

ANLT5070

Unit 4 Assignment 1 Tutorial

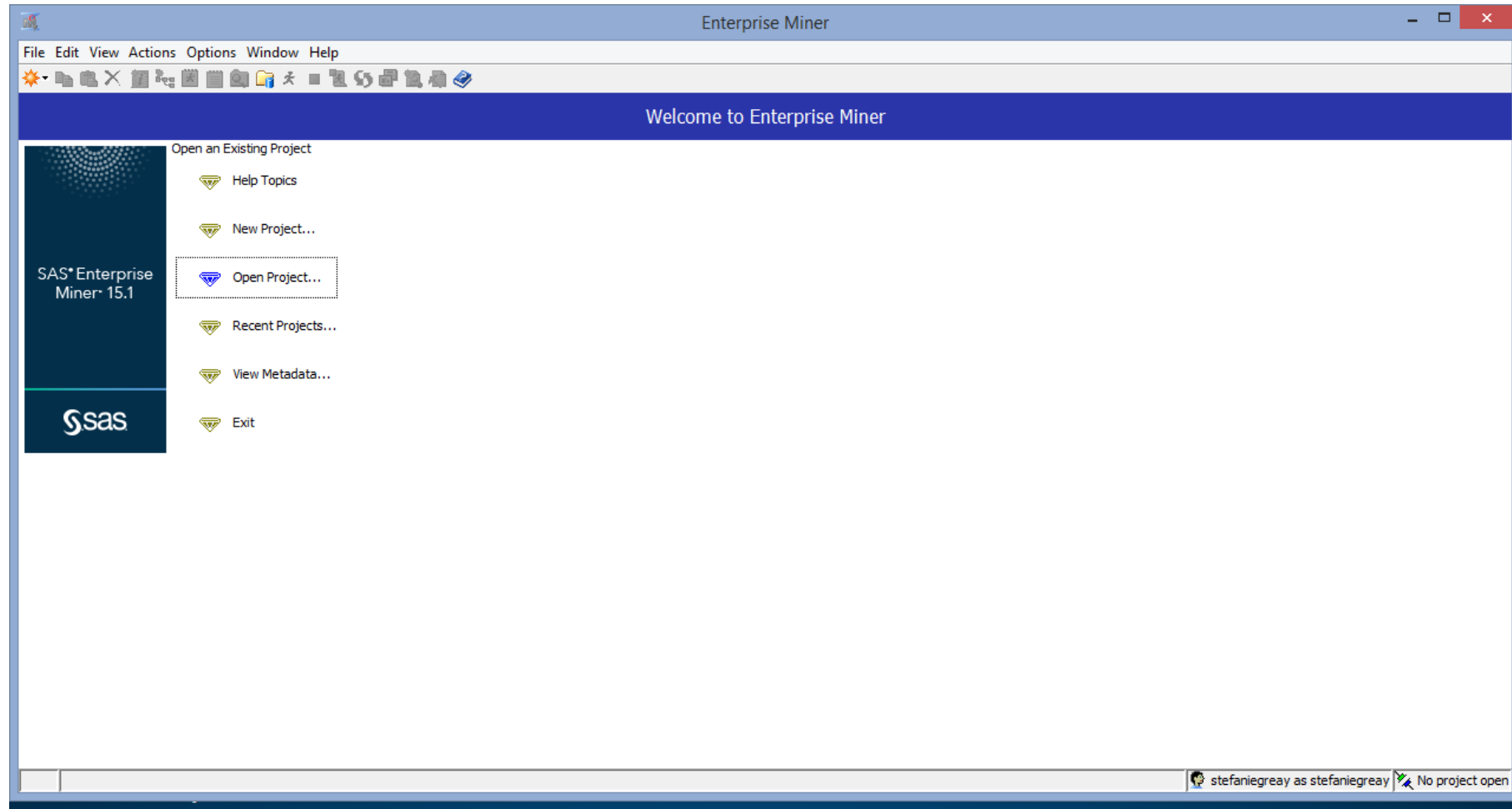


Open project created in Unit 1 and added to in Unit 2 and Unit 3.

