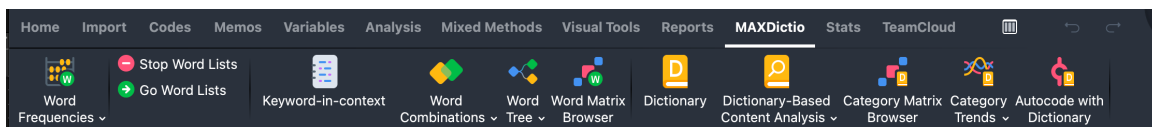


Exploring the Power of MAXDictio: Word-Based Analysis

Michael C Gizzi (mgizzi@gmail.com mgizzi@ilstu.edu)

MAXDictio is a quantitative text analysis module that offers tools for word-based analysis of text and PDF documents:

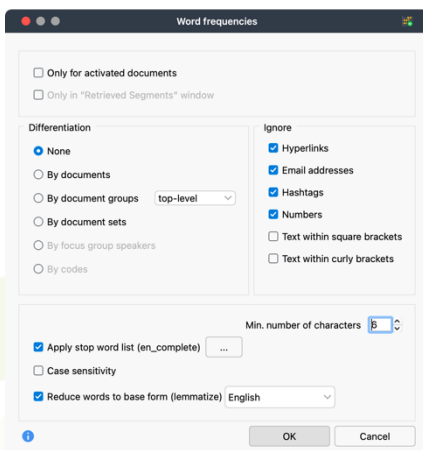
- Analysis of word frequencies and word combinations
- Analysis of Keyword-in-context (results tables and Interactive Word Tree)
- Dictionary-based content analysis with a dictionary



Explorations – Word Frequencies

Select MAXDictio > Word Frequencies to count and list words in all (or only in activated) text and PDF documents. Alternatively, you can restrict the search to the text currently displayed in the “Retrieved Segments” window.

The result table lists all words in the analyzed text by their absolute and relative frequencies. The columns “Documents” and “Documents %” indicate the number and percentage of documents in which a word occurs.



The screenshot shows the 'Word frequencies' results window. It displays a table with 20 documents (16666 analyzed words) and 1014 different words (type-token ratio = 0.0608). The table is sorted by frequency. The top 9 words are shown:

Word	Word length	Frequ...	%	Rank	Documents	Documents %
content	7	289	1.73	1	20	100.00
member	6	222	1.33	2	20	100.00
operate	7	216	1.30	3	20	100.00
revenue	7	203	1.22	4	20	100.00
netflix	7	196	1.18	5	20	100.00
growth	6	180	1.08	6	20	100.00
million	7	174	1.04	7	19	95.00
forecast	8	157	0.94	8	20	100.00
quarter	7	149	0.89	9	20	100.00

Stop and Go Lists

Stop lists are a place to collect uninteresting words, for example to or and, that you wish to exclude from your analysis. You can send words to the stop list with a click on the green icon in result tables or open MAXDictio > Stop List to customize, export, import and merge stop lists.

Differentiate by Document, Document Group or Document Set

Result tables can hold additional columns to display word frequencies for documents or document groups. Columns can display how often a word occur, the rank of a word in this document (group) or if the word occurs at all (1, 0).

Base Form (Lemmatization)

