



MyBox: Easy Tools Set

User Guide – Document Tools

Author: Mara
Version: 6.5.9
Date: 2022-8-31

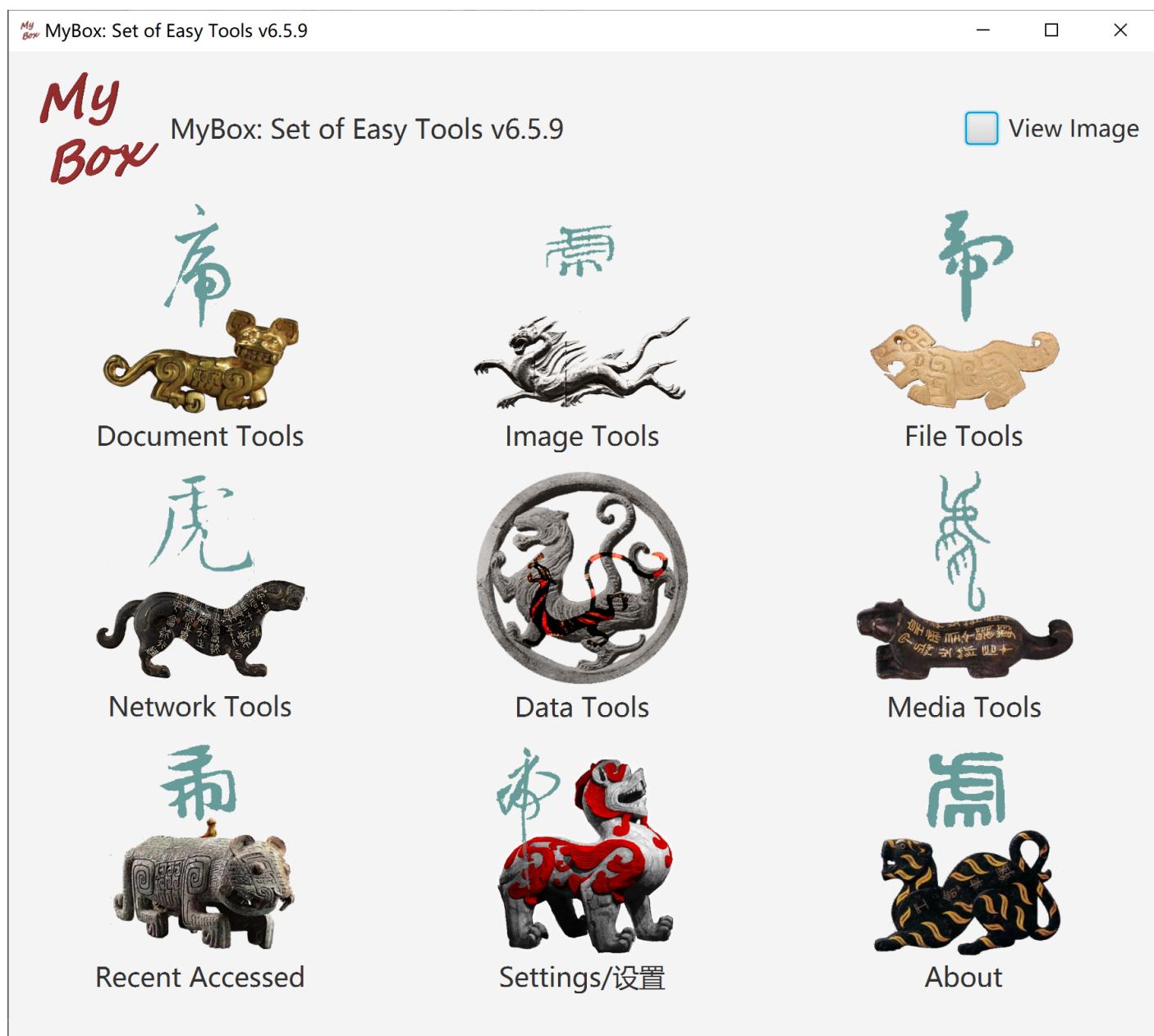
Contents

1	Introduction.....	3
1.1	Main Interface.....	3
1.2	Resources Addresses.....	4
1.3	Documents.....	5
1.4	Menu of Tools.....	5
2	Information in Tree.....	6
3	Notes.....	7
4	PDF Tools.....	9
4.1	View PDF file:.....	9
4.2	Convert PDF files in batch.....	9
4.3	Extract data in batch.....	10
4.4	Others.....	10
5	Base of Text Editors.....	11
6	Markdown Tools.....	12
6.1	Markdown Editor.....	12
6.2	Convert Markdown files in batch.....	12
7	Text Tools.....	13
7.1	Text Editor.....	13
7.2	Others.....	13
8	Html Tools	14
8.1	Html Editor	14
8.2	Find strings in html.....	15
8.3	Find elements in html.....	16
8.4	Snap html page	17
8.5	Others.....	17
9	Microsoft Documents	18
9.1	Handle PPT file.....	18
9.2	Handle Word file.....	19
9.3	Others.....	19
10	Edit Bytes.....	20
11	Text in System Clipboard.....	21
12	Text in MyBox Clipboard.....	22

1 Introduction

This is desktop application based on JavaFx to provide simple and easy functions. It's free and open sources.

1.1 Main Interface



1.2 Resources Addresses

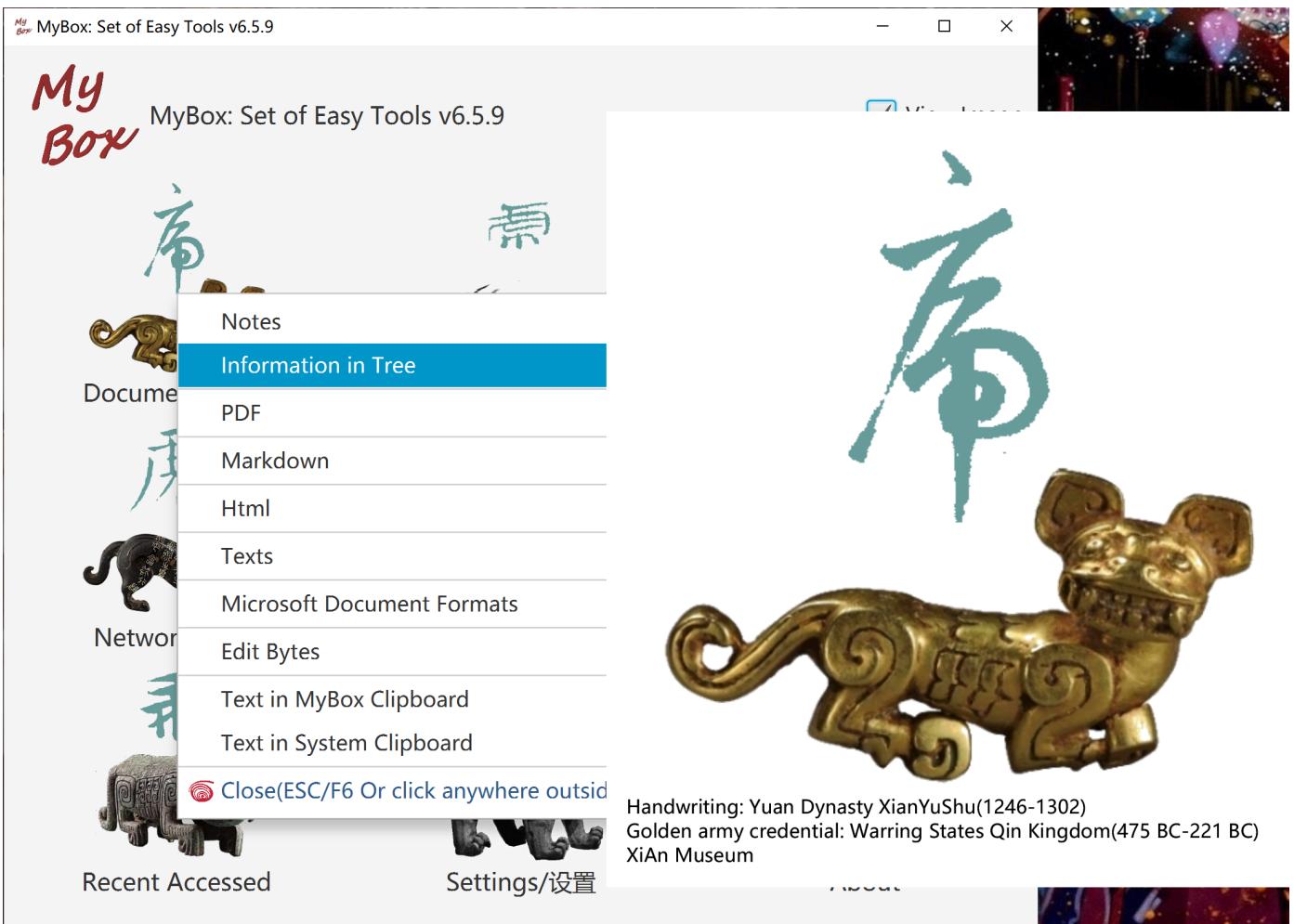
Contents	Link
Project Main Page	https://github.com/Mararsh/MyBox/
Source Codes and Compiled Packages	https://github.com/Mararsh/MyBox/releases
Submit Software Requirements and Problem Reports	https://github.com/Mararsh/MyBox/issues
Data	https://github.com/Mararsh/MyBox_data
Documents	https://github.com/Mararsh/MyBox_documents
Mirror Site	https://sourceforge.net/projects/mara-mybox/files/
Cloud Storage	https://pan.baidu.com/s/1fWMRzym_jh075OCX0D8yA#list/path=%2F