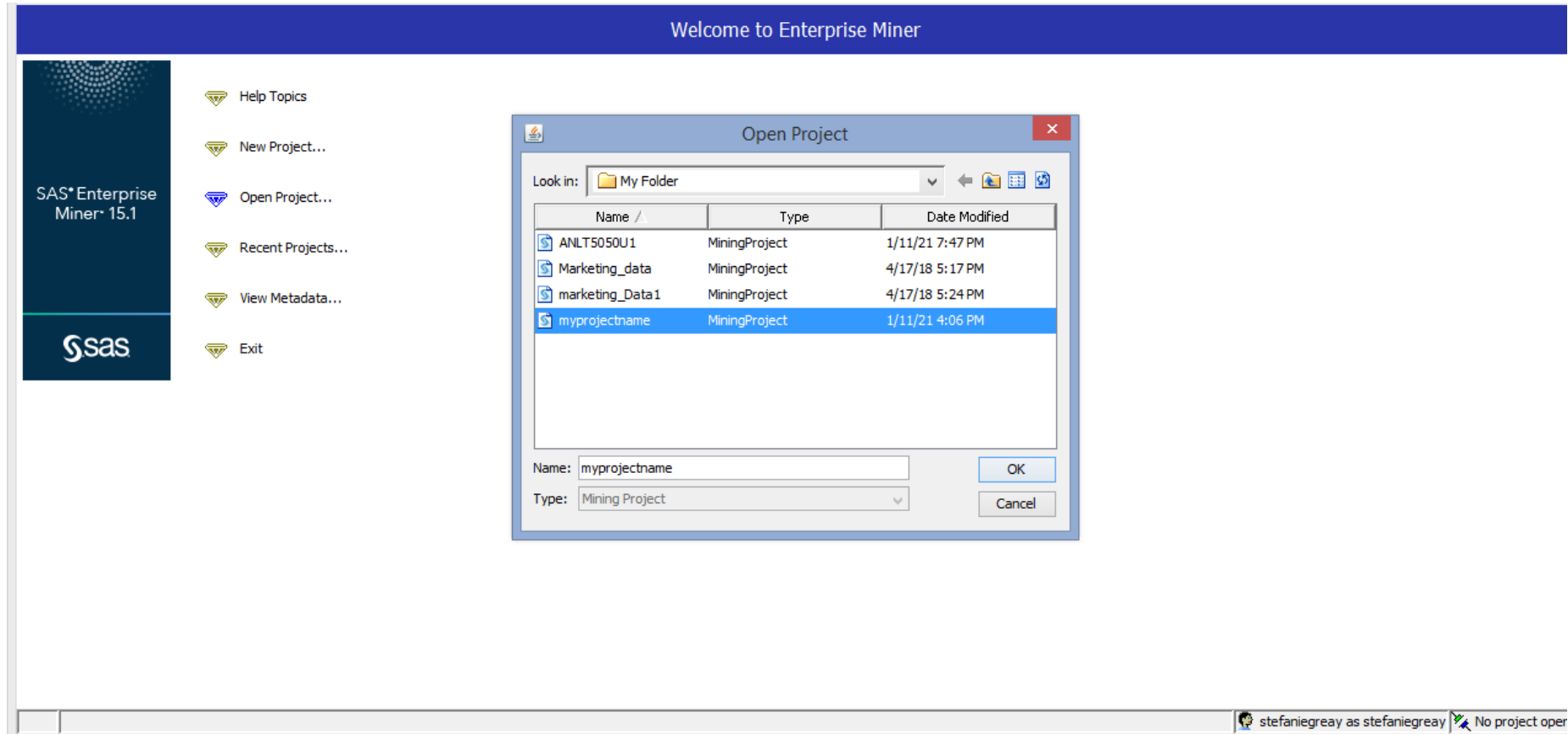
- This assignment is a continuation of the Unit 1 Assignment 1, Unit 2 Assignment 1, and Unit 3 Assignment 1 assignments. If you have not successfully completed these, or you have lost your work for them, please re-watch and re-complete those steps first.
- This assignment focuses on the exploring the results portion of the text analytics project.



