VesselAI, 2022: Information Extraction from PDFs, User Guide

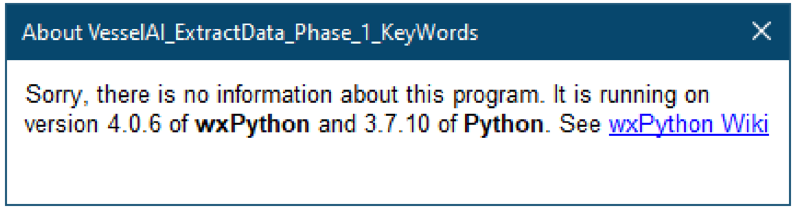
09/03/2023 VTT – beyond the obvious

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957237.

Trying to find marine engine specs & manuals from Internet sources (PDF format)

- We want to test how automatic extraction of information from PDF files would work
- The PDF files describe the performance of ship engines

- This presentation describes 4 Python programs, 3 first of which together implement a workflow for such information extraction from PDF files; NN version is separate
  - VesselAI_ExtractData_1.py
  - VesselAI_ExtractData_2.py
  - VesselAI_ExtractData_3.py
  - VesselAI_ExtractData_NN.py


About VesselAI_ExtractData_Phase_1_KeyWords

Sorry, there is no information about this program. It is running on version 4.0.6 of **wxPython** and 3.7.10 of **Python**. See wxPython Wiki

- The programs were written using Anaconda Python, and wxPython UI library

- Introduction / general information in slides 2-12, the workflow & programs are described in slides 13-19, and installation instructions are in slides 20-22.

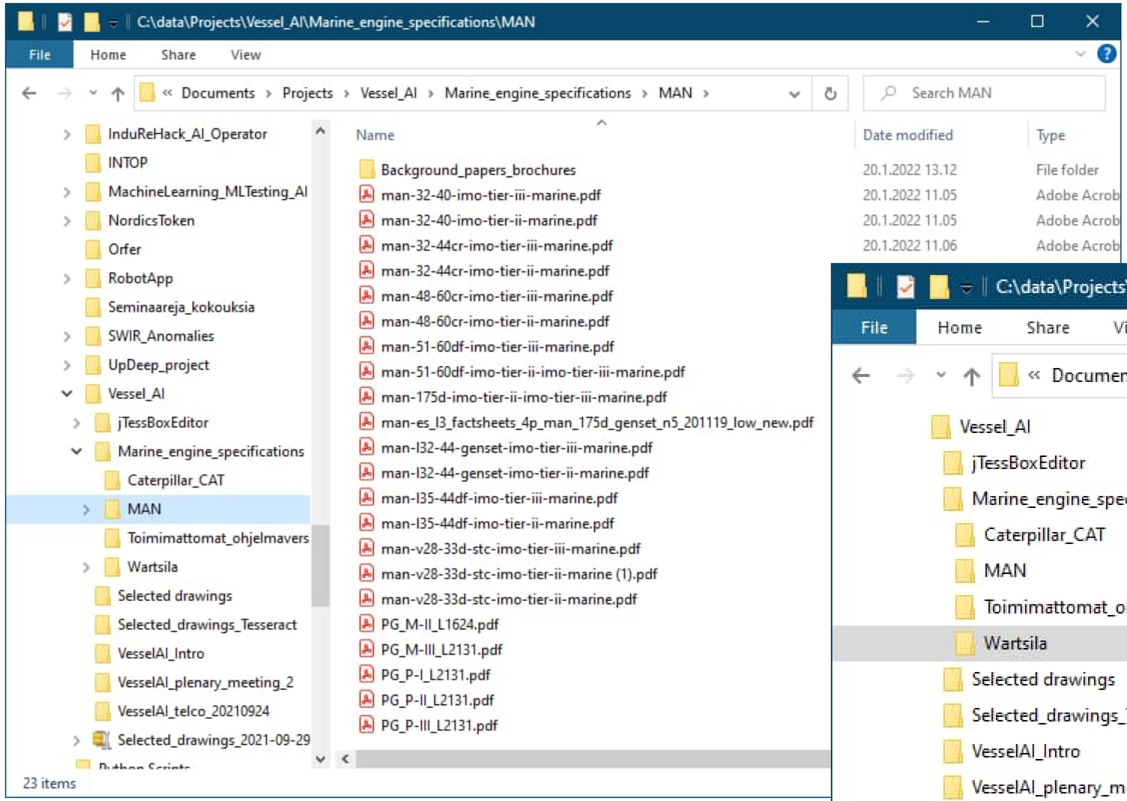
PDFs from largest ship engine manufacturers

- It was decided that we start with the 3 largest marine engine companies: Wärtsilä, MAN, and Caterpillar
- We downloaded all “technical looking” PDFs that consider marine engines
  - From Wärtsilä we downloaded 14 PDFs
  - From MAN we downloaded 22 PDFs
  - Alas, for Caterpillar / CAT there were no PDF specifications that could be freely downloaded…
- To make sure that our code is versatile enough, we later searched for PDF files in WinGD and HHI (Hyundai Heavy Industries) web pages also
  - Google search: engine specification “pdf” site:www.wingd.com
  - Google search: engine specification “pdf” site:english.hhi.co.kr
  - We downloaded 27 files from WinGD, and 11 files from HHI

- This was our test material!