The screenshot shows a GitHub repository page for 'Mararsh / MyBox'. The top navigation bar includes links for Pull requests, Issues, Marketplace, and Explore. Below the navigation bar, the repository name 'Mararsh / MyBox' is displayed along with its status as 'Public'. The main content area shows a list of files in the 'master' branch, including 'alpha/MyBox', 'docs', 'en', 'released/MyBox', '.gitignore', 'LICENSE', and 'README.md'. To the right of the file list is an 'About' section listing various tools and technologies. At the bottom of the page, there is a 'ReadMe in English' section and a link to the releases page at <https://github.com/Mararsh/MyBox/releases>.

1.3 Documents

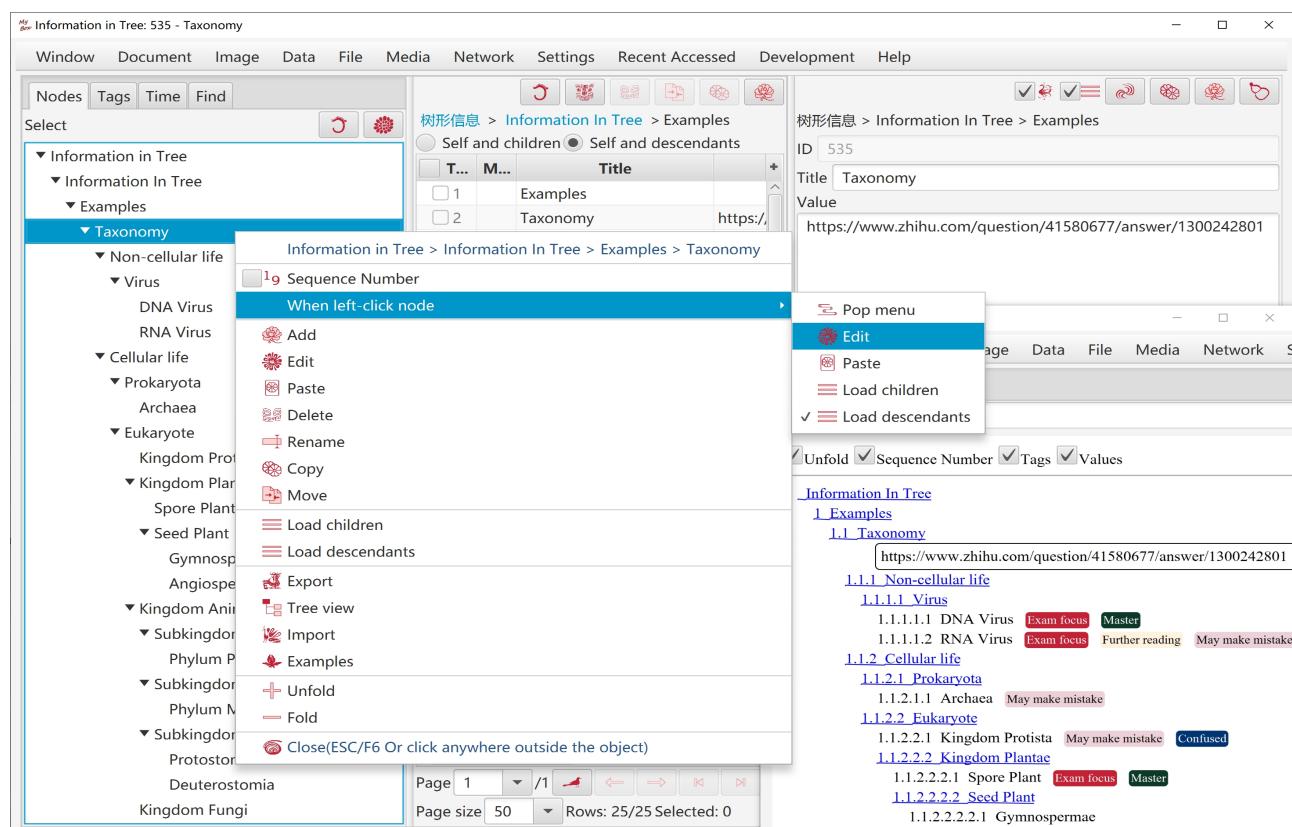
Name	Version	Time	English	Chinese
Development Logs	6.5.9	2022-8-31	html	html
Shortcuts	6.5.6	2022-6-11	html	html
Packing Steps	6.3.3	2020-9-27	html	html
Development Guide	2.1	2020-8-27	PDF	PDF
User Guide - Overview	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide – Document Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Image Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - File Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Network Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Data Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Media Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt
User Guide - Development Tools	6.5.9	2022-8-31	html PDF odt	html PDF odt