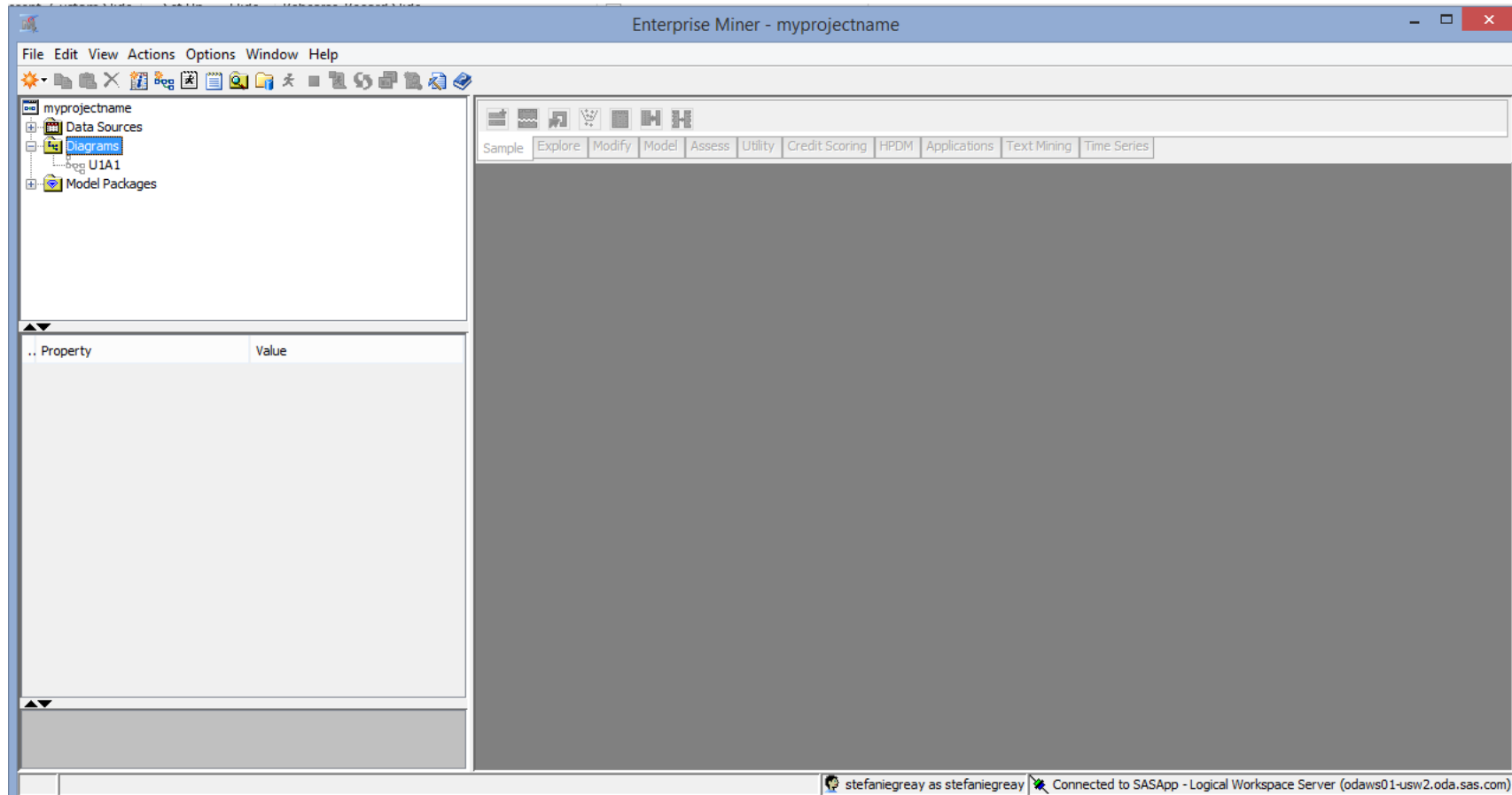
Click on open project



Select the project you created in U1A1, and click “OK” to open the project.



