

Management Information Systems (MIS) – 37C00100 – Spring 2022

Assignment 7 (10 points) – Research profiling with Scopus and text-mining the exported results with Leximancer

Summary of assignment: In this assignment you will **first search research from the Scopus database**, and after exporting and cleansing the data you will learn how to use an **automated content analyzer** tool called Leximancer to peruse the results. You can use these skills when conducting your thesis or similar research-based assignments. Moreover, Leximancer can be used to analyze many kinds of texts, and the technique you learn here (using excel/csv data) can be applied e.g. when analyzing open-text answers from surveys, or transcribed interview texts. The tool can analyze various types of text data, such as *doc*, *html*, *txt*, *pdf* and *xml* (from business data to social media). Leximancer has been applied in hundreds of research articles, see https://scholar.google.fi/scholar?hl=en&as_sdt=0,5&q=leximancer&scisbd=1.

Aalto BIZ has a campus license for Leximancer that can be used either in all Otaniemi U-wing computer classes (U-256, U-257, U-344, U-351),

OR from your own Mac or Win computer via Aalto’s **Virtual Desktop Infrastructure** at <https://vdi.aalto.fi> (*instructions follow later on how to get Leximancer into use either in computer class or via VDI*).

Detailed instructions:

Watch first the 20-minute introductory video to Leximancer: <https://www.youtube.com/watch?v=qnijLduQOGY>

Then read the brief overview of the tool and to what purposes it is used, and skim preliminarily through the step-by-step instructions how to start with the tool: <https://www.leximancer.com/getting-started-with-leximancer>

Note that the video shows the previous user interface, as the software was recently updated. There are only slight differences in UI, and many improvements (e.g., the theme colors added to concept list’s bars).

If you are doing this assignment via Aalto’s Virtual Desktop Infrastructure, set it up at this phase. *If you use OneDrive at Aalto for saving the data file, you can connect to VDI also after the literature search phase.* From your own Mac or Win computer go to <https://vdi.aalto.fi> (does not work in old Edge browsers!).

- Select Method 2: “VMWare Horizon HTML access”, log in with *Aalto ID* (and authenticate with *MS Authenticator*) and select “**Windows 10**”, and wait a minute or two while the desktop prepares itself. If your screen gets locked by screen saver, click ctrl-alt-del selector icon . More info at: <https://www.aalto.fi/en/services/vdiaaltofi-how-to-use-aalto-virtual-desktop-infrastructure>.

For gathering the text data needed for the assignment, conduct first a simple research profiling exploration at our library’s databases on scientific research articles with a **given topic below** (or alternatively with any MIS lecture, book chapter or trend report topic). Use a browser in VDI if you started using it above.

PHASE 1: See csv tip on the last page to change the **Regional settings format to English (UK) at this phase!!**

Your search should result in an adequate number of hits (100-2000) for the assignment to be meaningful. It is recommended to stay under the data export limit of 2000 (otherwise you need to export several files).

- 1) Go to our library’s Business guide at <http://libguides.aalto.fi/business> and click **Articles**
- 2) Scroll all the way down, select **Scopus** database from the list under **Reference databases**
 - In case you are *off-campus*, log in with your Aalto ID and password when prompted, in case you are not using a VPN connection.

- 3) Conduct a basic search (i.e., from *Article title, Abstract, Keywords*) with the following search phrase: **("information systems success") OR ("IS success")**
 ADD a date range field and limit the starting year from 2006 – the results will later be categorized into 3 time periods (2006-11, 2012-16, 2017-22)
 Or you may use another relevant search phrase related to MIS topic of your choice. Remember to use quotation marks in all "multi-word terms" such as examples above.

The screenshot shows the Scopus search interface. At the top, it says "Start exploring" and "Discover the most reliable, relevant, up-to-date research. All in one place." Below this are tabs for "Documents", "Authors", and "Affiliations". The search bar contains the query: "information systems success" or "IS success". Below the search bar, there are filters for "Published from" (2006) and "To" (Present). At the bottom right, there is a "Search Q" button.

- 4) From the left Refine results pane, scroll down to **Keywords**, and click **View more** and **View all**.