1.4 Menu of Tools



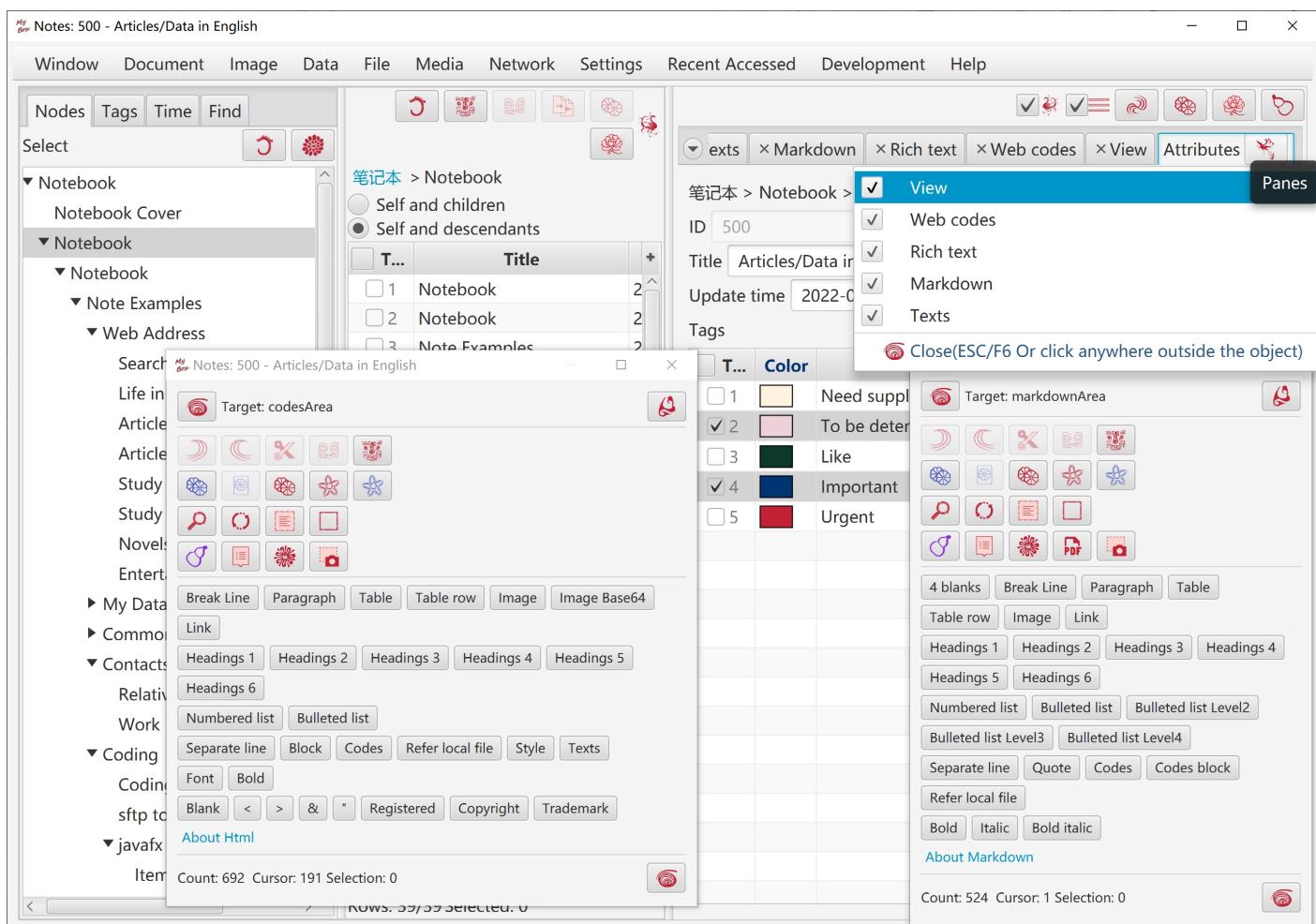
2 Information in Tree

1. Information is organized as a tree.
2. Following operations can be done against any node in the tree:
 - Add/Delete/Edit children, rename, move, copy, export, display tree view, unfold, and fold.
3. Edit node:
 - Name should not include string ">" (Blank before and after ">").
 - Each node can have multiple tags.
 - Option to show sequency numbers.
4. Export:
 - Formats: Text(for import), single html, html frameset, xml.
 - Options to export time, or tags.
 - Charset can be set.
 - Html style can be set.
5. Import. Example is provided.
6. Query:
 - Children or all descendants.
 - By tags.
 - By times.
 - By strings in title/contents.



3 Notes

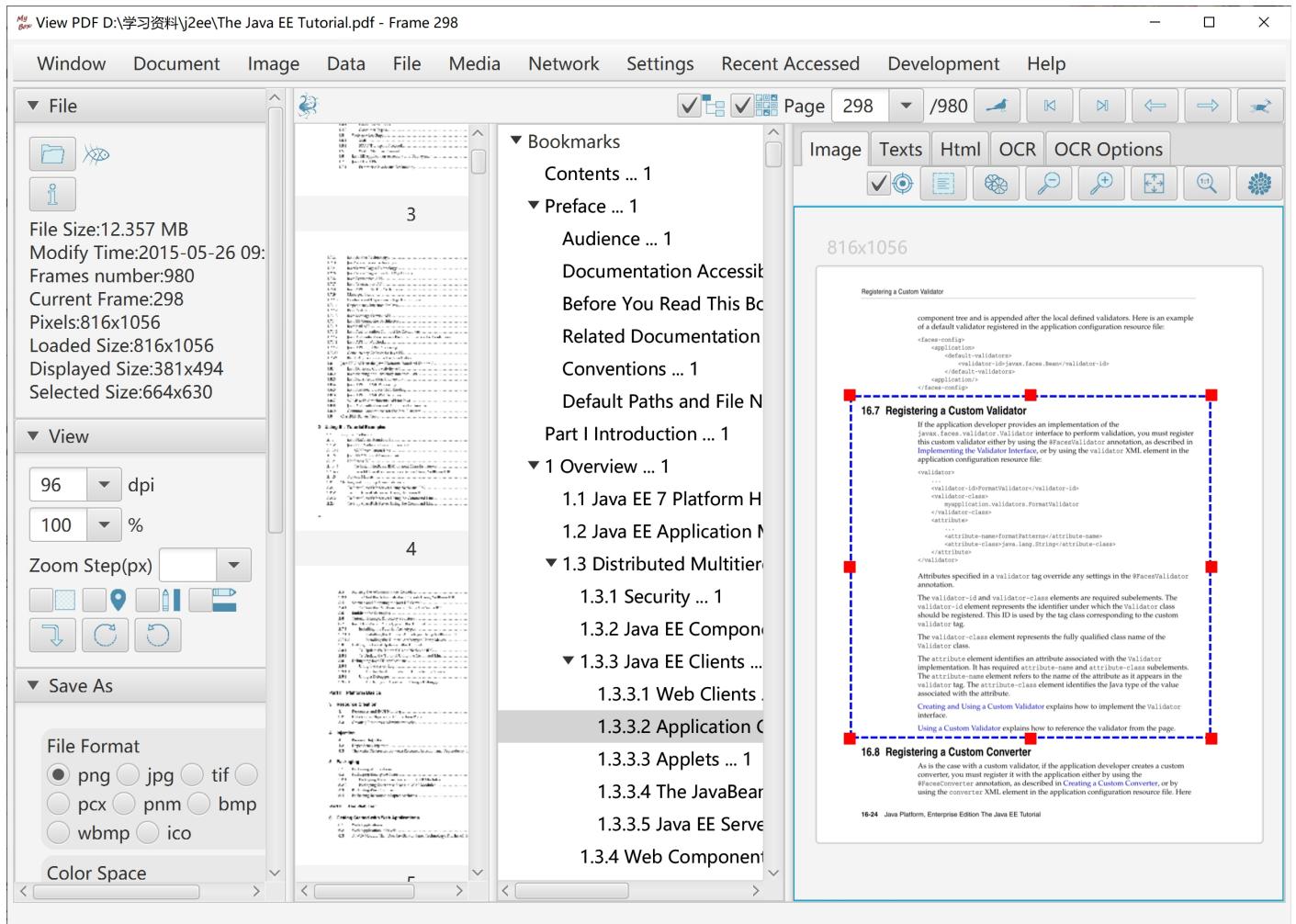
1. Note is a piece of information in html format:
 - Note can be edited in 4 modes: "html code", "Rich text", "Markdown", "Texts".
 - Each edit mode can change the html separately.
 - The mode in current tab is "current edit mode".
 - Click button "Synchronize"(F10) to apply updates in current edit mode to other modes.
 - Click button "Save"(F2) to save updates in current edit mode and synchronize automatically.
 - Note' html code should be contents of "body" and have not tags "html", "head" and "body".
 - Style can be set for note. Style is only for displaying and not saved in note's codes.
2. Notebook is information tree of notes.