Click on “Diagrams,” navigate to the diagram you created, and double click on it to open it.



Double click on the “Text Filter” node to edit the options for filtering the text data.

The screenshot displays the SAS Enterprise Miner interface. On the left, a project tree shows 'myprojectname' with sub-items 'Data Sources', 'Diagrams', and 'Model Packages'. The 'Diagrams' folder is expanded, showing 'U1A1'. Below the tree is a table of properties for the selected 'TextFilter' node.

Property	Value
General	
Node ID	TextFilter
Imported Data	
Exported Data	
Notes	
Train	
Variables	
Spelling	
Check Spelling	Yes
Dictionary	
Weightings	
Frequency Weighting	Default
Term Weight	Default
Term Filters	
Minimum Number of Documents	4
Maximum Number of Terms	.
Import Synonyms	
Document Filters	
Dictionary	
Specify a data set of correctly spelled terms	

The main workspace shows a workflow diagram for 'U1A1'. It consists of four nodes: 'Text Import', 'Text Parsing', 'Text Filter', and 'SAS Code'. Arrows indicate the flow from 'Text Import' to 'Text Parsing', and from both 'Text Parsing' and 'SAS Code' to 'Text Filter'. Each node has a green checkmark icon. The bottom status bar shows the user 'stefaniegreay' is connected to the SASApp - Logical Workspace Server.



Click on the elipses (3 dots) next to “Filter Viewer” to open the interactive filter viewer.

The screenshot shows the SAS Enterprise Miner interface. On the left, a project tree lists 'myprojectname', 'Data Sources', 'Diagrams', 'U1A1', and 'Model Packages'. Below this is a table of properties for the selected diagram.

Property	Value
Term Weight	Default
Term Filters	
Minimum Number of Documents	2
Maximum Number of Terms	100
Import Synonyms	...
Document Filters	
Search Expression	
Subset Documents	...
Results	
Filter Viewer	...
Spell-Checking Results	EMWS1.TextFilter_spellDS
Exported Synonyms	...
Report	
Terms to View	All
Number of Terms to Display	20000
Status	
Create Time	1/23/21 1:04 AM
Run ID	4ac57bf8-R514-f541-b401-a976340a5aa7
General	

The main workspace displays a workflow diagram for 'U1A1' with the following steps: 'Text Import' → 'Text Parsing' → 'Text Filter'. A 'SAS Code' step is also connected to the 'Text Filter' step. All steps have a green checkmark indicating they are successful. The bottom status bar shows 'Diagram U1A1 opened', the user 'stefaniegreay as stefaniegreay', and the connection 'Connected to SASApp - Logical Workspace Server (odaws01-usw2.oda.sas.com)'.



Type “football” into the search box and click “Apply” to search for the term “football.”

The screenshot shows the 'Interactive Filter Viewer' window. A blue oval highlights the search bar containing the text 'football' and the 'Apply' button. Below the search bar is a table of documents. At the bottom of the window is a 'Terms' section with a table of search results.