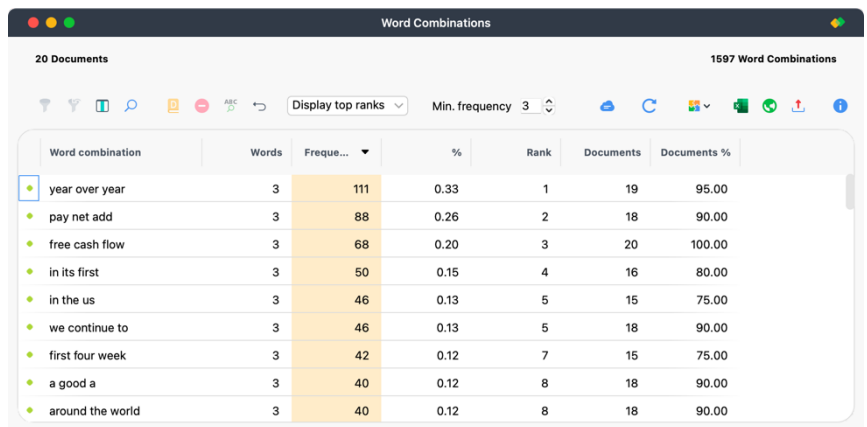
Activate the “Reduce Words to Base Form (lemmatization)” option to count words from the same stem together. For example, the frequency for “give” will include occurrences of give, gave, or given. Lemmatization is currently available for 14 languages.

Exploration: Word Combinations

Select MAXDictio > Word Combinations works similar to the Word Frequencies feature but counts and lists word combinations rather than single word

occurrences. You can search for combinations of up to 5 words.

You can decide if word combinations should occur within the same sentence or if word combinations should also be counted across paragraphs and full stops.



Word Combinations

20 Documents 1597 Word Combinations

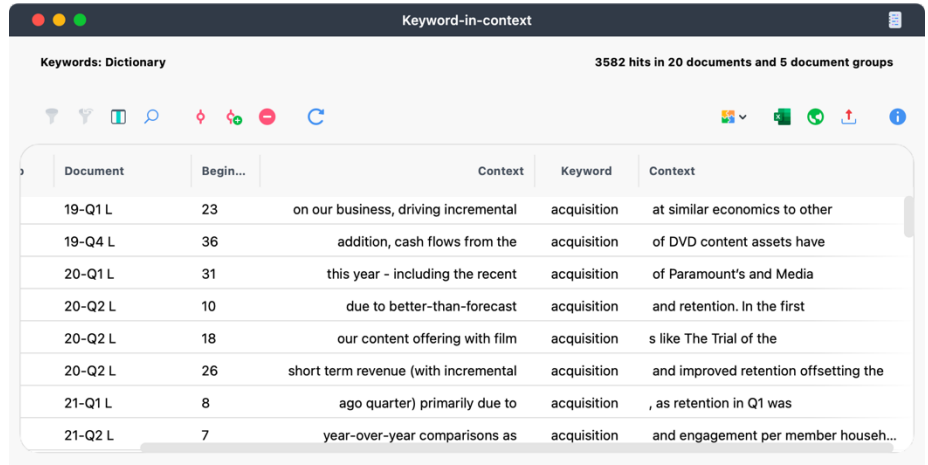
Display top ranks Min. frequency 3

Word combination	Words	Frequ...	%	Rank	Documents	Documents %
year over year	3	111	0.33	1	19	95.00
pay net add	3	88	0.26	2	18	90.00
free cash flow	3	68	0.20	3	20	100.00
in its first	3	50	0.15	4	16	80.00
in the us	3	46	0.13	5	15	75.00
we continue to	3	46	0.13	5	18	90.00
first four week	3	42	0.12	7	15	75.00
a good a	3	40	0.12	8	18	90.00
around the world	3	40	0.12	8	18	90.00

Tip: If you hover with a mouse over a row in the results table, all the lemmatized words or word combinations in this row will be displayed.

Exploration: Keyword-in-context

Select MAXDictio > Keyword-in-context to view selected keywords in a tabular view along with the words that appear before and after the word. The table also displays document information.



