

## Acronym Eliminator Instructions

This script removes acronyms from a given list in VantagePoint (VP) by converting them to their full word form (if a match can be made between an acronym and its full word form in the same list). For example: ‘MBS’ converts to ‘Manchester Business School’ if both of these co-occur in the same list.

After the initial script completes the user will be given output that looks like this:

	A	B	C	D	E	F	G	H	I	J	K
1	MATCHES		MATCH LIKELIHOOD		Done			Generate Acronym Finder Code		Collect Acronym Finder Results	
2	<b>dye solar cells</b>										
3	DSC	<input type="checkbox"/>	Medium								
4	<b>dye-sensitized solar cell</b>										
5	DSSC	<input type="checkbox"/>	High								
6	dye-sensitized solar cell (DSSC	<input type="checkbox"/>	Medium								
7	<b>electron transport</b>	<input type="checkbox"/>									
8	ET	<input type="checkbox"/>	Medium								
9	ETA-solar cell	<input type="checkbox"/>	Low								
10	<b>solar cells</b>										
11	sc	<input type="checkbox"/>	High								
12	DSC	<input type="checkbox"/>	Medium								

**Figure 1: Initial Output**

The user will then be prompted to check boxes in Column B (of Figure 1) that connect full words to (true) acronym matches. Column C provides guidance (in the form of High/Medium/Low scores<sup>1</sup>) for matches that are likely to be true positives. The match likelihood scores that appear in Column C provide some assistance on which matches to select, but the user may still need additional guidance.

To obtain additional guidance (on which matches are true positives), the user has the option<sup>2</sup> of selecting “Generate Acronym Finder Code” to create a block of iMacro<sup>3</sup> code that will download a list of full word possibilities (for the acronyms in the user’s dataset) from [www.acronymfinder.com](http://www.acronymfinder.com). After this block of code is run select “Collect Acronym Finder Results” to obtain a list of which Acronym → Full Word matches (from Figure 1) are confirmed by [acronymfinder.com](http://acronymfinder.com). Results will appear like this:

<sup>1</sup> Methodology for the High/Medium/Low ranking scheme appears in Appendix A.

<sup>2</sup> Note: Clicking “Generate Acronym Finder Code” and “Collect Acronym Finder Results” are entirely optional and should only be done if the user needs additional guidance on which matches are legitimate.

<sup>3</sup> See Appendix B for an overview of iMacros.

	A	B	C	D	E	F
1	Full Word	Acronym	File	Full Word Confirmed by Acronym Finder	Total # Acronym Finder matches (for Science & Medicine)	Acronym Finder Rank
2	dye solar cells	DSC	File1	yes - partial match		27 *****
3	dye-sensitized solar cell	DSSC	File2	yes - exact match		6 *
4	dye-sensitized solar cell	dye-sensitized solar cell (DSSC)	File3	no		0 NA
5	dye-sensitized solar cell	dye-sensitized solar cells (DSSCs)	File4	no		0 NA
6	electron transport	ET	File5	no		52 NA
7	electron transport	ETA-solar cell	File6	no		0 NA
8	solar cells	DSSC	File7	yes - partial match		6 *
9	solar cells	sc	File8	yes - partial match		73 *****
10	solar cells	DSC	File9	yes - partial match		27 *****
11	solar cells	dye-sensitized solar cell (DSSC)	File10	no		0 NA
12	solar cells	dye-sensitized solar cells (DSSCs)	File11	no		0 NA

**Figure 2: Acronym Finder Results**

Columns D thru F (in Figure 2) are taken from acronymfinder.com<sup>4</sup>. Column D indicates whether a given Full Word (Column A) to Acronym (Column B) match is confirmed by acronymfinder.com. Column E indicates the number of total Full Word match possibilities (for Science & Medicine acronyms) exist for a given acronym on acronymfinder.com. Column F is a measure provided by acronymfinder.com for how common/popular/relevant a given Acronym → Full Word match is. Armed with this information the user is in a better position to identify which matches (from Figure 1) are true and, after selecting the appropriate checkboxes (from Column B of Figure 1) should then select “Done”.

The user is then provided the option to data-mine the Abstract field (of the VP file he or she is working with) for additional Acronym → Full Word match possibilities. This part of the macro proceeds by searching for all sets of open and closed parentheses in each Abstract (if an Abstract field exists). If the macro finds open & closed parentheses whose contents do not contain any spaces, and whose contents do not consist entirely of numbers, it calls this an acronym. After counting the number of letters in this acronym it identifies the group of words immediately preceding the acronym whose count equals the number of letters in the acronym. If the first letter of the first word doesn't equal the first letter of the acronym the script searches for the first word after this whose letter does equal the first letter of the acronym. If it doesn't find one it drops this potential match. After collecting all potential Abstract acronym matches the macro provides the user with an option to drop non-exact matches—i.e. matches whose full words do not have first letters that exactly match the potential acronym match (for example: “photoelectrochemical measurements” and “PECM”). If the user opts to data-mine all Abstracts for all potential acronym matches it's possible that a large amount of noise (i.e. false positives) will result, but if the user opts to data-mine Abstracts for all exact matches it's possible a few true positives will be omitted. Another consideration is that this procedure can involve a significant time investment—i.e. running this on 50 Abstracts took approximately 10 minutes on a fast machine—ergo when working with a large number of Abstracts the user is probably best served to run this code overnight. If the user opts to data-mine the Abstract field for acronym matches he or she will be presented with results that look like this:

<sup>4</sup> Note: results from this website are taken only from their dataset of Science & Medicine acronyms.