TEXT	ACCESSED	CREATED	EXTENSI...	FILTERED	FILTE...	LANGUAGE	MODIFIED	NAME	OMITTED
1 Paper 10360-2016 Nine Frequently Asked Questions about Getting Started with SAS® Visual	2021-01-1...	2021-01-1...	.pdf	/home/ste...	54492.0	English	2021-01-1...	10360-201...	0.0
1 Paper 10401-2016 Responsible Gambling Model at Veikkaus Tero Kallioniemi, Veikkaus Oy	2021-01-1...	2021-01-1...	.pdf	/home/ste...	15834.0	English	2021-01-1...	10401-201...	0.0
1 Paper 10600-2016 You Can Bet On It, The Missing Rows are Preserved with PRELOADFMT	2021-01-1...	2021-01-1...	.pdf	/home/ste...	16612.0	English	2021-01-1...	10600-201...	0.0
1 Paper 10740-2016 Developing an On-Demand Web Report Platform Using Stored Processes	2021-01-1...	2021-01-1...	.pdf	/home/ste...	22946.0	English	2021-01-1...	10740-201...	0.0
1 Paper 11221-2016 Leads and Lags: Static and Dynamic Queues in the SAS® DATA STEP	2021-01-1...	2021-01-1...	.pdf	/home/ste...	29382.0	English	2021-01-1...	11221-201...	0.0
1 Paper 11775 - 2016 Using SAS® Text Miner for Automatic Categorization of Blog Posts on a	2021-01-1...	2021-01-1...	.pdf	/home/ste...	18223.0	English	2021-01-1...	11775-201...	0.0
1 Paper 2101-2016 Using the Kaplan-Meier Product-Limit Estimator to Adjust NFL Yardage	2021-01-1...	2021-01-1...	.pdf	/home/ste...	46710.0	English	2021-01-1...	2101-2016...	0.0
1 Paper 2480-2016 Performing Pattern Matching by Using Perl Regular Expressions Arthur Li,	2021-01-1...	2021-01-1...	.pdf	/home/ste...	33364.0	English	2021-01-1...	2480-2016...	0.0
3940: Fantasizing about the Big Data of NFL Fantasy Football, or Time to Get a Life Clint	2021-01-1...	2021-01-1...	.pdf	/home/ste...	199.0	English	2021-01-1...	3940-2016...	0.0
SAS 5580-2016 MACRO VARIABLES IN SAS® ENTERPRISE GUIDE® Khoi To, Office of Planning	2021-01-1...	2021-01-1...	.pdf	/home/ste...	9118.0	English	2021-01-1...	5580-2016...	0.0
1 SAS 5581-2016 USING PROC TABULATE AND LAG(n) FUNCTION FOR RATES OF CHANGE	2021-01-1...	2021-01-1...	.pdf	/home/ste...	10957.0	English	2021-01-1...	5581-2016...	0.0
6500: Research Problems Arising in Sports Statistics Tim Swartz Content for this session has	2021-01-1...	2021-01-1...	.pdf	/home/ste...	163.0	English	2021-01-1...	6500-2016...	0.0
1 Paper: 7020-2016 Three Methods to Dynamically Assign Colors to Plots Based on Group	2021-01-1...	2021-01-1...	.pdf	/home/ste...	4682.0	English	2021-01-1...	7020-2016...	0.0

TERM	FREQ	# DOCS	KEEP ▼	WEIGHT	ROLE	ATTRIBUTE
product	194	28	✓	0.33		Alpha
program	169	21	✓	0.391		Alpha
option	240	20	✓	0.568		Alpha
output	205	20	✓	0.373		Alpha
add	194	19	✓	0.403		Alpha
generate	113	17	✓	0.415		Alpha
version	61	17	✓	0.473		Alpha
file	261	16	✓	0.439		Alpha
email	75	15	✓	0.442		Alpha
global	167	14	✓	0.667		Alpha
line	87	14	✓	0.453		Alpha
job	67	12	✓	0.507		Alpha



Right click on the results to choose to Toggle to full text or to view the original. If the original will not display, you can reference the documents using the document numbers in the text of the first field.