4 PDF Tools

4.1 View PDF file:

1. Bookmarks and thumbnails.
2. Each page is converted as an image. DPI can be set to adjust resolution.
3. Extract text in page.
4. Convert page as a html page.
5. Recognize texts in PDF(OCR).



4.2 Convert PDF files in batch

1. Convert pages of PDF as image files. Options like format, density, color space, compression, quality, etc.
2. Convert images in selected pages of PDF files, and save as new PDF files.
3. Compress images in PDF files, and save as new PDF files. JPEG quality or threshold of black-white can be set.
4. Convert PDF files as html files in batch. Options: Whether one html for each page or one html for each PDF; The way to handle fonts/images: embed, save separately, or ignore.

4.3 Extract data in batch

1. Extract images in PDF file.
2. Extract texts in PDF file. Splitting line can be customized.
3. Recognize texts(OCR) in images of PDF files.

4.4 Others

1. Split a PDF file into multiple PDF files, by pages number, by files number, or by start-end list.
2. Merge multiple PDF files.
3. Combine multiple images as PDF file.
4. Options to write PDF like page size, image attributes, font file, margin size, header, author, etc.
5. Modify PDF file's attributes like title, author, modify time, user password, owner password, user permissions, etc.

5 Base of Text Editors

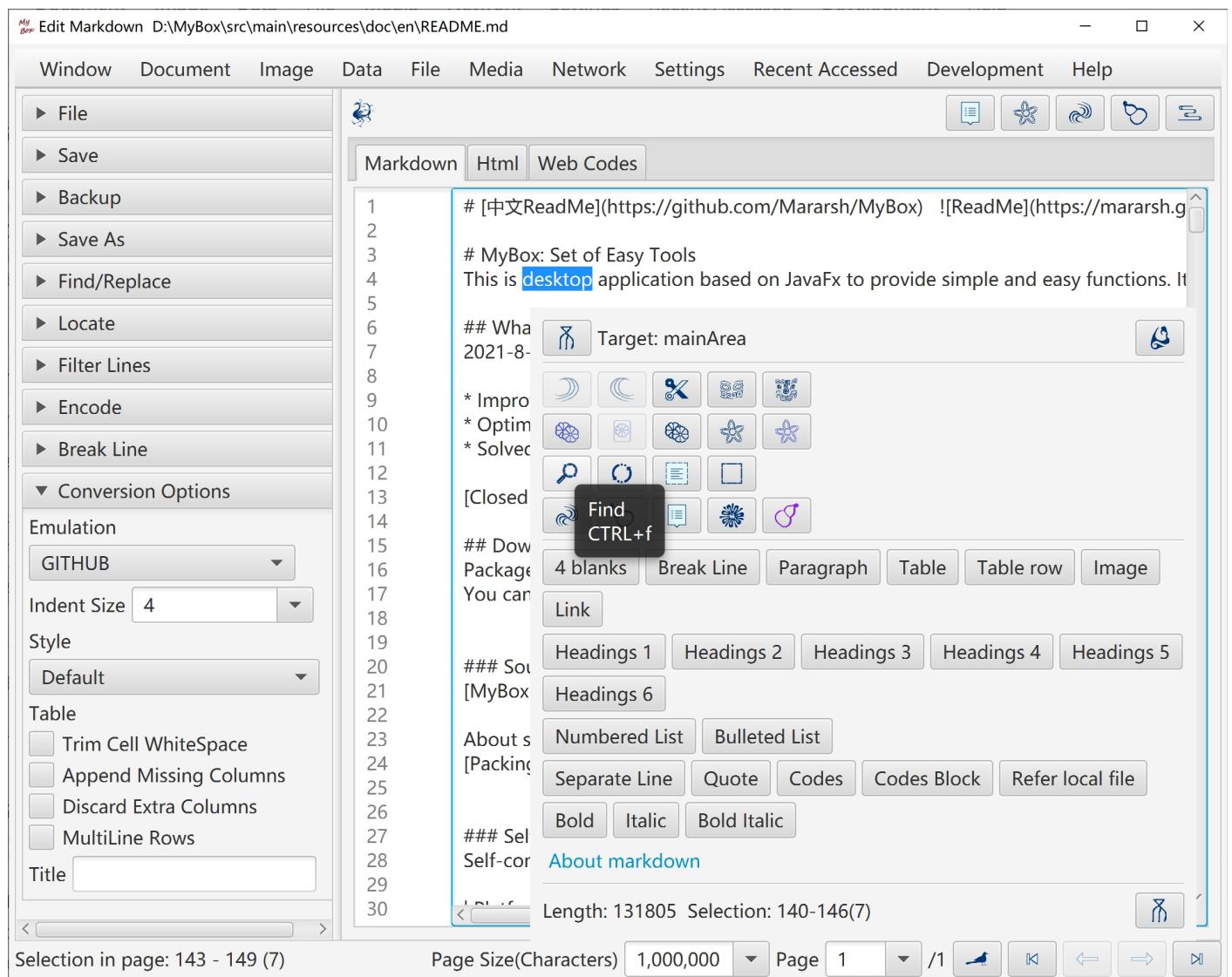
1. General functions of editing, like copy/paste/cut/delete/selectAll/undo/redo/recover. And their shortcuts.
2. Find and replace:
 - Options: Case-insensitive, Wrap around.
 - Both Find String and Replace String can be multiple lines. Their line breaks are handled as file's definition.
 - Support regular expression. Provide examples.
 - Counting.
 - Due to limitation of algorithm, to find regular expression in pages, assume maximum length of matched string is less than 1/16 of current JVM available memory.
3. Locate:
 - Select character at specified position.
 - Select line at specified position.
 - Select lines of specified range.
 - Select characters of specified range.
4. Filter lines:
 - Conditions: Include/Not include One, Include/Not Include All, Include/Not Include Regular Expression, Match/Not Match Regular Expression.
 - Cumulative filter.
 - Filtered results can be saved. Select whether include lines number.
5. Paginate. Good at viewing or editing very large file, such as logs in size of several GBs.
 - Set page size.
 - Pages navigation bar
 - Load and display first page, and scan the file in background to count characters number and lines number.
Part of functions are unavailable while counting. Interface will be refreshed automatically after counting process is complete.
 - Make sure correction of finding, replacing, and filtering of strings that are across pages.