PDF specifications from Wärtsilä and MAN

- These are the Wärtsilä and MAN PDF files that we considered:



C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN

« Documents › Projects › Vessel_AI › Marine_engine_specifications › MAN ›

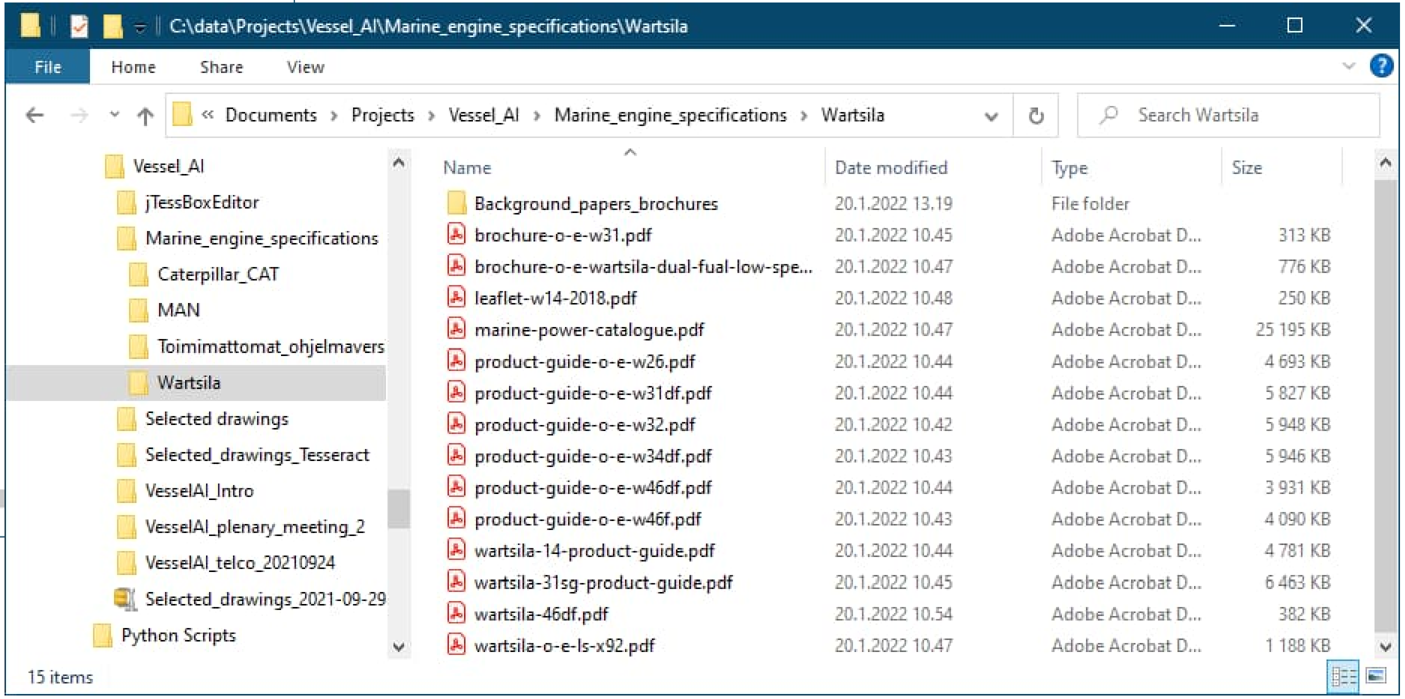
Search MAN

InduReHack_AI_Operator
INTOP
MachineLearning_MLTesting_AI
NordicsToken
Orfer
RobotApp
Seminaareja_kokouksia
SWIR_Anomalies
UpDeep_project
Vessel_AI
jTessBoxEditor
Marine_engine_specifications
Caterpillar_CAT
MAN
Toimimattomat_ohjelmavers
Wartsila
Selected drawings
Selected_drawings_Tesseract
VesselAI_Intro
VesselAI_plenary_meeting_2
VesselAI_telco_20210924
Selected_drawings_2021-09-29

Name	Date modified	Type
Background_papers_brochures	20.1.2022 13.12	File folder
man-32-40-imo-tier-iii-marine.pdf	20.1.2022 11.05	Adobe Acrob
man-32-40-imo-tier-ii-marine.pdf	20.1.2022 11.05	Adobe Acrob
man-32-44cr-imo-tier-iii-marine.pdf	20.1.2022 11.06	Adobe Acrob
man-32-44cr-imo-tier-ii-marine.pdf		
man-48-60cr-imo-tier-iii-marine.pdf		
man-48-60cr-imo-tier-ii-marine.pdf		
man-51-60df-imo-tier-iii-marine.pdf		
man-51-60df-imo-tier-ii-imo-tier-iii-marine.pdf		
man-175d-imo-tier-ii-imo-tier-iii-marine.pdf		
man-es_l3_factsheets_4p_man_175d_genset_n5_201119_low_new.pdf		
man-l32-44-genset-imo-tier-iii-marine.pdf		
man-l32-44-genset-imo-tier-ii-marine.pdf		
man-l35-44df-imo-tier-iii-marine.pdf		
man-l35-44df-imo-tier-ii-marine.pdf		
man-v28-33d-stc-imo-tier-iii-marine.pdf		
man-v28-33d-stc-imo-tier-ii-marine (1).pdf		
man-v28-33d-stc-imo-tier-ii-marine.pdf		
PG_M-II_L1624.pdf		
PG_M-III_L2131.pdf		
PG_P-I_L2131.pdf		
PG_P-II_L2131.pdf		
PG_P-III_L2131.pdf		