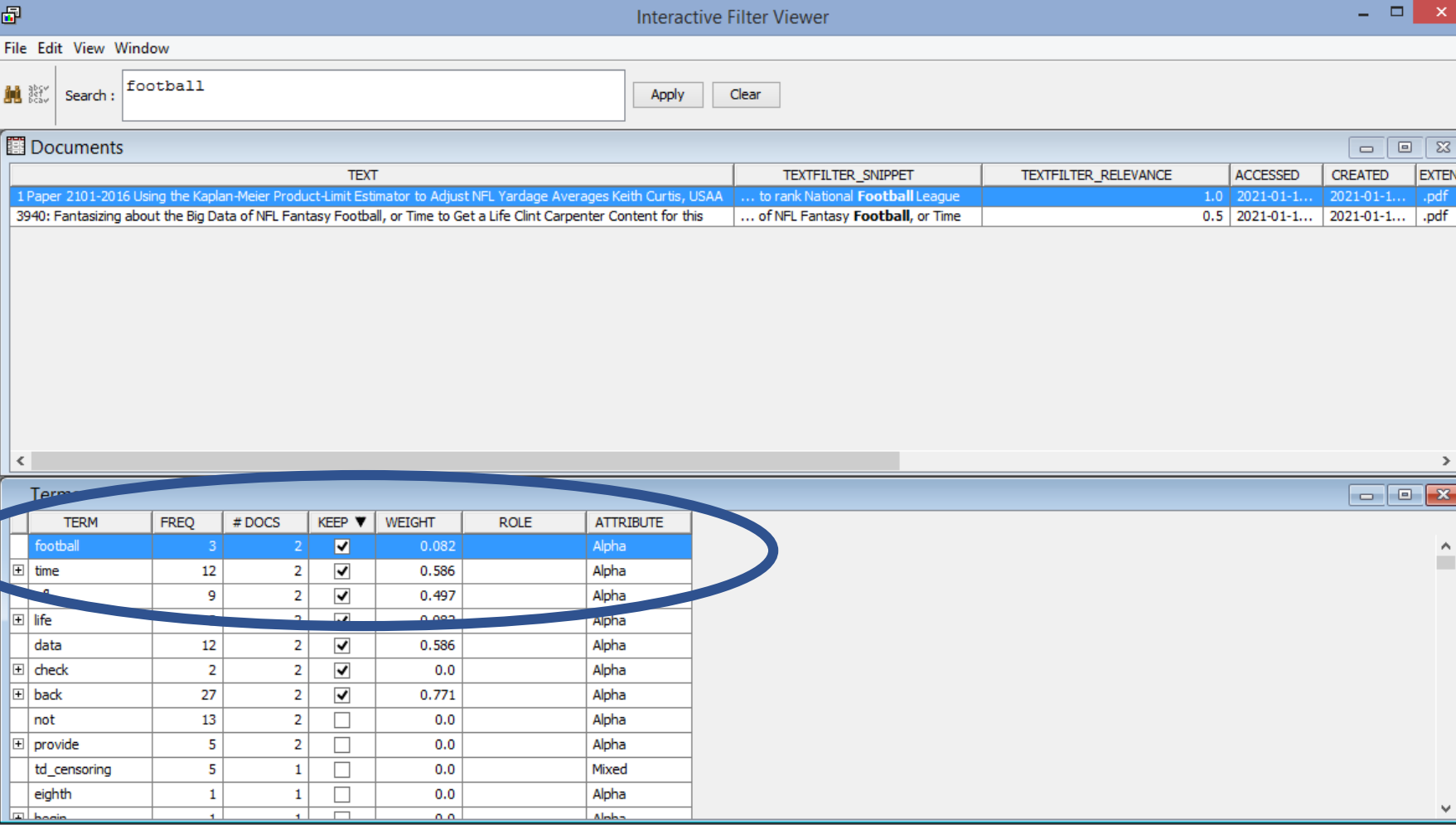
The screenshot shows the 'Interactive Filter Viewer' application. At the top, there is a search bar with the text 'football' and buttons for 'Apply' and 'Clear'. Below the search bar is a table of search results. The first row is highlighted in blue. A right-click context menu is open over the first row, with the 'Toggle Show Full Text' option highlighted. The menu also includes options like 'Find', 'Repeat Find', 'Clear Selection', 'Print...', 'View Input File...', and 'View Original...'. Below the search results table is a 'Terms' section with a table of terms and their frequencies.