6. Auto-save periodically.
7. Auto-backup when save.

6 Markdown Tools

6.1 Markdown Editor

1. Provide buttons to help inputting formats.
2. Display converted html and html codes synchronously.



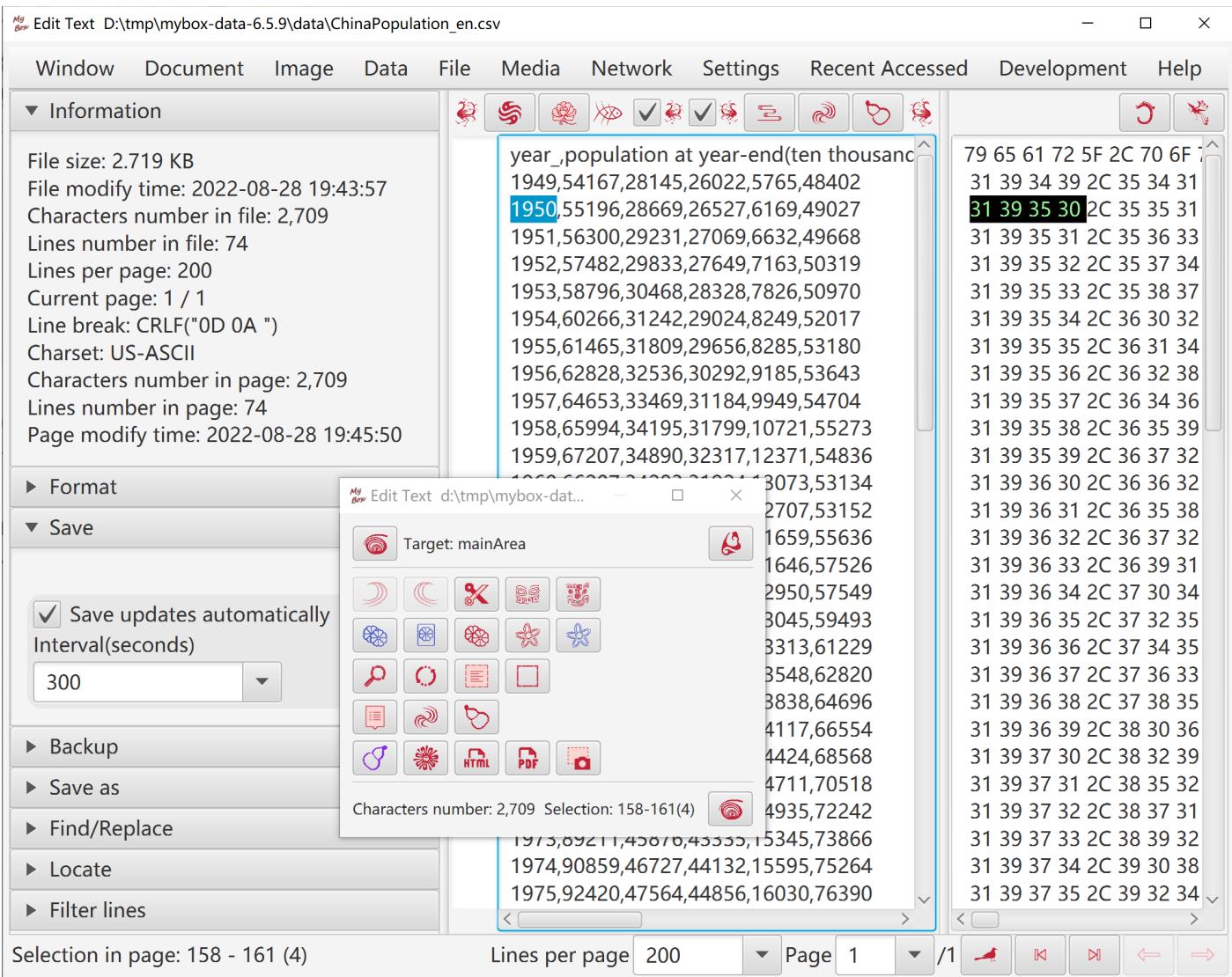
6.2 Convert Markdown files in batch

1. Convert Markdown as html in batch.
2. Convert Markdown as texts in batch.
3. Convert Markdown as PDFs in batch.

7 Text Tools

7.1 Text Editor

1. File charset can be either detected automatically or set manually. Target file charset can be selected to implement encoding conversion. BOM setting is supported.
2. Detect line break automatically. Convert line break. Show lines number.
Support LF(Unix/Linux), CR(Apple), and CRLF(Windows).
3. Hexadecimal codes according file's charset can be viewed, scrolled, and selected synchronously.



7.2 Others

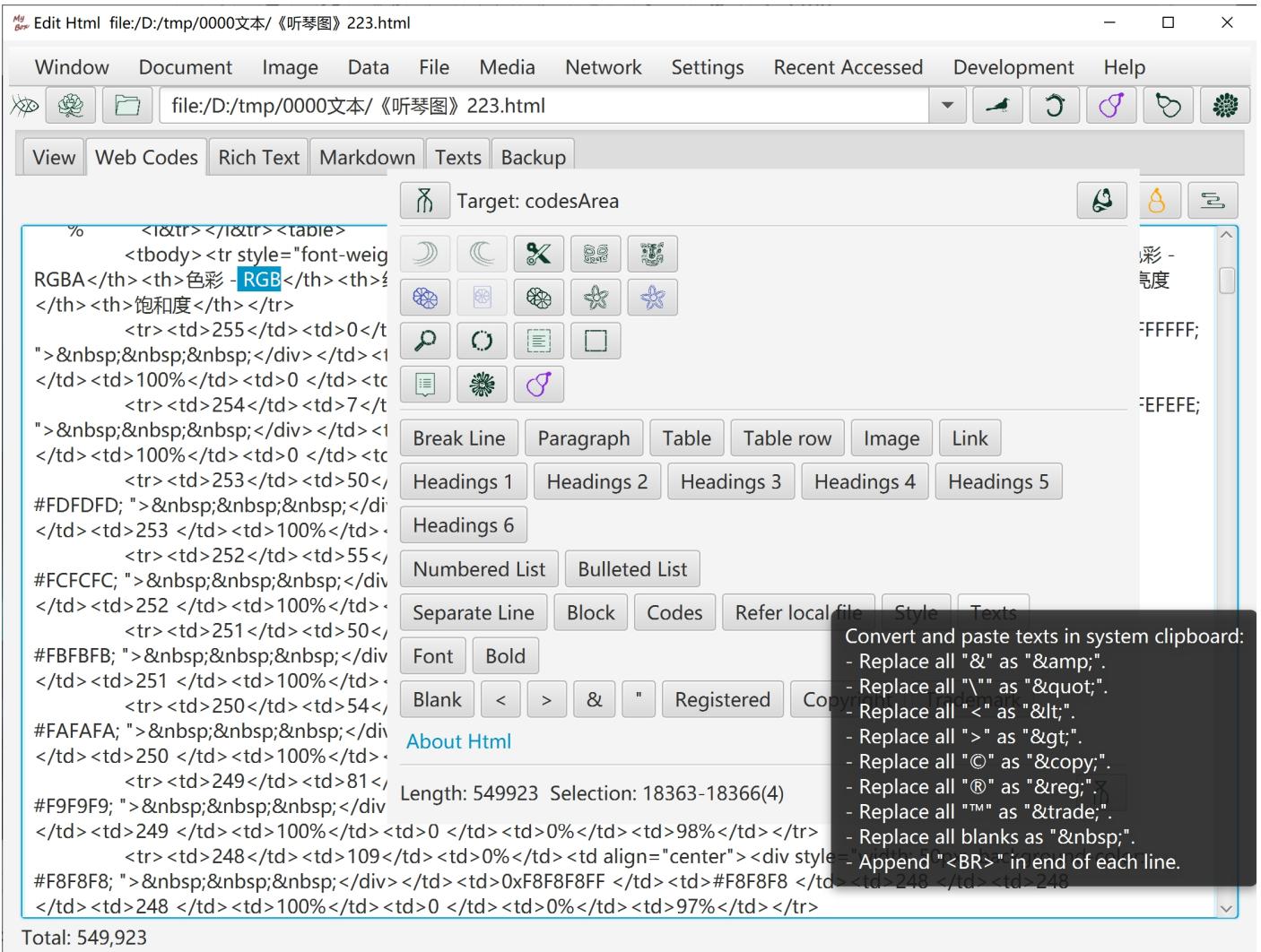
1. Convert/Split text files in batch.
2. Merge text files.
3. Convert text as htmls/PDF in batch.
4. Replace strings in text files in batch.