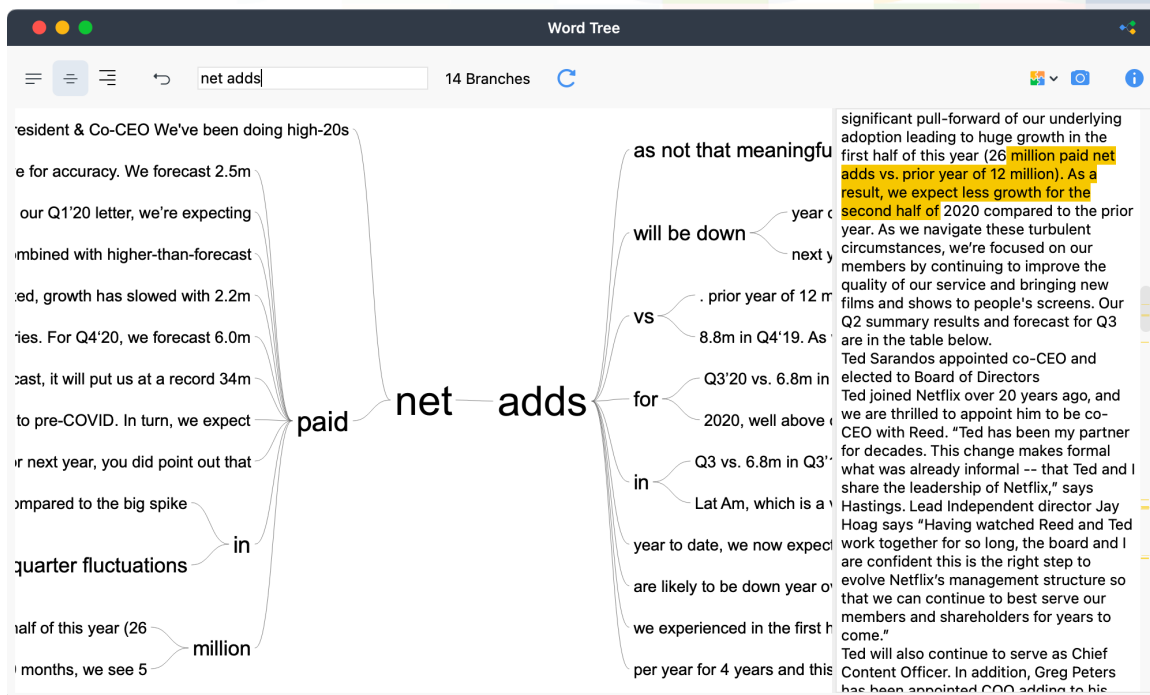
Document	Begin...	Context	Keyword	Context
19-Q1 L	23	on our business, driving incremental	acquisition	at similar economics to other
19-Q4 L	36	addition, cash flows from the	acquisition	of DVD content assets have
20-Q1 L	31	this year - including the recent	acquisition	of Paramount's and Media
20-Q2 L	10	due to better-than-forecast	acquisition	and retention. In the first
20-Q2 L	18	our content offering with film	acquisition	s like The Trial of the
20-Q2 L	26	short term revenue (with incremental	acquisition	and improved retention offsetting the
21-Q1 L	8	ago quarter) primarily due to	acquisition	, as retention in Q1 was
21-Q2 L	7	year-over-year comparisons as	acquisition	and engagement per member househ...

Tip: This table is interactively linked to the original source data. Click into a table row to highlight the search item in the original text in the Document Browser. Each column of a result table can be alphabetically sorted with a click on the column header. You can rearrange columns via drag & drop, and right click on a column header to filter for characters.

Exploration: Interactive Word Tree

Select MAXDictio > Interactive Word Tree to visually explore word and word combinations in your data). Drag all the documents, document groups or sets into the dialog window that you want to use to create a Word Tree.

The word “net adds” is the root of the tree with branches to the left and right showing the words before and after that word. The root can be placed in the middle (as shown above) or as the starting or ending point. The window to the right contains the complete analyzed text. Yellow highlighted passages are currently displayed in the tree view.