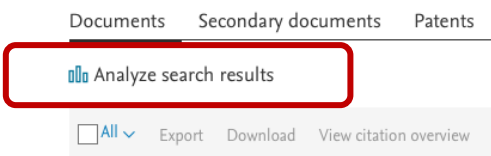
The screenshot shows the "Refine results" pane. It has tabs for "Limit to" and "Exclude". Below are various filters: Open Access, Year, Author name, Subject area, Document type, Publication stage, Source title, and Keyword. The Keyword section is expanded, showing a list of keywords with their respective counts: Information Systems (589), IS Success Model (352), Information Quality (265), IS Success (261), and Information Use (240). A "View more" button is highlighted at the bottom of the list.

You should get the overall view on the following page. **Read through the keywords** – this presents a simple frequency-based list and contains synonyms such as 'information systems success model' and 'IS success model' (you would need to combine synonyms in case you would make a top-10 table of keywords, but it is not needed in this assignment). By glancing the keywords that Scopus provides, you can get some grasp of a (previously unknown) research topic. For instance, you may find that surveys and questionnaires, SEM (Structural Equation Modeling) and PLS (Partial Least Squares) methods are quite often used in IS success research (optional: check the IS theory wiki for a synopsis of IS Success model: https://is.theorizeit.org/wiki/Delone_and_McLean_IS_success_model)

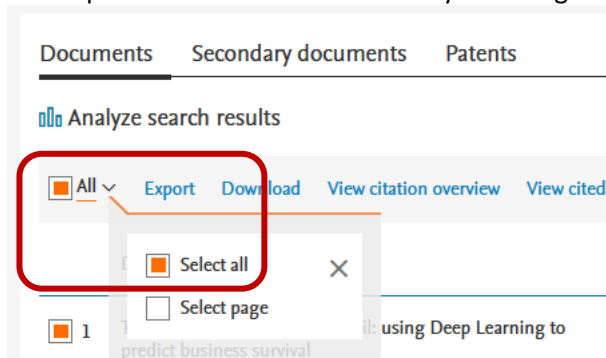
Close then the keywords view from the **X** on top right, and continue with the next phase.



- 5) Next, take a quick glance of the graphical results overview of other data fields than the keywords, by clicking the “**Analyze search results**” icon above the search results list: you will see the annual publication trend chart, documents by source annually, top-10 authors, top-10 affiliations and top-10 countries, documents by subject area, etc. You can click any of the smaller graphs to get them under focus. The graphs are interactive, and you will see more results by clicking numbers or bars.



- 6) After the quick research profiling phase, go back to the search results by clicking < **Back to results**, and export the bibliometric results by choosing **All/Select all**, and then **Export**.



Select first the **CSV file** format, and the tick the following fields to export (see also pic on next page) **Year, citation count, DOI and Abstract**, and click EXPORT. Save the file to Downloads (or OneDrive). DOI is Digital Object Identifier is used to uniquely identify an article or document. In case you find interesting results in the next phase, you are able to locate the article easily just by googling the DOI number (note that Leximancer changes / to _ and it should be corrected first, e.g. to 10.1057/ejis.2008.15 or a newer DOI type 10.1109/ICAIE53562.2021.00141).

Export document settings ✕

You have chosen to export 1977 documents

Select your method of export

MENDELEY ExLibris RefWorks RIS Format EndNote, Reference Manager CSV Excel BibTeX Plain Text ASCII in HTML

What information do you want to export?