TEXT	TEXTFILTER_SNIPPET	TEXTFILTER_RELEVANCE	ACCESSED	CREATED	EXTENSI...	FILTERED
1 Paper 2101-2016 Using the Kaplan-Meier Product-Limit Estimator to rank National Football League	to rank National Football League	1.0	2021-01-1...	2021-01-1...	.pdf	/home/ste...
3940: Fantasizing about the Big Data of NFL Fantasy Football, or	of NFL Fantasy Football, or Time	0.5	2021-01-1...	2021-01-1...	.pdf	/home/ste...

TERM	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
football	3	2	<input checked="" type="checkbox"/>	0.082		Alpha
time	12	2	<input checked="" type="checkbox"/>	0.586		Alpha
nfl	9	2	<input checked="" type="checkbox"/>	0.497		Alpha
life	3	2	<input checked="" type="checkbox"/>	0.082		Alpha
data	12	2	<input checked="" type="checkbox"/>	0.586		Alpha
check	2	2	<input checked="" type="checkbox"/>	0.0		Alpha
back	27	2	<input checked="" type="checkbox"/>	0.771		Alpha
not	13	2	<input type="checkbox"/>	0.0		Alpha
provide	5	2	<input type="checkbox"/>	0.0		Alpha
td_censoring	5	1	<input type="checkbox"/>	0.0		Mixed
eighth	1	1	<input type="checkbox"/>	0.0		Alpha
begin	1	1	<input type="checkbox"/>	0.0		Alpha



You can see the number of times that the term “football” shows up across all 31 documents, as well as the number of documents it is found in.



The screenshot shows the 'Interactive Filter Viewer' application. At the top, there is a search bar with the text 'football' and buttons for 'Apply' and 'Clear'. Below this is a 'Documents' section with a table showing search results. The table has columns: TEXT, TEXTFILTER_SNIPPET, TEXTFILTER_RELEVANCE, ACCESSED, CREATED, and EXTEN. Two documents are listed, both containing the word 'Football' in bold. Below the documents section is a 'Terms' section with a table showing the frequency and document count for various terms. The 'football' term is highlighted with a blue oval. The 'Terms' table has columns: TERM, FREQ, # DOCS, KEEP, WEIGHT, ROLE, and ATTRIBUTE.

TERM	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
football	3	2	<input checked="" type="checkbox"/>	0.082		Alpha
time	12	2	<input checked="" type="checkbox"/>	0.586		Alpha
life	9	2	<input checked="" type="checkbox"/>	0.497		Alpha
data	12	2	<input checked="" type="checkbox"/>	0.586		Alpha
check	2	2	<input checked="" type="checkbox"/>	0.0		Alpha
back	27	2	<input checked="" type="checkbox"/>	0.771		Alpha
not	13	2	<input type="checkbox"/>	0.0		Alpha
provide	5	2	<input type="checkbox"/>	0.0		Alpha
td_censoring	5	1	<input type="checkbox"/>	0.0		Mixed
eight	1	1	<input type="checkbox"/>	0.0		Alpha
begin	1	1	<input type="checkbox"/>	0.0		Alpha



Right click on the “football” record and choose “View Concept Links” to create the concept link map/diagram.