	A	B	C	D	E	F	G	H	I	J
1	MATCHES		# Abstract Co-Occurrences	Done			Generate Acronym Finder Code		Collect Acronym Finder Results	
2	atomic force microscopy									
3	AFM	<input type="checkbox"/>	2							
4	Raman spectroscopy									
5	RS	<input type="checkbox"/>	1							
6	dye sensitized solar cell									
7	DSSC	<input type="checkbox"/>	5							
8	Electrochemical Impedance Spectroscopy									
9	EIS	<input type="checkbox"/>	1							
10	cationic surfactant cetyltrimethylammonium bromide									
11	CTAB	<input type="checkbox"/>	1							
12	fluorine-doped tin oxide									
13	FTO	<input type="checkbox"/>	2							
14	Nickel oxide									
15	NiO	<input type="checkbox"/>	3							

**Figure 3: Abstract Datamining Results**

Column A of this figure provides all potential matches. Column B provides checkboxes for each match, allowing the user to identify which ones are valid. Column C provides the number of abstracts in which a given full word co-occurs (or co-appears) with its potential acronym match. The higher this number the more likely it is that a given full word is representative of all (or most) instances in which its potential acronym match is used. As with Figure 1 results, the user is provided the option to select “Generate Acronym Finder Code” to create a block of iMacro code that downloads a list of full word possibilities (for the acronyms in the user’s dataset) from [www.acronymfinder.com](http://www.acronymfinder.com). After this block of code is run can select “Collect Acronym Finder Results” to view a list of the Acronym → Full Word matches (from Figure 3) that are confirmed by [acronymfinder.com](http://acronymfinder.com). Using these results the user can then identify which matches (from Figure 3) are valid and, after selecting the appropriate checkboxes (in Column B) should then select “Done”.

By default, this script is designed to convert acronyms to their full word form. The user is given the option, however, of converting full words to their acronym form. After all Acronym → Full Word matches are made the user is also given the option to combine groups, provided two or more groups exist in the selections that were made—i.e. if a group labeled “solar cells” exists, and a group labeled “dye sensitized solar cells” exists, the user is given the option to absorb the former into the latter (or v.v.).

After all selections are made a thesaurus (**Acronym\_Thesaurus.the**) is saved to the user’s Documents folder. If this is run on the (keywords) list that was active when the user first ran **Acronym\_Eliminator.vpm** acronyms are converted to their full word form (unless the user opts to convert full words to their acronym form). The new list will consolidate a given acronym and its full word form into one list item and combine their records. In the event that duplicate Sub Items exist the user will be notified of such and is advised to resolve these using VP’s Thesaurus Editor. It is advisable to run an iterative List Cleanup after running **Acronym\_Eliminator.vpm**.

## Appendix A: Match Likelihood Methodology

Acronyms can be exactly or partially matched to full words. An example the former is DFT and Density Function Theory. An example of the latter is TDDFT and Density Function Theory. In addition, acronyms can co-occur in the same (i) Abstract or (ii) Abstract sentence. On the basis of these possibilities the following Match Likelihood scheme is used:

<b>MATCH LIKELIHOOD</b>	<b>ACRONYM EQUIVALENT IS</b>	<b>A LIST ITEM AND ACRONYM EQUIVALENT CO-OCCUR IN THE SAME ABSTRACT</b>	<b>A LIST ITEM AND ACRONYM EQUIVALENT CO-OCCUR IN THE SAME ABSTRACT SENTENCE</b>
High	Exact	yes	no
High	Exact	yes	yes
NA	Exact	no	yes
Medium	Exact	no	no
Medium	Loose	yes	no
High	Loose	yes	yes
NA	Loose	no	yes
Low	Loose	no	no

## Appendix B: iMacros

iMacros provides a means to collect data from the Internet in automated fashion. These can be installed (for Firefox) at:

<https://addons.mozilla.org/en-us/firefox/addon/imacros-for-firefox/>

After installation completes a folder will be created in the user's Documents folder called 'iMacros'. A subfolder in this folder, called 'Macros', contains a list of Internet scripts. To run a given iMacro the user can copy a block of iMacro code to a Notepad file and save it (type: All Files) to this directory with the extension ".iim". After launching Firefox a small icon will appear at the top of the user's screen that looks like this:



When the user clicks this icon an iMacros control panel will appear containing a list of all scripts saved to the user's iMacro's folder<sup>5</sup>. To run a given script the user can click its name and then click "Play".

A more thorough overview of this tool is provided at:

[http://wiki.imacros.net/Main\\_Page](http://wiki.imacros.net/Main_Page)

---

<sup>5</sup> Libraries\Documents\iMacros\Macros