Citation information Bibliographical information Abstract & keywords Funding details Other information

Author(s) Author(s) ID Document title Year EID Source title volume, issue, pages Citation count Source & document type Publication Stage DOI Open Access

Affiliations Serial identifiers (e.g. ISSN) PubMed ID Publisher Editor(s) Language of original document Correspondence address Abbreviated source title

Abstract Author keywords Index keywords

Number Acronym Sponsor Funding text Tradenames & manufacturers Accession numbers & chemicals Conference information Include references

Cancel Export

- 7) Open the csv file with Excel. **Select all** (ctrl-A or cmd-A) and select **Data > Sort**, sort your data in alphabetical order using **the Abstract column**. There are (probably) **records with '[No abstract available]' in the Abstract column - delete those rows altogether from the CSV file first**

	A	B	C
1	Year	Link	Abstract
2	2019	https://w	[No abstract available]
3	2019	https://w	[No abstract available]
4	2019	https://w	[No abstract available]
5	2019	https://w	[No abstract available]
6	2018	https://w	[No abstract available]
7	2018	https://w	[No abstract available]
8	2018	https://w	[No abstract available]
9	2021	https://w	"Smart Contracts" are tools based on distributed lec
10	2019	https://w	5G is a revolutionary technology in mobile telecomn

Delete also the extra Link column as it would cause anomalies to the csv data in the next phase.

Next, select/activate the Abstract column. Select **Data: Text to Columns**, select **Delimited** and Deselect Tab, tick **Other**, and copy-paste © to the Other field as shown below, click Finish. **This split procedure separates the unnecessary copyright sentences from the end of the abstract cells (that also confound the text-mining results, if not removed).**

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how

Delimiters

Tab Semicolon Comma Space Other: ©

Treat consecutive delimiters as one

Text qualifier: " ▾

NOTE: some copyright sentences have two instances of © at the end of the abstract, such as "© Copyright © 2021 Li and Wang.", and these cause extra columns (i.e. make two new split columns, see below)

Abstract	new split column-1	new split-column-2
Government	Published under licence by IOP Publishing Ltd.	
Cloud fina	Copyright	2021 Li and Wang.

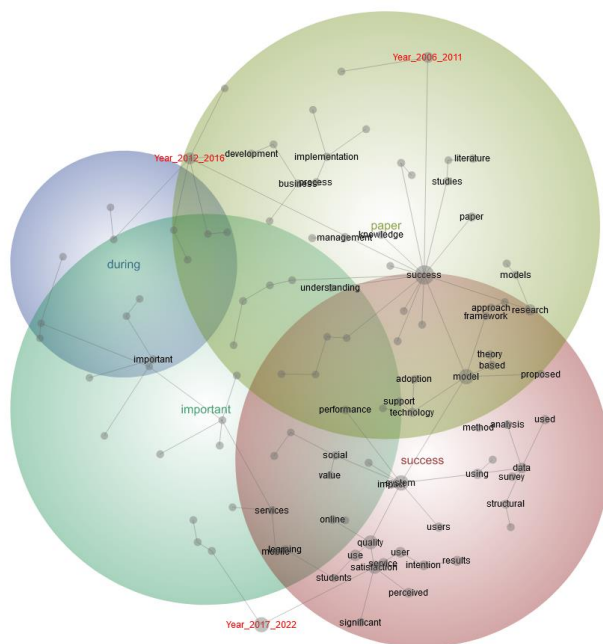
THUS, remove all extra columns the split has made AFTER the Abstract column!

Next, select all (ctrl-A) again **and sort the data again by year from smallest to largest**. Then, dividing the years into 3 time periods, **edit the year column cells as follows**. That is, copy-paste the bolded value to all cells belonging to that particular year range (select all those cells and paste the new value, e.g. 2006-11. Repeat still 2 times):

years 2006, 2007, 2008, 2009, 2010 and 2011
 years 2012, 2013, 2014, 2015 and 2016
 years 2017, 2018, 2019, 2020, 2021 and 2022

> **2006-11**
 > **2012-16**
 > **2017-22**

This edit will allow you to add these **three time periods** as **TAGs** to the Leximancer theme map later on (see example pic below), to study the text concepts also temporally. Note that in case the text data used would be survey or interview data, the TAG could be a respondent category, or interviewee name/class, etc.