The screenshot shows the 'Interactive Filter Viewer' application. At the top, there is a search bar with the text 'football' and buttons for 'Apply' and 'Clear'. Below the search bar is a 'Documents' section containing a table with columns: TEXT, TEXTFILTER_SNIPPET, TEXTFILTER_RELEVANCE, ACCESSED, CREATED, and EXTEN. The table has two rows of document entries. Below the documents is a 'Terms' section with a table containing columns: TERM, FREQ, # DOCS, KEEP, WEIGHT, ROLE, and ATTRIBUTE. The 'football' term is selected in the 'TERM' column. A right-click context menu is open over the 'football' term, with the option 'View Concept Links' highlighted. A blue oval is drawn around the 'Terms' table and the context menu.

TEXT	TEXTFILTER_SNIPPET	TEXTFILTER_RELEVANCE	ACCESSED	CREATED	EXTEN
1 Paper 2101-2016 Using the Kaplan-Meier Product-Limit Estimator to Adjust NFL Yardage Averages Keith Curtis, USAA	... to rank National Football League	1.0	2021-01-1...	2021-01-1...	.pdf
3940: Fantasizing about the Big Data of NFL Fantasy Football, or Time to Get a Life Clint Carpenter Content for this	... of NFL Fantasy Football , or Time	0.5	2021-01-1...	2021-01-1...	.pdf

TERM	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
football				0.082		Alpha
time				0.586		Alpha
nfl				0.497		Alpha
life				0.082		Alpha
data				0.586		Alpha
check				0.0		Alpha
back				0.751		Alpha
not				0.0		Alpha
provide				0.0		Alpha
td_censoring				0.0		Mixed
eighth				0.0		Alpha
begin				0.0		Alpha



The concept map will display in the bottom right pane of your current screen. (You will need to include your screenshot without the white circle shown below.)

Interactive Filter Viewer

File Edit View Window

Search:

Apply Clear

Documents

TEXT	TEXTFILTER_SNIPPET	TEXTFILTER_RELEVANCE	ACCESSED	CREATED	EXTEN
1 Paper 2101-2016 Using the Kaplan-Meier Product-Limit Estimator to Adjust NFL Yardage Averages Keith Curtis, USAA	... to rank National Football League	1.0	2021-01-1...	2021-01-1...	.pdf
3940: Fantasizing about the Big Data of NFL Fantasy Football, or Time to Get a Life Clint Carpenter Content for this	... of NFL Fantasy Football , or Time	0.5	2021-01-1...	2021-01-1...	.pdf

Terms

TERM	FREQ	# DOCS	KEEP ▼	WEIGHT	ROLE	ATTRIBUTE
football	3	2	<input checked="" type="checkbox"/>	0.082		Alpha
time	12	2	<input checked="" type="checkbox"/>	0.586		Alpha
nfl	9	2	<input checked="" type="checkbox"/>	0.497		Alpha
life	3	2	<input checked="" type="checkbox"/>	0.082		Alpha
data	12	2	<input checked="" type="checkbox"/>	0.586		Alpha
check	2	2	<input checked="" type="checkbox"/>	0.0		Alpha
back	27	2	<input checked="" type="checkbox"/>	0.771		Alpha
not	13	2	<input type="checkbox"/>	0.0		Alpha
provide	5	2	<input type="checkbox"/>	0.0		Alpha
td_censoring	5	1	<input type="checkbox"/>	0.0		Mixed
eighth	1	1	<input type="checkbox"/>	0.0		Alpha
begin	1	1	<input type="checkbox"/>	0.0		Alpha

Concept Linking

```
graph TD; football((football)) --- check[+ check]; football --- data[+ data]; football --- life[+ life]; football --- time[+ time]; football --- back[+ back];
```



Follow the steps outlined in the text below to evaluate the success of the import of the pdf files.

- Chakraborty, G., Pagolu, M., & Garla, S. (2014). Text mining and analysis: practical methods, examples, and case studies using SAS. SAS Institute.
<https://capella.skillport.com/skillportfe/main.action?path=summary/BOOKS/59026>
- Direct link to Chapter 5 – Data Transformation:
<https://capella.skillport.com/skillportfe/assetNonSSOLaunch.action?courseName= ss chapter:59026-350295958&courseType=7>