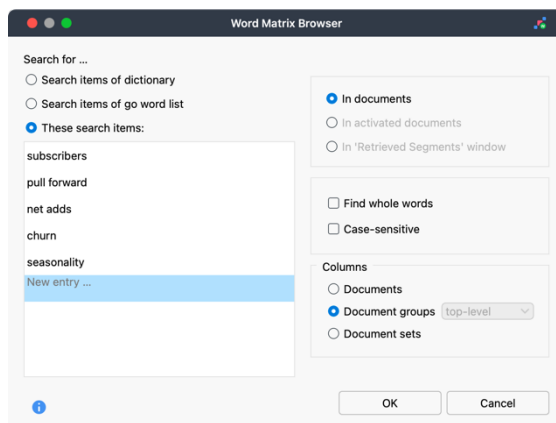


Tip: Click on another word to use it as the new root or search for word (combinations) in the search field.

Word Matrix Browser

The Word Matrix browser lets you produce a table like the code matrix browser, but instead of showing the distribution of codes in documents, document groups or document sets, it lets you see how words are used.

Word Matrix browser can use a set of search terms, words in a “go list,” or words in a dictionary.



MAXDAYS

MAXQDA VIRTUAL CONFERENCE

Word Matrix Browser

	2018	2019	2020	2021	2022
▼ competition					
● competition	13	38	17	22	22
● competitors	7	6	8	7	10
● competing		7	2	5	4
● streamers		1		1	
● streaming	12	48	45	58	72
● media	15	16	8	7	12
● linear	7	23	7	25	20
● broadcast	5	7	1		4
● competitive noise					
● bundling	7	6			

If you click on one of the cells, you will see the search results of that word or phrase and can then auto-code it if you want.

Search results

48 hits in 8 documents and 1 document groups

Preview	Document group	Document name	Se
s - Chief Product Officer Yes. I think we'll see sort of what the right solution is for consumers as we shift to this online streaming world. And I anticipate that there are models that make sense where they will bundle multiple content services	2019	19-Q4 T	sti
ar period and ahead of our forecast of 7.6m, fueled by our broad slate of original programming and the worldwide adoption of streaming video. We generated Q4-record paid net adds in each of the EMEA, LATAM and APAC regions, while UCAN paid net adds	2019	19-Q4 L	sti
21% year over year while streaming ARPU increased 9% year over year. Excluding a -\$133 million impact from foreign currency, streaming ARPU rose 12% year over year. Operating income in Q4 amounted to \$459m (vs. \$216m last year). While this was	2019	19-Q4 L	sti
on our early progress. Q4 Results and Q1 Forecast In Q4, average streaming paid memberships grew 21% year over year while streaming ARPU increased 9% year over year. Excluding a -\$133 million impact from foreign currency, streaming ARPU rose 12%	2019	19-Q4 L	sti

Dictionary-based Content Analysis

To conduct a dictionary based quantitative context analysis, your first step is to create a dictionary via MAXDictio > Dictionary. A dictionary may contain any number of categories with an unlimited number of search items. For example, a simple dictionary may look like this:

The screenshot shows the 'Dictionary' window in MAXDictio. It features a sidebar on the left with 'Dictionaries of project' and 'Global dictionaries'. The main area is divided into 'Categories' and 'Search items'. The 'Categories' list includes 'competition', 'Competitors', 'Finances', 'Global', 'Partnerships', 'Content', 'Service', 'Strategy', 'Customers', 'COVID', and 'Forecast and Guidance'. The 'Search items' table lists various terms with checkboxes for 'Whole word', 'Case sensitivity', and 'Starting letters'.