23 items

- 22 PDFs from MAN, 14 from Wärtsilä…
- 27 files from WinGD, 11 files from HHI



C:\data\Projects\Vessel_AI\Marine_engine_specifications\Wartsila

« Documents › Projects › Vessel_AI › Marine_engine_specifications › Wartsila

Search Wartsila

Vessel_AI
jTessBoxEditor
Marine_engine_specifications
Caterpillar_CAT
MAN
Toimimattomat_ohjelmavers
Wartsila
Selected drawings
Selected_drawings_Tesseract
VesselAI_Intro
VesselAI_plenary_meeting_2
VesselAI_telco_20210924
Selected_drawings_2021-09-29
Python Scripts

Name	Date modified	Type	Size
Background_papers_brochures	20.1.2022 13.19	File folder	
brochure-o-e-w31.pdf	20.1.2022 10.45	Adobe Acrobat D...	313 KB
brochure-o-e-wartsila-dual-fual-low-spe...	20.1.2022 10.47	Adobe Acrobat D...	776 KB
leaflet-w14-2018.pdf	20.1.2022 10.48	Adobe Acrobat D...	250 KB
marine-power-catalogue.pdf	20.1.2022 10.47	Adobe Acrobat D...	25 195 KB
product-guide-o-e-w26.pdf	20.1.2022 10.44	Adobe Acrobat D...	4 693 KB
product-guide-o-e-w31df.pdf	20.1.2022 10.44	Adobe Acrobat D...	5 827 KB
product-guide-o-e-w32.pdf	20.1.2022 10.42	Adobe Acrobat D...	5 948 KB
product-guide-o-e-w34df.pdf	20.1.2022 10.43	Adobe Acrobat D...	5 946 KB
product-guide-o-e-w46df.pdf	20.1.2022 10.44	Adobe Acrobat D...	3 931 KB
product-guide-o-e-w46f.pdf	20.1.2022 10.43	Adobe Acrobat D...	4 090 KB
wartsila-14-product-guide.pdf	20.1.2022 10.44	Adobe Acrobat D...	4 781 KB
wartsila-31sg-product-guide.pdf	20.1.2022 10.45	Adobe Acrobat D...	6 463 KB
wartsila-46df.pdf	20.1.2022 10.54	Adobe Acrobat D...	382 KB
wartsila-o-e-ls-x92.pdf	20.1.2022 10.47	Adobe Acrobat D...	1 188 KB

15 items

PDF files NOT included!

- There are 73 PDF files in total, containing 18 019 pages
- These files are (or at least were in spring 2022) publicly available
- However, due to copyright issues we cannot distribute them
- Lists of these files are presented in the next slides
- You must try to find them from the vendors' web sites and download them yourself
- Google searches for MAN, Wärtsilä, HHI, WinGD PDF files (**or:** "engine manual"…):
  - engine specification "pdf" site: www.man-es.com/marine/
  - engine specification "pdf" site: www.wartsila.com
  - engine specification "pdf" site:english.hhi.co.kr
  - engine specification "pdf" site:www.wingd.com
- Save the found files into subfolders 'MAN', 'Wartsila', 'Hyundai_HHI', and 'WinGD'
- However, the Python programs contain hard-wired lists of the files to be processed. If you cannot find all these listed files, you must modify these lists
- But the NN (neural network) program version also contains the indices of documents and pages that the human expert considered relevant ("Ground Truth"). These index lists must be modified also, which is tedious…

PDF files NOT included! MAN files to find

- MAN/man-175d-imo-tier-ii-imo-tier-iii-marine.pdf
- MAN/man-32-40-imo-tier-ii-marine.pdf
- MAN/man-32-40-imo-tier-iii-marine.pdf
- MAN/man-32-44cr-imo-tier-ii-marine.pdf
- MAN/man-32-44cr-imo-tier-iii-marine.pdf
- MAN/man-48-60cr-imo-tier-ii-marine.pdf
- MAN/man-48-60cr-imo-tier-iii-marine.pdf
- MAN/man-51-60df-imo-tier-ii-imo-tier-iii-marine.pdf
- MAN/man-51-60df-imo-tier-iii-marine.pdf
- MAN/man-es_l3_factsheets_4p_man_175d_genset_n5_201119_low_new.pdf
- MAN/man-l32-44-genset-imo-tier-ii-marine.pdf
- MAN/man-l32-44-genset-imo-tier-iii-marine.pdf
- MAN/man-l35-44df-imo-tier-ii-marine.pdf
- MAN/man-l35-44df-imo-tier-iii-marine.pdf
- MAN/man-v28-33d-stc-imo-tier-ii-marine.pdf
- MAN/man-v28-33d-stc-imo-tier-iii-marine.pdf
- MAN/PG_M-III_L2131.pdf
- MAN/PG_M-II_L1624.pdf
- MAN/PG_P-III_L2131.pdf
- MAN/PG_P-II_L2131.pdf
- MAN/PG_P-I_L2131.pdf