8 Html Tools

8.1 Html Editor

1. Html can be loaded by opening file, creating new file , or inputting address.
2. Html can be edited in 4 modes which are "html code", "Rich text", "Markdown", "Texts":
 - o Each edit mode can change the html separately.
 - o The mode in current tab is "current edit mode".

- Click button "Synchronize"(F10) to apply updates in current edit mode to other modes.
 - Click button "Save"(F2) to save updates in current edit mode and do "Synchronize" automatically.
 - Click button "Pop"(CTRL/ALT+p) to display contents in current edit mode in a new window.
3. Click button "Menu"(F12) or right click edit area to pop buttons menu.
 4. Click button "MyBox Clipboard"(CTRL/ALT+m) to pop texts clipboard for pasting.
 5. Select frame in frameset to edit.



Total: 549,923

8.2 Find strings in html

The screenshot shows the 'Find String In Web Page' tool window. The menu bar includes 'Window', 'Document', 'Image', 'Data', 'File', 'Media', 'Network', 'Settings', 'Recent Accessed', and 'Developmen'. The toolbar has icons for search, stop, and refresh. The main area is titled 'Edit Bytes' and contains a numbered list of instructions. To the right are three panels: 'Conditions' (with checkboxes for 'Regular expression' and 'Case-insensitive', and a 'String' field containing 'MyBox'), 'Display' (with color and font settings), and 'Results' (showing 'Found: 8' items, item 4 selected, and navigation arrows).

Edit Bytes

- Bytes are expressed as 2 hexadecimal characters. All blanks, line breaks, and invalid values are ignored.
- Input boxes of general ASCII characters.
- Break lines, which is only for display and has not actual effect. By bytes number or by some defined bytes.
- Select charset to decode bytes which can be viewed, scrolled, and selected synchronously.
- Paginate. When break lines by bytes number, crossing pages need not concerned.

Text in MyBox Clipboard

- Add/Delete/View texts in MyBox Clipboard.
- Add text in System is onl text input controls, MyBox Clipboard can be popped for selecting text to paste.

Text in System Clipboard

Load/Refresh/Delete text in System Clipboard:

- After button is clicked, new texts in System Clipboard are monitored and saved in MyBox Clipboard.
- Monitor interval can be set.
- Monitored texts can be accumulated in this interface. Separator can be choiced.
- This monitoring stops when one of following happens:
 - User clicks button “Stop”
 - MyBox exits.
 - “Copy to MyBox Clipboard” is not checked and this interface is closed.
- Option “Start monitoring when MyBox reboots”.

LI

8.3 Find elements in html

MyBox User Guide – Document Tools v6.5.9

Find Elements in Web Page

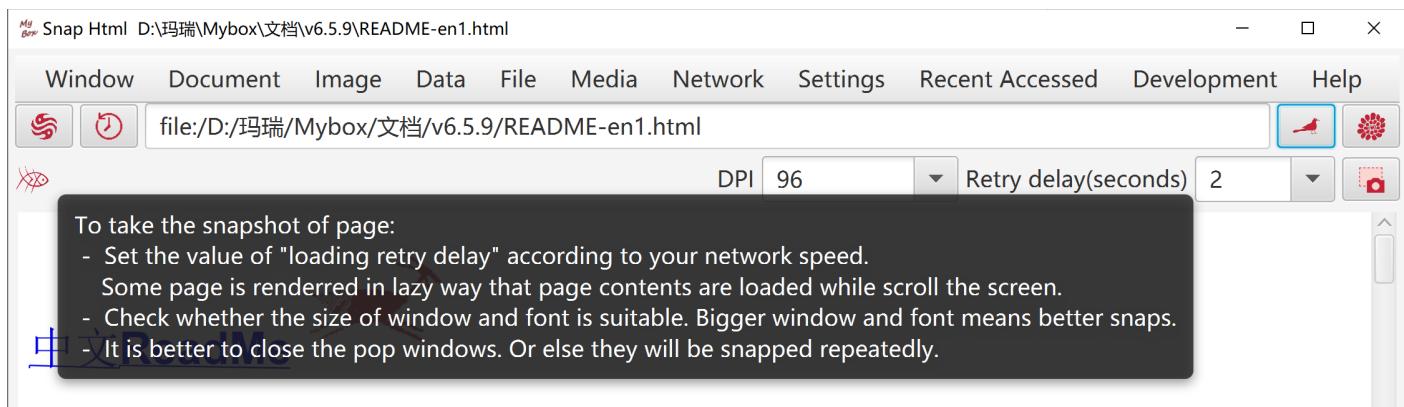
Window Document Image Data File Media Network Settings Recent Accessed Dev

Index	Tag	Name	Type	Value	Attributes
1	IMG	IMG	1		Target: elementInput p img a div li ul ol h1 h2 h3 button input label form table tr th td script style font span b hr br frame pre
2	IMG	IMG	1		
3	IMG	IMG	1		imageViewer en.jpg alt=Snap-imageViewer
4	IMG	IMG	1		src=https://mararsh.github.io/MyBox/snap-imageBrowser-en.jpg alt=Snap-imageBrowser
5	IMG	IMG	1		src=https://mararsh.github.io/MyBox/snap-imageAnanlyse-en.jpg alt=Snap-imageAnanlyse