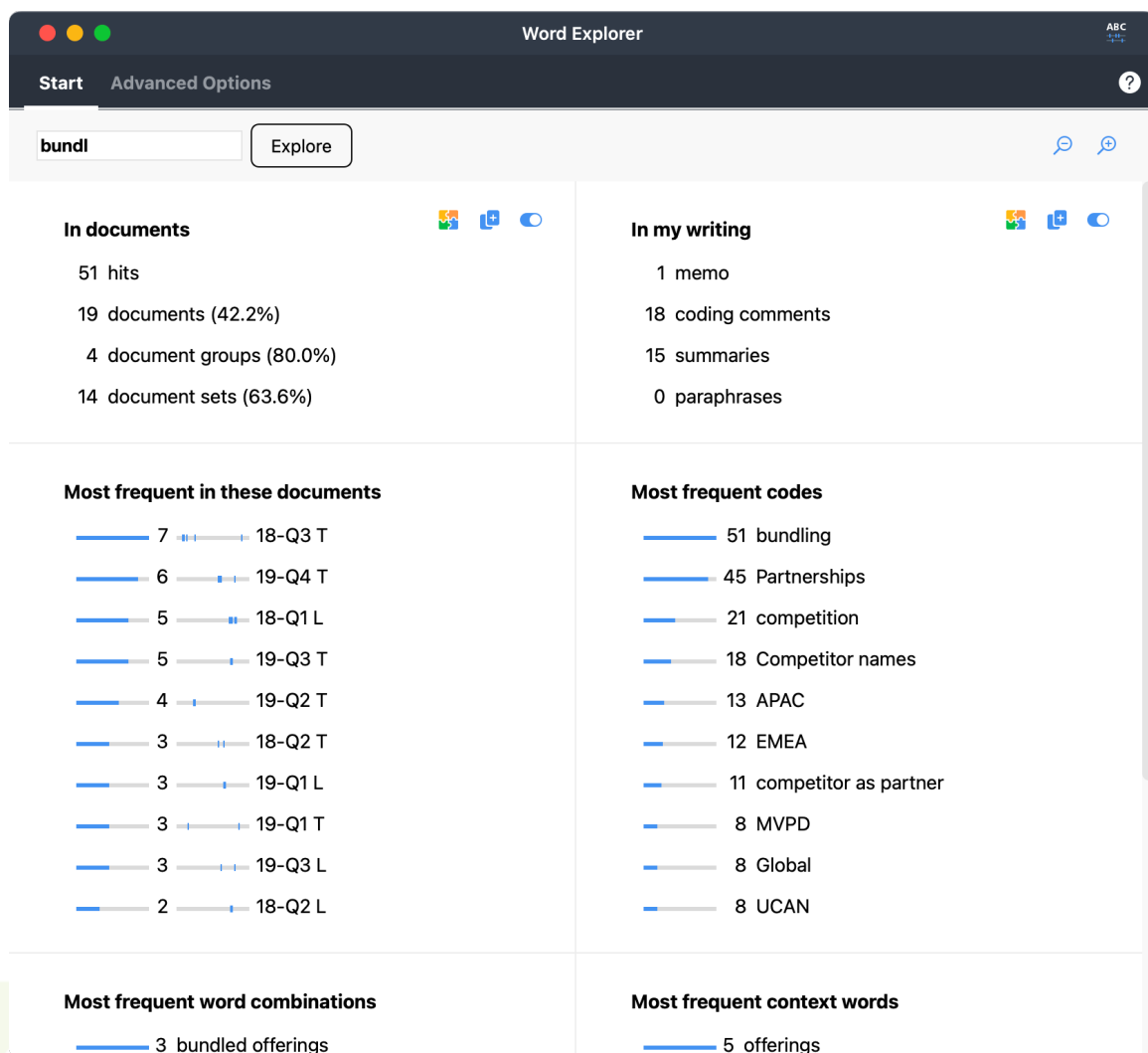
Search item	Whole word	Case sensitivity	Starting letters
cash flow	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
revenue	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
M&A	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
acquisition	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
financ	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
stock price	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pric	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
margin	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
tax	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
spend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
buyback	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
FCF	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
bond	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

This dictionary contains categories, like “Finances” or Partnerships”, each of which includes search items. A search item can also consist of multiple words or phrases. As you can see, you can specify if you want whole word, case sensitivity, or starting letters. Like stop-lists, dictionaries can be imported, exported, and merged. By default, dictionaries and stop lists are saved within a MAXQDA project, to ensure that if you share your project data with your team, everyone can use them.