PDF files NOT included! Wärtsilä files to find

- Wartsila/brochure-o-e-w31.pdf
- Wartsila/brochure-o-e-wartsila-dual-fual-low-speed.pdf
- Wartsila/leaflet-w14-2018.pdf
- Wartsila/marine-power-catalogue.pdf
- Wartsila/product-guide-o-e-w26.pdf
- Wartsila/product-guide-o-e-w31df.pdf
- Wartsila/product-guide-o-e-w32.pdf
- Wartsila/product-guide-o-e-w34df.pdf
- Wartsila/product-guide-o-e-w46df.pdf
- Wartsila/product-guide-o-e-w46f.pdf
- Wartsila/wartsila-14-product-guide.pdf
- Wartsila/wartsila-31sg-product-guide.pdf
- Wartsila/wartsila-46df.pdf
- Wartsila/wartsila-o-e-ls-x92.pdf

PDF files NOT included! HHI files to find

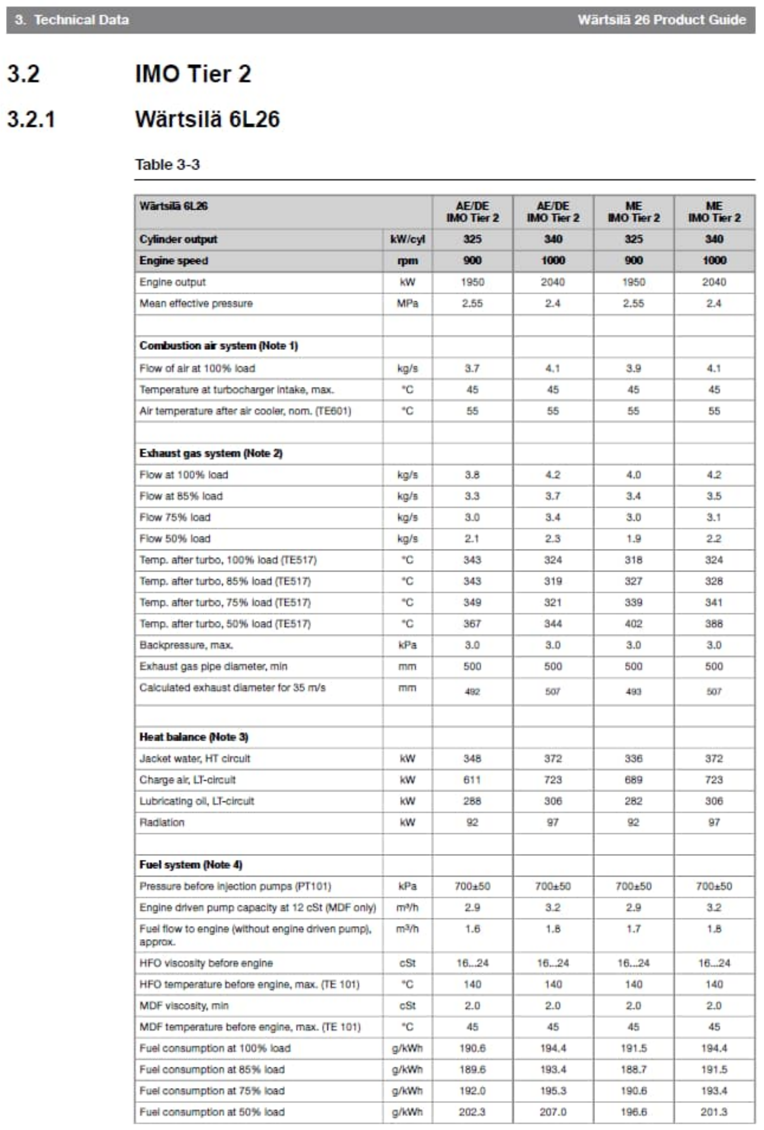
- Hyundai_HHI/2016_HHI_en.pdf
- Hyundai_HHI/2017_HHI_en.pdf
- Hyundai_HHI/2018_HHI_en.pdf
- Hyundai_HHI/2019_HHI_en.pdf
- Hyundai_HHI/2020_KSOE_IR_EN.pdf
- Hyundai_HHI/2021_KSOE_IR_EN.pdf
- Hyundai_HHI/HHI_EMD_brochure2017_1.pdf
- Hyundai_HHI/HHI_EMD_brochure2017_2.pdf
- Hyundai_HHI/HHI_EMD_brochure2017_3.pdf
- Hyundai_HHI/HHI_EMD_brochure2019_1.pdf
- Hyundai_HHI/speclal_NSD2020.pdf