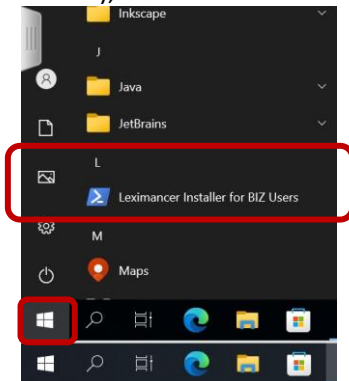


8) **Save the data file** (keep original CSV format!): it should look similar than below after all edit steps:

	A	B	C	D
1	Year	Cited by	DOI	Abstract
2	2006-11	2	10.2523/9	A hydraulic fracturing technology is put forward in clay-carbonate reservoirs w
3	2006-11	727	10.1287/is	Although DeLone, McLean, and others insist that system usage is a key variable
4	2006-11	172	10.1080/0	Although online communities can significantly facilitate collaboration among I
5	2006-11	21	10.1142/Si	Although several previous studies have investigated the success factors and in
6	2006-11	8		Appraising the success of information systems has long been a difficulty for m
7	2006-11	2	10.4324/9	At the heart of the frenetic activity of the record industry and of all the conflict
8	2006-11	8	10.4018/9	Business organizations are still struggling to improve the quality of informatio

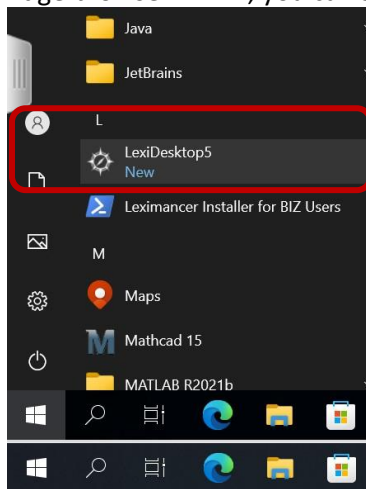
PHASE 2:

- 9) After the data file is cleansed and prepared, it is finally time to work with Leximancer. Either from your VDI screen, or class computer, open Windows Start menu (window icon on VDI's bottom left corner), and select **Leximancer Installer for BIZ Users** under L.

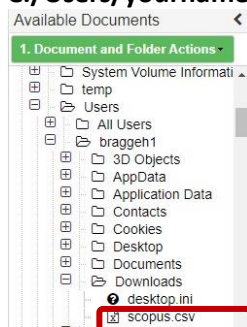


WAIT a minute or two while the installation proceeds (there are no signs of progress).

- 10) Open the same menu again, and you should see that LexiDesktop5 has appeared to the start menu under L. You need to do this each time you start working *anew* with Leximancer (due to BIZ license control issues). **Click LexiDesktop5** and **Leximancer should open after some initialization time into the browser** (you probably had a browser open already when using Scopus). If Leximancer opens Edge browser in VDI, you can select “No, sign me out”, and close the extra tab to see Leximancer)



- 11) When Leximancer is open in the browser, activate “Leximancer Projects” row from the left panel and right-click on top of it: select “+ Create Project” from the dropdown menu. Give a name and description to your project and click OK. Refer also to Leximancer’s getting started instructions again, if needed: <https://www.leximancer.com/getting-started-with-leximancer>
- 12) Click **Select Documents** icon in the main window. From the left panel, select the directory where you exported the scopus.csv (if you used VDI when saving it, it might be under the path **C:/Users/yourname/Downloads**, as shown below):