Selecting Search Terms

The hard work is in creating a functional dictionary that will capture the terms you are looking to include. Here you can draw on other MAXDictio tools, such as Word Combinations, and Word Frequencies to drag words into search terms in a dictionary.

Use the Analysis > Word Explorer or MAXDictio > Word Matrix browser to first explore the terms you are thinking of including in a dictionary. Word Explorer can let you see the usage of words or phrases in documents, groups, word combinations, context words, and associated codes. It is probably the best place to “beta test” words to include as dictionary search terms.



Word Explorer

Start Advanced Options

Search: **bundl** Explore

In documents	In my writing
51 hits	1 memo
19 documents (42.2%)	18 coding comments
4 document groups (80.0%)	15 summaries
14 document sets (63.6%)	0 paraphrases

Most frequent in these documents	Most frequent codes
7 18-Q3 T	51 bundling
6 19-Q4 T	45 Partnerships
5 18-Q1 L	21 competition
5 19-Q3 T	18 Competitor names
4 19-Q2 T	13 APAC
3 18-Q2 T	12 EMEA
3 19-Q1 L	11 competitor as partner
3 19-Q1 T	8 MVPD
3 19-Q3 L	8 Global
2 18-Q2 L	8 UCAN

Most frequent word combinations	Most frequent context words
3 bundled offerings	5 offerings

MAXDAYS

MAXQDA VIRTUAL CONFERENCE

Search results

ALL: bundl 51 hits in 19 documents and 4 document groups

Autocode

Preview	Document group	Document name	Se
= Last year, we launched bundle offers with Proximus (in Belgium), SFR Altice (France) and T-Mobile (US).	2018	18-Q1 L	t
= They have proven to be very successful and we are now adding similar bundle offers with additional MVPD partners.	2018	18-Q1 L	bu
= Recently, we announced that we are bundling the Netflix service with packages from Sky which will begin later this year and with Comcast in the US, which are currently being rolled out.	2018	18-Q1 L	bu
= We believe that the lower churn in these bundles offsets the lower Netflix ASP.	2018	18-Q1 L	bu
= We remain primarily a direct-to-consumer business, but we see our bundling initiative as an attractive supplemental channel.	2018	18-Q1 L	bu
= And based on what we've seen with these new bundle models that we referred to with both Comcast and Sky announcing in the last quarter, we've seen the economics of those, we take in the retention, the acquisition characteristics to be very, very	2018	18-Q1 T	bu

Frequencies of Dictionary Categories

Select MAXDictio > Dictionary-based Content Analysis to create a frequency table similar to the ones we discussed earlier. This time MAXQDA only searches for search items in active dictionaries.

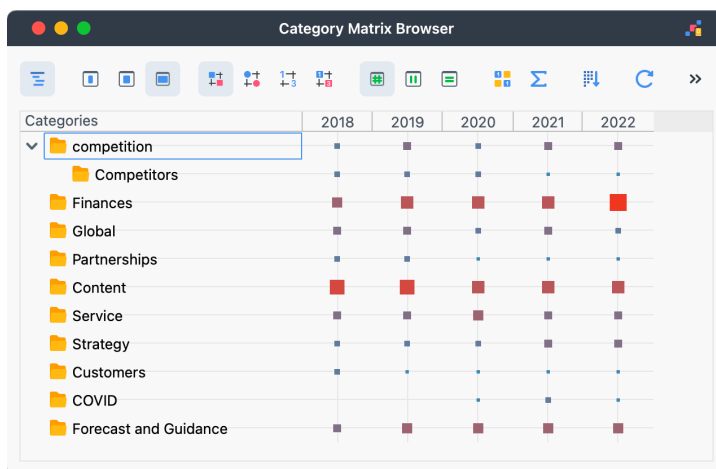
Dictionary-Based Content Analysis

All 22 Units of analysis

Document set	Words	competition	Competitors	Finances	Global	Partnerships	Content
19-Q3	10595	90	26	124	52	24	131
19-Q4	11163	51	26	100	67	12	108
20-Q1	10058	42	7	79	29	11	93
20-Q2	10917	23	12	115	18	13	97
20-Q3	10044	21	18	66	31	11	97
20-Q4	10306	51	34	95	48	2	81
21-Q1	10092	44	8	80	47	4	81
21-Q2	10271	64	22	113	24	1	94