PDF files NOT included! WinGD files to find

- WinGD/16-06-pres-kit_x92.pdf
- WinGD/cimac2016_120_kyrtatos_paper_thedevelopmentofthemodern2strokemarinedieselengine.pdf
- WinGD/cimac2016_173_brueckl_paper_virtualdesign-and-simulation-in2strokemarineenginedevelopment.pdf
- WinGD/cimac2016_233_ott_paper_x-df.pdf
- WinGD/dominikschneiter_tieriii-programme.pdf
- WinGD/Fuel-Flexible-Injection-System-How-to-Handle-a-Fuel-Specturm-from-Diesel_CIMAC2019_paper_404_Andreas-Schmid.pdf
- WinGD/marcspahni_generation-x-engines.pdf
- WinGD/MIM_WinGD-RT-flex48T-D.pdf
- WinGD/MIM_WinGD-X72.pdf
- WinGD/MM_WinGD-RT-flex58T-D.pdf
- WinGD/MM_WinGD-X35-B_2021-09.pdf
- WinGD/MM_WinGD-X52.pdf
- WinGD/MM_WinGD-X72.pdf
- WinGD/MM_WinGD-X72DF.pdf
- WinGD/MM_WinGD-X82-B.pdf
- WinGD/Motorship-May-2018-VP-R-D-Leader-Brief.pdf
- WinGD/OM_WinGD-X62_2021-09.pdf
- WinGD/OM_WinGD-X72DF_2021-09.pdf
- WinGD/OM_WinGD-X82-B.pdf
- WinGD/OM_WinGD_RT-flex50DF.pdf
- WinGD/WinGD-12X92DF-Development-of-the-Most-Powerful-Otto-Engine-Ever_CIMAC2019_paper_425_Dominik-Schneiter.pdf
- WinGD/wingd-paper_engine_selection_for_very_large_container_vessels_201609.pdf
- WinGD/WinGD-WiDE-Brochure.pdf
- WinGD/WinGD_Engine-Booklet_2021.pdf
- WinGD/wingd_moving-inlet-ports-concept-for-optimization-of-2-stroke-uni-flow-engines_patrick-rebecchi.pdf
- WinGD/X-DF-Engines-by-WinGD.pdf
- WinGD/X-DF-FAQ.pdf

Wärtsilä PDF files: Relevant tables

- Then a few of the Wärtsilä PDF files were inspected
- The first one that contains relevant numerical information was ‘product-guide-o-e-w26.pdf’
- It contains many tables of the same format
- The page that contains the first such table is shown here:
- The four last lines give **fuel consumption** at 100% / 85 % / 75 % / 50 % engine load [g/kWh]
- The four first lines in group “Exhaust gas system” give **exhaust gas flow** at 100% / 85 % / 75 % / 50 % engine load [kg/s]
- There was no **charge air flow**; however, the first line in group “Combustion air system” gives “Flow of air at 100 % load” [kg/s]; we will use these numbers instead
- Header data above is crucial; we want to know the engine model etc. to which the numbers apply!