Count: 5

Examples

8.4 Snap html page



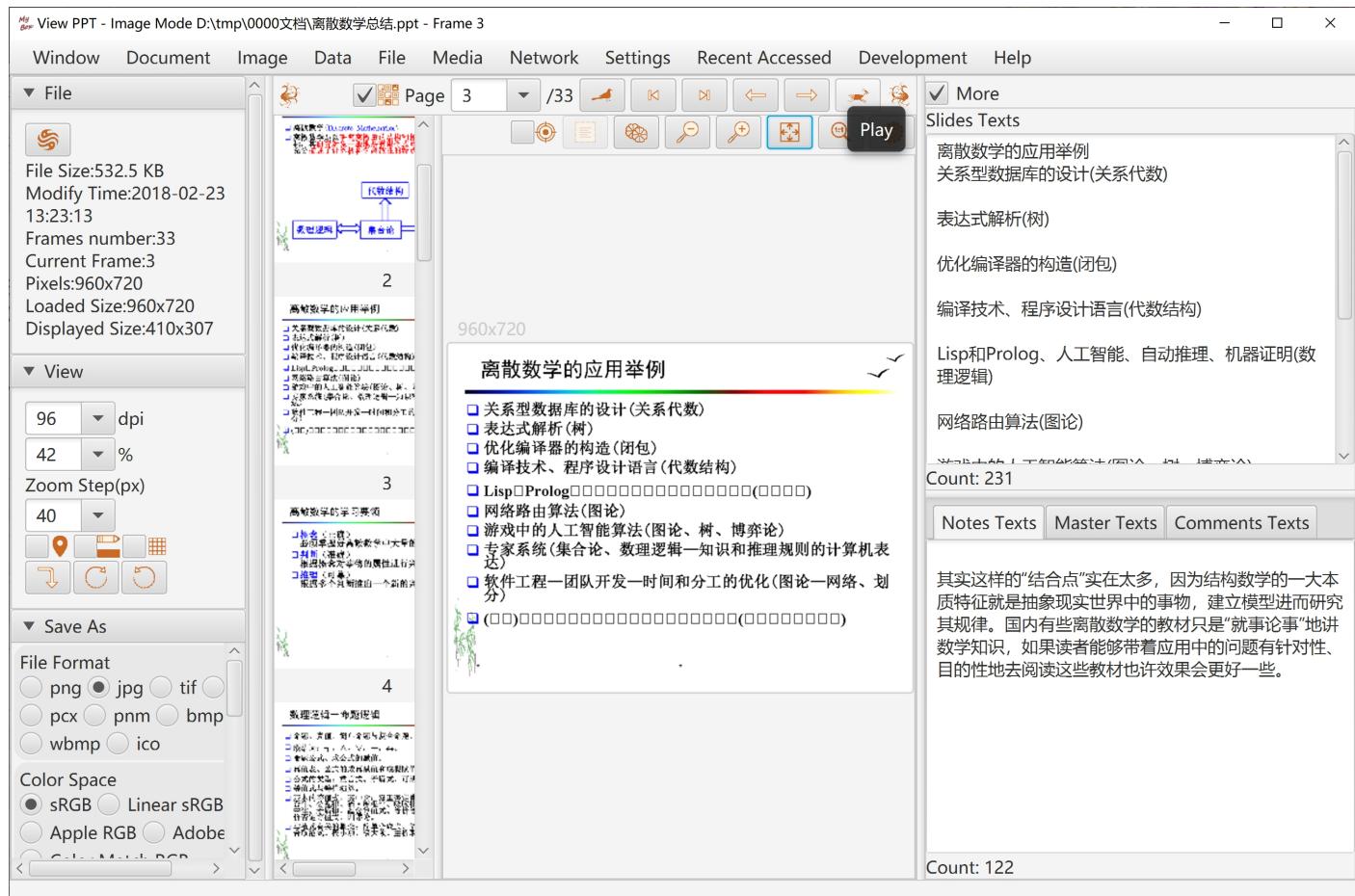
8.5 Others

1. Extract table data in html in batch.
2. Convert html as Markdown/text/PDF in batch.
3. Convert charset/style in html files in batch.
4. Merge multiple pages as one html/Markdown/texts/PDF.
5. Generate frameset file for multiple files.

9 Microsoft Documents

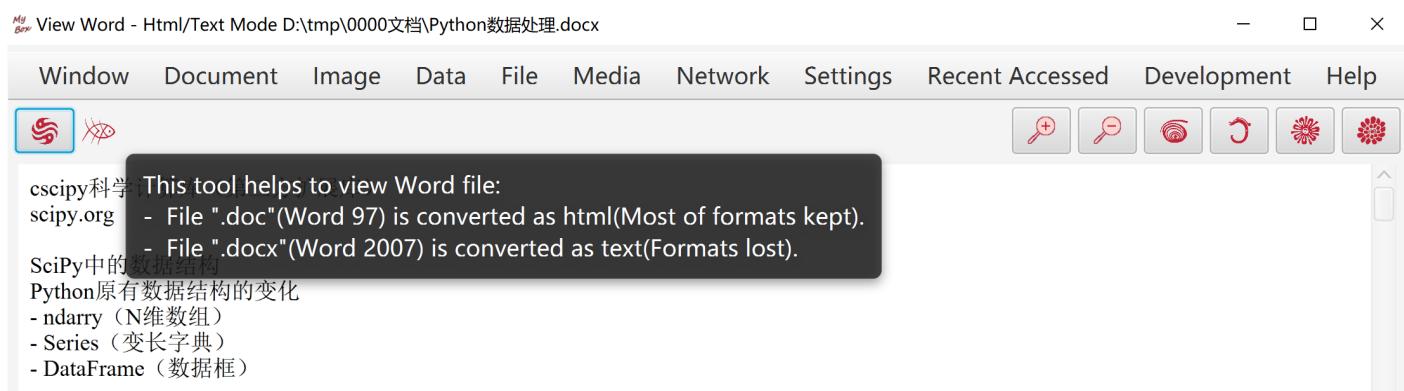
9.1 Handle PPT file

1. Formats:
 - File ".ppt"(PPT 97, OLE 2) .
 - File ".pptx"(PPT 2007, OOXML).
- If no comment, both formats are supported.
2. View PPT file. Shown page by page:
3. Page is converted as an image
4. Slide text and Note text are displayed
5. Convert PPT files as images/PDF in batch
6. Extract objects in PPT file in batch:
 - Selections: Slide text, Note text, Master text, Comments text, image, sound, OLE(Word/Excel)
 - Not support to extract sounds in ".pptx".
7. Split PPT files in batch
8. Merge ".pptx" files
9. Combine images as ".ppt" file.
10. Play PPT file



9.2 Handle Word file

1. Formats:
 - File ".doc"(Word 97, OLE 2) is converted as html. Most of formats kept.
 - File ".docx"(Word 2007, OOXML) is converted as text. Formats lost.
2. View Word file.
3. Convert Word files as html/PDF in batch.

