Zobel. (1992). *In:* Proceedings of the 18th *VLDB Conference Vancouver, British Columbia, Canada*. p 352-363.

[3] American Society of Indexers. (2025). LM-Generated Book Indexes: Can They Replace Professionally Created Indexes White paper. https://asindexing.org/ai-news/white-paper-ai-index/).

[4] Society of Indexers. (2025). Indexing and Publishing Technology. https://www.indexers.org.uk/?id=104].

[5] Park, Sehwa., Park, Seog. (2019). Reverse collective spatial keyword query processing on road networks with G-tree index structure. *Information Systems*, Volume 84, Pages 49-62.

[6] Wentao, Xu., Chen, Haoyu., Huan, Yidong., Hu, Xuedong., Nong, Ge. (2022). Full-text search engine with suffix index for massive heterogeneous data*, Information Systems*, Volume 104, 101893

[7] Traina, J. M., Agma., Brinis, Safia., Pedrosa, V. Glauco., Avalhais, P. S. Letricia., Traina, Caetano. (2019). Querying on large and complex databases by content: Challenges on variety and veracity regarding real applications, *Jr Information Systems*, Volume 86, December Pages 10-27

[8] Arroyuelo, Diego., Bonacic, Carolina., VGil-Costa, Verónica., Marin, Mauricio., Navarro, Gonzalo. (2014). Distributed text search using suffix arrays, *Parallel Computing*, Vol 40 (9), October Pages 471-495

[9] Cambazoglu, B. B., Baeza-Yates, R. (2016). The Indexing System. In: Scalability Challenges in Web Search Engines. *Synthesis Lectures on Information Concepts, Retrieval, and Services. Springer,*

[10] Hendriksen, G. *et al*. (2024). The Open Web Index. In: Goharian, N., *et al*. Advances in Information Retrieval. ECIR 2024. *Lecture Notes in Computer Science,* vol 14 (6), p. 12. Springer, Cham).

[11] Formal, T., Piwowarski, B., Clinchant, S. (2021). SPLADE: Sparse Lexical and Expansion Model for First Stage Ranking, SIGIR 2021, p. 2288-2292. *Association for Computing Machinery, New York, Isbn* 978145 0380379.

[12] Fröbe, M., et al. (2023). The Information Retrieval Experiment Platform. *In:* Proceedings of the 46th International ACM *SIGIR Conference on Research and Development in Information Retrieval.*

[13] Fröbe, M., et al. (2023). Continuous integration for reproducible shared tasks with TIRA.io. In: Advances in Information Retrieval. 45th European Conference on IR Research (ECIR 2023). LNCS. *Springer*. https://doi.org/10.1007/978-3-031-28241-6_20

[14] Gao, L., et al. (Dec 2020). The Pile: An 800GB Dataset of Diverse Text for Language Modeling.

[15] Goel, S., Broder, A. Z., Gabrilovich, E., Pang, B. (2010). Anatomy of the long tail: ordinary people with extraordinary tastes. *In:* Davison, B.D., Suel, T., Craswell, N., Liu, B. (eds.) Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, 4-6 February 2010, p. 201–210. *ACM, New York*.

[16] Gollub, T., Potthast, M., Stein, B. (2018). Shaping the Information Nutrition Label. In: Albakour, D., Corney, D., Gonzalo, J., Martinez, M., Poblete, B., Valochas, A. (eds.) 2nd International Workshop on Recent Trends in News Information Retrieval (NewsIR 2018) at ECIR. *CEUR Workshop Proceedings,* vol. 2079, p. 9–11, ISSN 1613-0073.

[17] Granitzer, M., Voigt, S., et al. (2023) Impact and Development of an Open Web Index for Open Web Search. *J. Assoc. Inform. Sci. Technol.*

[18] Guha, R. V., Brickley, D., MacBeth, S. (2015). Schema.org: evolution of structured data on the web: big data makes common schemas even more necessary. *Queue* 13 (9), 10–37, ISSN 1542-7730.

[19] Kamphuis, C., Hasibi, F., Lin, J., de Vries, A. P. (2022). REBL: entity linking at scale. In: Alonso, O., Baeza-Yates, R., King, T.H., Silvello, G. (eds.) Proceedings of the Third International Conference on Design of Experimental Search Information Retrieval Systems, San Jose, CA, USA, 30-31 August 2022. *CEUR Workshop Proceedings,* vol. 3480, p. 68–75. *CEUR-WS.org*.

[20] Khattab, O., Zaharia, M. (2020). ColBERT: efficient and effective passage search via contextualized late interaction over BERT. *In:* Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR , p. 39-48. *Association for Computing Machinery, New York* ISBN 9781450380164.

[21] Koster, M., Illyes, G., Zeller, H., Sassman, L. (2022). RFC 9309 Robots Exclusion Protocol.

[22] Kreutzer, J., et al. (2021). Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets.

[23] Lewandowski, D. (2019). The web is missing an essential part of infrastructure: an open web index. Commun. *ACM* 62 (4), 24

[24] Li, H., Su, Y., Cai, D., Wang, Y., Liu, L. (2022). A Survey on Retrieval-Augmented Text Generation. arXiv preprint arXiv:2202.01110

[25] Lin, J., et al. (2020). Supporting interoperability between open-source search engines with the common index file format. *In:* Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, p. 2149–2152.

[26] Middleton, S. E., Kordopatis-Zilos, G., Papadopoulos, S., Kompatsiaris, Y. (2018). Location extraction from social media: geoparsing, location disambiguation, and geotagging. ACM Trans. *Inform. Syst*. (TOIS) 36(4), 1–27.