3. Technical Data | Wärtsilä 26 Product Guide

3.2 IMO Tier 2

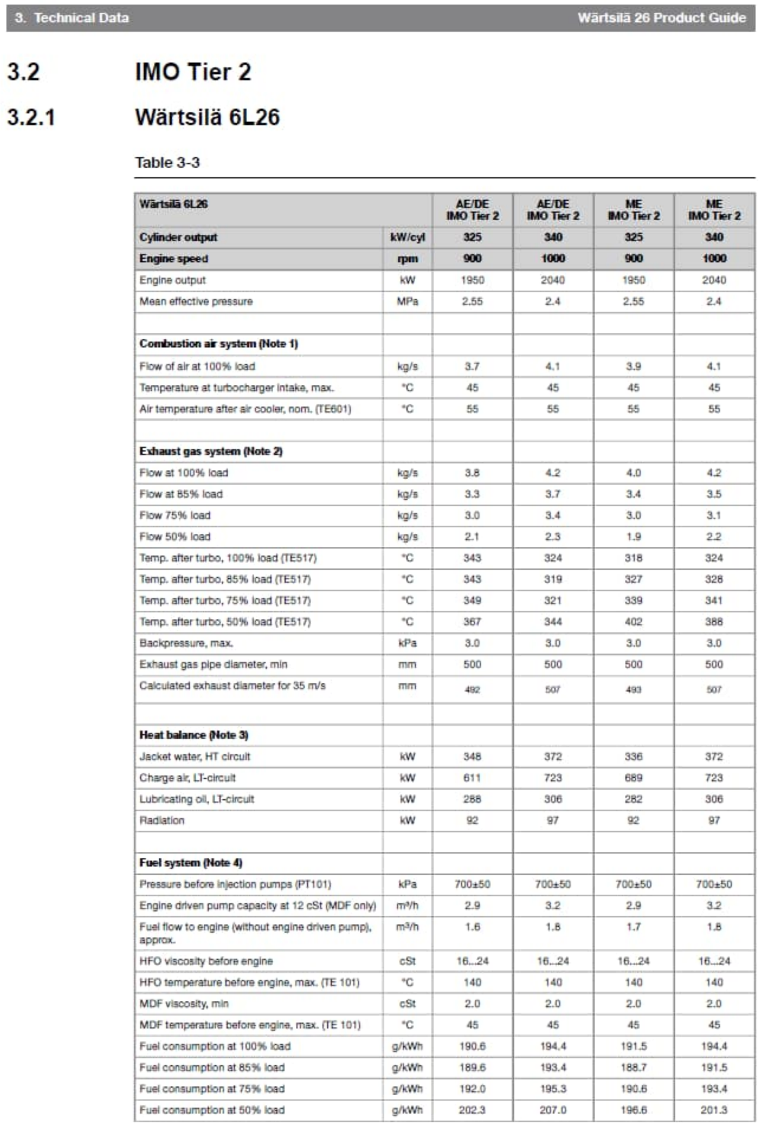
3.2.1 Wärtsilä 6L26

Table 3-3

Wärtsilä 6L26		AE/DE IMO Tier 2	AE/DE IMO Tier 2	ME IMO Tier 2	ME IMO Tier 2
Cylinder output	**kW/cyl**	**325**	**340**	**325**	**340**
Engine speed	**rpm**	**900**	**1000**	**900**	**1000**
Engine output	kW	1950	2040	1950	2040
Mean effective pressure	MPa	2.55	2.4	2.55	2.4
Combustion air system (Note 1)					
Flow of air at 100% load	kg/s	3.7	4.1	3.9	4.1
Temperature at turbocharger intake, max.	°C	45	45	45	45
Air temperature after air cooler, nom. (TE601)	°C	55	55	55	55
Exhaust gas system (Note 2)					
Flow at 100% load	kg/s	3.8	4.2	4.0	4.2
Flow at 85% load	kg/s	3.3	3.7	3.4	3.5
Flow 75% load	kg/s	3.0	3.4	3.0	3.1
Flow 50% load	kg/s	2.1	2.3	1.9	2.2
Temp. after turbo, 100% load (TE517)	°C	343	324	318	324
Temp. after turbo, 85% load (TE517)	°C	343	319	327	328
Temp. after turbo, 75% load (TE517)	°C	349	321	339	341
Temp. after turbo, 50% load (TE517)	°C	367	344	402	388
Backpressure, max.	kPa	3.0	3.0	3.0	3.0
Exhaust gas pipe diameter, min	mm	500	500	500	500
Calculated exhaust diameter for 35 m/s	mm	492	507	493	507
Heat balance (Note 3)					
Jacket water, HT circuit	kW	348	372	336	372
Charge air, LT-circuit	kW	611	723	689	723
Lubricating oil, LT-circuit	kW	288	306	282	306
Radiation	kW	92	97	92	97
Fuel system (Note 4)					
Pressure before injection pumps (PT101)	kPa	700±50	700±50	700±50	700±50
Engine driven pump capacity at 12 cSt (MDF only)	m³/h	2.9	3.2	2.9	3.2
Fuel flow to engine (without engine driven pump), approx.	m³/h	1.6	1.8	1.7	1.8
HFO viscosity before engine	cSt	16...24	16...24	16...24	16...24
HFO temperature before engine, max. (TE 101)	°C	140	140	140	140
MDF viscosity, min	cSt	2.0	2.0	2.0	2.0
MDF temperature before engine, max. (TE 101)	°C	45	45	45	45
Fuel consumption at 100% load	g/kWh	190.6	194.4	191.5	194.4
Fuel consumption at 85% load	g/kWh	189.6	193.4	188.7	191.5
Fuel consumption at 75% load	g/kWh	192.0	195.3	190.6	193.4
Fuel consumption at 50% load	g/kWh	202.3	207.0	196.6	201.3

PDF files: Relevant tables

- After much testing we settled for the following rules:
- We first search for these keywords (can be inside () or []):
  - g/kWh
  - kJ/kWh
  - kg/s
  - kg/kWh
- But with two upper keywords the following other words must also be found in the same page:
  - fuel **AND** consumption
  - **OR** sfoc **OR** bsfc
- And with two lower keywords the following other words must also be found in the same page:
  - charge air **OR** combustion air **OR** exhaust gas
  - **AND** flow
- This produced good enough results!