- 13) **Select the file** from the left panel **and right-click on it and select “Add this File or Folder to the Project”**. Click OK after the file name shows in the center window.
- 14) Click to open **“Text Processing Settings”** and un-tick there **“Identify Name-Like Concepts”**. Read the other settings. Click OK.
- 15) Open next the **“Concept seeds settings”**, and un-tick **“Mark Tags as Attributes”** (this will allow you to use the three time periods as TAGs later on). Read the other settings. Click OK.
- 16) Press **Generate Concept Seeds** and *wait a minute until the text-processing finishes and the box turns Green*.
- 17) This is not a necessary phase, but for your information if there would appear odd concepts/partial words (this happens often with PDF form data): Click to open **“Concept Seeds”**, and you may tick non-meaningful concepts and select **Remove**. You may also **merge singular and plural concept terms** here, e.g. user & users, by ticking them both and then selecting Merge. Note: if you repeat the merge procedure, **untick the previously merged term, which stays active** (or use unmerge)! Click OK when you are ready.
- 18) Click the third blue box: **Generate Thesaurus**, and wait until the box turns **green**. Open the **“Concept Coding Settings”**. Search for the three ‘Year: 20xx_yy’ period concepts under Available Names list (scroll to bottom), tick the three boxes, and add them with the arrow <- to the mapping concepts. Click OK.
- 19) Click the fourth blue box: **Generate Concept map**, and WAIT until all four process boxes are **green**.
- 20) After that, click **Concept Map** from the Results area. Note, if you use a small laptop screen, the Results pane may appear hidden below the model building boxes and not on the right of them.
- 21) Explore with the three sliders to see how the map changes, 1) change the **“Concepts 33%”** first to **0%**, 2) the **“Theme Size 60%”** to **50%**, and finally, 3) the **“Rotation 0”** to any number where the **figure fits best to page**, and which minimizes the words appearing on top of each other.
- 22) Then start sliding the concept-% gradually upward, to see the concepts appear in the order of importance to the map, which may help you to better interpret the data or build a story behind it.
- 23) Continue working with the tool in order to answer the following 5 tasks/questions for your report.

Report the following 5 points in your assignment report (in ca. 3 pages):

- 1) A screenshot of the theme map and related Analyst Synopsis, and an explanation how the map is interpreted (what do the **colors** mean, what are the **gray nodes** called, and what are they, etc.). Include some of the names of the gray nodes (using slider, e.g. 33-100%) in the screenshot, and the **red year TAGs**
- 2) A screenshot of the concepts list (partial list from the top is sufficient)
- 3) Explain a) how you would query with the tool all instances where *perceived* is discussed in connection to *usefulness*? Hint: note that *perceived* is a concept in the map/concept list, whereas *usefulness* is just a word that appears in the text. Hint2: you can access the User Guide from the Welcome tab in Leximancer. AND b) How would you drill down to see/open the text excerpts containing these two terms for more detailed examination? (you can include screenshots here too). Query with other terms if you have used an alternative topic than IS success for this assignment!
*Although you can search single or multi-item words (such as perceived usefulness) easily in texts, it is not easy to find instances where they appear in the same paragraph but not next to each other, or with a word variant (e.g. “Hence, it is very important to work to get the highest **perception** of **usefulness** in users by enhancing their e-skills.”). Leximancer helps in this kind of examination.*
- 4) How would you utilize Leximancer for other purposes than described in this assignment, e.g. when reading a course e-book or other study materials, such as research article PDFs?
- 5) Finally, how many text sentences per block Leximancer analyzes by default/normally? And what is the range of available options provided for this in the Project control’s settings?

Note that due to the license management issues, it is advised to complete the assignment in one sitting, as the program does not persist on computers.

When you are done, close the browser in VDI, and then sign out from VDI by clicking the gray side handle on the left edge, and clicking the Ctrl-alt-del selector (3 boxes icon) on the top of the screen; select **Sign out**.



Once you are disconnected, log out also from VMware Horizon HTML connection server: **Open the menu** (3 lines icon) and **Log out**.



INSTRUCTIONS FOR RETURNING THE ASSIGNMENT

Return the assignment report to MyCourses – Assignments **by 8.4. at 12:00** (noon).

- Note that -1 point is deducted from every late 24 hours of return.

Note: Check that the CSV file you have edited and saved is actually having commas (,) as item separators and not semicolons (;). You can check this by right-clicking above the Scopus.csv file in the folder, and selecting > Open with > Notepad. If you see semicolons as separators, the file won't work properly in Leximancer as the CSV (Comma separated value) file is not understood correctly!

One remedy is to change already at the beginning of Phase 1 from VDI's Windows's Region settings the Regional format to **English (UK), in order to have dots as decimal separators and not commas (as Finnish regional settings do have) - before you do the export in Scopus.**

An indication of problems with the comma issue is that the file does not open neatly in separate columns when opening it in Excel. See screenshots below, if you are unfamiliar with how to change the regional format settings in Windows (do this in VDI's Windows, if you are using VDI). You can return the settings back to normal, once you have finished working with Leximancer