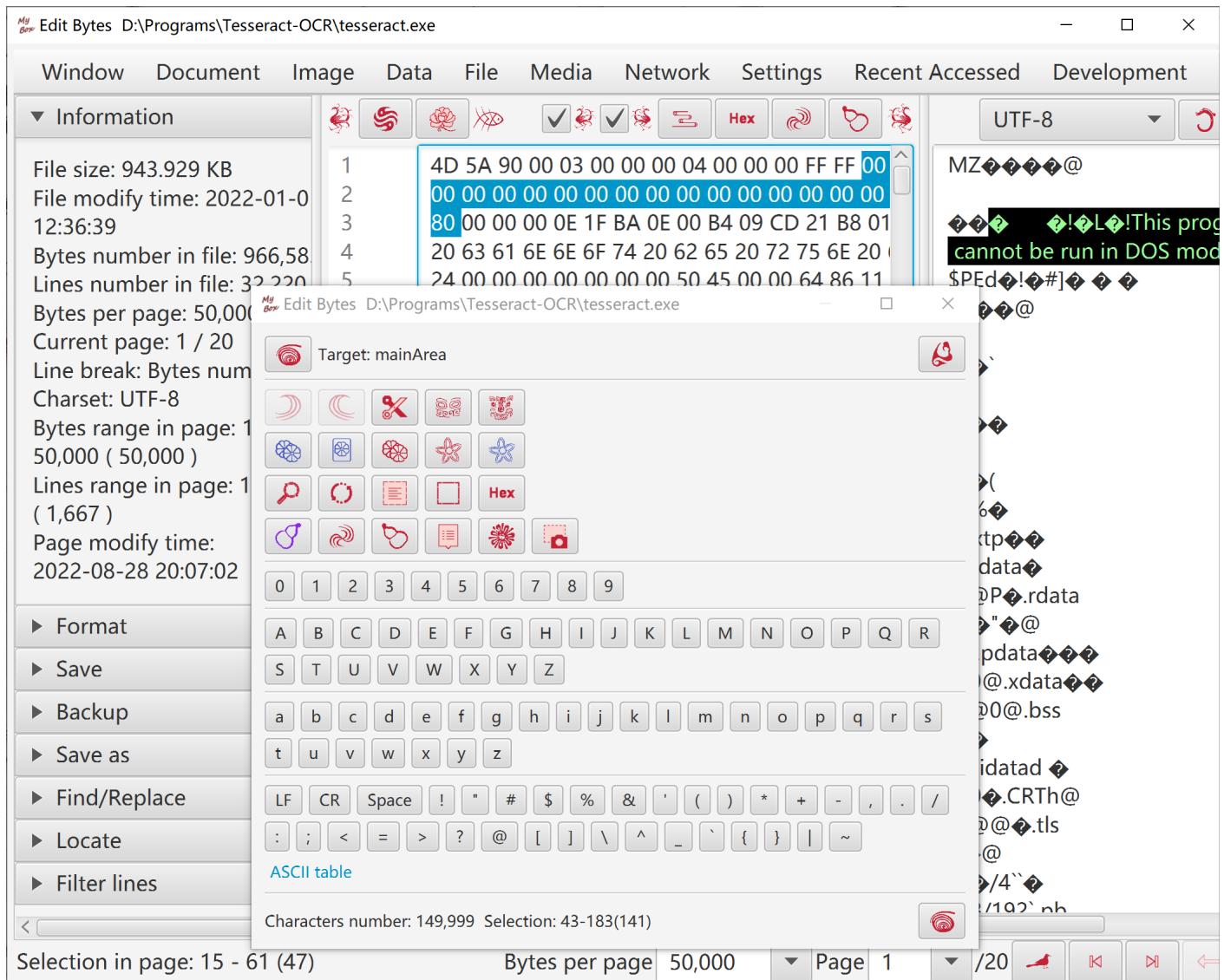


9.3 Others

Extract texts in Excel/Word/PowerPoint/Publisher/Visio files in batch.

10 Edit Bytes

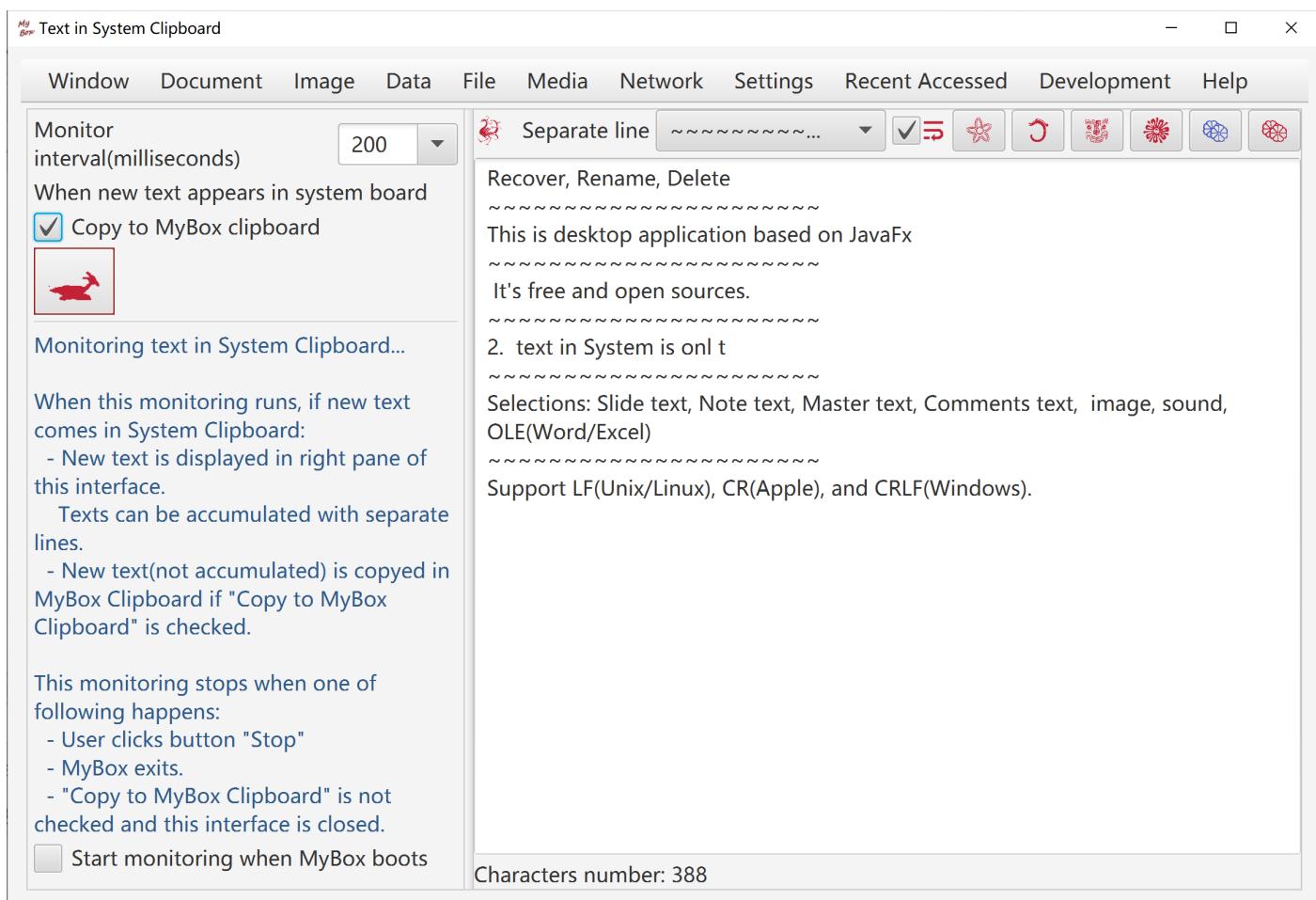
- Bytes are expressed as 2 hexadecimal characters. All blanks, line breaks, and invalid values are ignored.
- Input boxes of general ASCII characters.
- Break lines, which is only for display and has not actual effect. By bytes number or by some defined bytes.
- Select charset to decode bytes which can be viewed, scrolled, and selected synchronously.
- Paginate. When break lines by bytes number, crossing pages need not concerned.



11 Text in System Clipboard

Load/Refresh/Delete text in System Clipboard:

1. After button is clicked, new texts in System Clipboard are monitored and saved in MyBox Clipboard.
2. Monitor interval can be set.
3. Monitored texts can be accumulated in this interface. Separator can be choiced.
4. This monitoring stops when one of following happens:
 - User clicks button "Stop"
 - MyBox exits.
 - "Copy to MyBox Clipboard" is not checked and this interface is closed.
5. Option "Start monitoring when MyBox reboots".



12 Text in MyBox Clipboard

1. Add/Delete/View texts in MyBox Clipboard.
2. Add text in System is onl text input controls.
3. Copy selected texts in System clipboard.
4. MyBox Clipboard can be popped for selecting text to paste.

The screenshot shows the 'Text in MyBox Clipboard' application window. The menu bar includes 'Window', 'Document', 'Image', 'Data', 'File', 'Media', 'Network', 'Settings', 'Recent Accessed', 'Development', and 'Help'. The toolbar contains various icons for file operations. A tooltip is displayed over the 'No duplication' checkbox in the toolbar, explaining its function:

If "No duplication" is checked:
 - New arrived text is checked whether same as an existed item.
 - If existed, no new item is added. The existed item is updated as new time.
 - All existed items are not checked whether duplicated.

Table row	Text
<input type="checkbox"/> 1	Support LF(Unix/Linux), CR(Apple), a...
<input checked="" type="checkbox"/> 2	Selections: Slide text, Note text, Mas...
<input type="checkbox"/> 3	2. text in System is onl t

Below the table, there are page navigation buttons (left, right, first, last) and a 'Page size' dropdown set to '50'. The status bar at the bottom shows 'Characters number: 94' and 'Monitoring text in System Clipboard...'.

<End of Document>