3. Technical Data   Wärtsilä 26 Product Guide

3.2 IMO Tier 2

3.2.1 Wärtsilä 6L26

Table 3-3

Wärtsilä 6L26		**AE/DE IMO Tier 2**	**AE/DE IMO Tier 2**	**ME IMO Tier 2**	**ME IMO Tier 2**
Cylinder output	**kW/cyl**	**325**	**340**	**325**	**340**
Engine speed	**rpm**	**900**	**1000**	**900**	**1000**
Engine output	kW	1950	2040	1950	2040
Mean effective pressure	MPa	2.55	2.4	2.55	2.4
Combustion air system (Note 1)					
Flow of air at 100% load	kg/s	3.7	4.1	3.9	4.1
Temperature at turbocharger intake, max.	°C	45	45	45	45
Air temperature after air cooler, nom. (TE601)	°C	55	55	55	55
Exhaust gas system (Note 2)					
Flow at 100% load	kg/s	3.8	4.2	4.0	4.2
Flow at 85% load	kg/s	3.3	3.7	3.4	3.5
Flow 75% load	kg/s	3.0	3.4	3.0	3.1
Flow 50% load	kg/s	2.1	2.3	1.9	2.2
Temp. after turbo, 100% load (TE517)	°C	343	324	318	324
Temp. after turbo, 85% load (TE517)	°C	343	319	327	328
Temp. after turbo, 75% load (TE517)	°C	349	321	339	341
Temp. after turbo, 50% load (TE517)	°C	367	344	402	388
Backpressure, max.	kPa	3.0	3.0	3.0	3.0
Exhaust gas pipe diameter, min	mm	500	500	500	500
Calculated exhaust diameter for 35 m/s	mm	492	507	493	507
Heat balance (Note 3)					
Jacket water, HT circuit	kW	348	372	336	372
Charge air, LT-circuit	kW	611	723	689	723
Lubricating oil, LT-circuit	kW	288	306	282	306
Radiation	kW	92	97	92	97
Fuel system (Note 4)					
Pressure before injection pumps (PT101)	kPa	700±50	700±50	700±50	700±50
Engine driven pump capacity at 12 cSt (MDF only)	m³/h	2.9	3.2	2.9	3.2
Fuel flow to engine (without engine driven pump), approx.	m³/h	1.6	1.8	1.7	1.8
HFO viscosity before engine	cSt	16...24	16...24	16...24	16...24
HFO temperature before engine, max. (TE 101)	°C	140	140	140	140
MDF viscosity, min	cSt	2.0	2.0	2.0	2.0
MDF temperature before engine, max. (TE 101)	°C	45	45	45	45
Fuel consumption at 100% load	g/kWh	190.6	194.4	191.5	194.4
Fuel consumption at 85% load	g/kWh	189.6	193.4	188.7	191.5
Fuel consumption at 75% load	g/kWh	192.0	195.3	190.6	193.4
Fuel consumption at 50% load	g/kWh	202.3	207.0	196.6	201.3

PDF files: Learning Program

- Another version of the original Python program was written
- Instead of selecting keywords (e.g. “kJ/kWh”, “fuel”, “consumption”; see previous slide) we select a few of our data pages as “Model Pages”
- For example, some Model Pages could be related to fuel consumption, some to charge air / exhaust gas / heat balance
- Then the program computes the relative frequencies of all the words in each Model Page
- And when a candidate page is processed, a similar word frequency histogram is computed, and compared to the frequency histograms of all the Model Pages
- Words that can be converted to floating-point numbers are excluded from the diff. sums
- If the best difference is small enough (e.g. <=0.5), i.e. that candidate page matches at least one of the Model Pages, then that candidate page is accepted

PDF files: Whole workflow

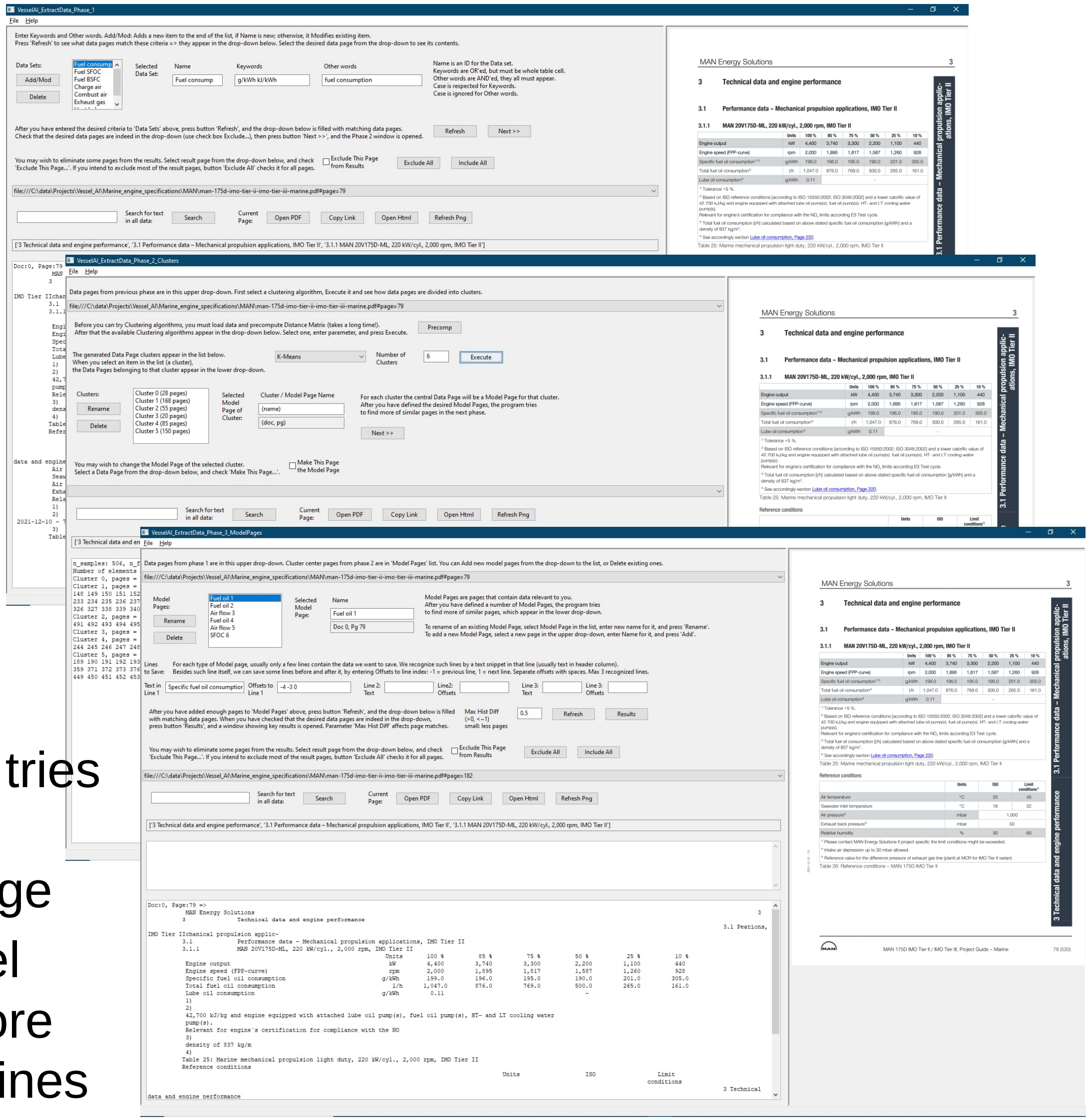
- **New versions of the Python programs discussed previously were written, so that together the new programs create a whole workflow for information extraction:**