Visualize Frequencies of Dictionary Categories

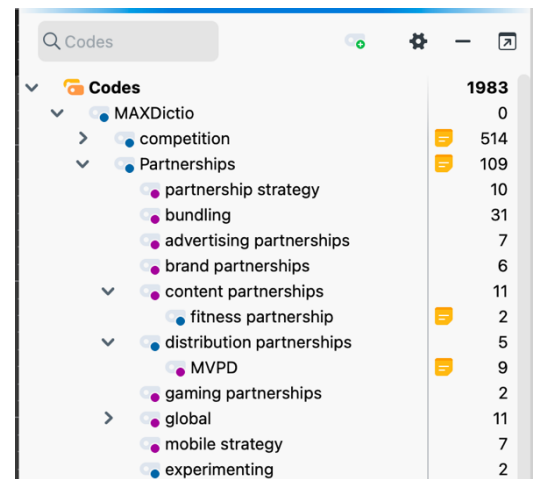
Select MAXDictio > Category Matrix Browser to visualize the same type of frequencies in a visual matrix. The image below visualizes how often search items from different categories occur in different fairy tales. You can also start this feature by clicking the matrix browser icon in the result table of a quantitative content analysis.



Like in the code matrix browser, or code relations browser, you can choose to view the results with symbols (circles or squares), numbers, or numbers in a heat map.

Auto-code documents with Dictionary Categories

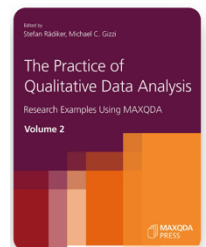
Select MAXDictio > Auto code with dictionary to automatically search for occurrences of category search items. Each search result can be coded with the name of the category, that holds this search item. All the usual auto code options of MAXQDA are available, e.g. to code just the search item, or the (adjoining) sentence(s) or paragraph(s) it occurs in.



Reviewing Auto-coded segments in a dictionary-based content analysis

While dictionary-based auto-coding is powerful, you need to be sure that what you are coding is relevant, and that it codes all of the relevant segments in your documents. There are multiple ways of doing this:

- Manual review of the coded segments in your documents. Read through the documents to see if what you coded is comprehensive or even relevant. One way to assist in this is to activate the new codes and use the Document Browser settings to only display activated coded segments.
- Use **the Smart Coding tool** to review coded segments. This can be both a check for accuracy, and a powerful tool to do secondary coding. If you treat the coded categories like themes in a thematic analysis, you can use smart coding tool to easily add sub-codes. You could also use the **Word Explorer** to double check some of categories you coded to determine if your results appear to be accurate and complete.
- A good workflow for dictionary building might be to begin with **Word Explorer** and **Word Frequencies/Combinations**, and **Word Matrix Browser** to develop a dictionary. Test it out by running the Dictionary Content Analysis and Category Matrix Browser to see if the results seem to be capturing what you are hoping to capture. Then auto-code and review the results in the document browser and smart coding tool.
- For more information, download a free copy of *The Practice of Qualitative Data Analysis With MAXQDA: Research Examples Using MAXQDA, Volume 2* from [MAXQDA-Press](https://www.maxqda.com/). Then read chapter 3, “Using MAXQDA for Dictionary-Based Content Analysis: An Example Exploring the Business Strategy of Netflix Over Time”



For More information contact. Dr. Michael Gizzi,
Certified MAXQDA Trainer

<https://www.4stagesofresearch.com> mgizzi@ilstu.edu or mgizzi@gmail.com