1. We extend the first version of our program (using keywords) so that the user enters some keywords (e.g. fuel consumption). Then the program finds those *.pdf file **pages** that match these keywords. Optionally the user can eliminate some (irrelevant) pages. The remaining pages go to phase 2.
2. We cluster the found pages into groups (i.e. clusters); group center pages will be the proposed Model Pages. Optionally the user can eliminate some (irrelevant) Model Pages. The remaining pages go to phase 3.
3. We extend the second version of our program (learning version) to use these Model Pages, and then to find more of similar pages within the PDF files. Besides, for each model page the user can define rules for lines where the relevant data values are, thus reducing the amount of irrelevant information even further.

- And now we have a toolchain that can extract information from a large number of PDF files: Return desired parameter values for desired marine engine types.

Implementation of the workflow (1/4)

- Still using our set of downloaded PDF files
- Can be later extended to web crawlers…
- Three separate Python programs to realize the workflow:
  - VesselAI_ExtractData_1.py
  - VesselAI_ExtractData_2.py
  - VesselAI_ExtractData_3.py
- In Phase 1 the user enters keywords and other words, and the program finds the matching files/pages; links to these files/pages are saved
- The Phase 2 program reads these files, and tries to group into clusters (4 algorithms); central page in each cluster is saved as a Model Page
- In Phase 3 the user can edit this set of Model Pages; and then the program tries to find more of similar pages, as well as extract relevant lines



Enter Keywords and Other words. Add/Mod: Adds a new item to the end of the list, if Name is new; otherwise, it Modifies existing item.
Press 'Refresh' to see what data pages match these criteria => they appear in the drop-down below. Select the desired data page from the drop-down to see its contents.

Data Sets: Fuel consump, Fuel SFOC, Fuel BSFC, Charge air, Combust air, Exhaust gas

Selected Data Set: Name: Fuel consump; Keywords: g/kWh kJ/kWh; Other words: fuel consumption

Name is an ID for the Data set.
Keywords are OR'ed, but must be whole table cell.
Other words are AND'ed, they all must appear.
Case is respected for Keywords.
Case is ignored for Other words.

After you have entered the desired criteria to 'Data Sets' above, press button 'Refresh', and the drop-down below is filled with matching data pages.
Check that the desired data pages are indeed in the drop-down (use check box Exclude...), then press button 'Next >>', and the Phase 2 window is opened.

You may wish to eliminate some pages from the results. Select result page from the drop-down below, and check 'Exclude This Page...'. If you intend to exclude most of the result pages, button 'Exclude All' checks it for all pages. Exclude This Page from Results

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Search for text in all data: Current Page:

['3 Technical data and engine performance', '3.1 Performance data – Mechanical propulsion applications, IMO Tier II', '3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II']

Data pages from previous phase are in this upper drop-down. First select a clustering algorithm, Execute it and see how data pages are divided into clusters.

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Before you can try Clustering algorithms, you must load data and precompute Distance Matrix (takes a long time!).
After that the available Clustering algorithms appear in the drop-down below. Select one, enter parameter, and press Execute.

The generated Data Page clusters appear in the list below.
When you select an item in the list (a cluster),
the Data Pages belonging to that cluster appear in the lower drop-down.

K-Means; Number of Clusters: 6

Clusters: Cluster 0 (28 pages), Cluster 1 (168 pages), Cluster 2 (55 pages), Cluster 3 (20 pages), Cluster 4 (85 pages), Cluster 5 (150 pages)

Selected Model Page of Cluster: Cluster / Model Page Name: (name); (doc, pg)

For each cluster the central Data Page will be a Model Page for that cluster.
After you have defined the desired Model Pages, the program tries
to find more of similar pages in the next phase.

You may wish to change the Model Page of the selected cluster.
Select a Data Page from the drop-down below, and check 'Make This Page...'. Make This Page the Model Page

```
n_samples: 506, n_f
Number of elements
Cluster 0, pages =
Cluster 1, pages =
148 149 150 151 152
233 234 235 236 237
326 327 338 339 340
Cluster 2, pages =
491 492 493 494 495
Cluster 3, pages =
Cluster 4, pages =
244 245 246 247 248
Cluster 5, pages =
189 190 191 192 193
359 371 372 373 376
449 450 451 452 453
```

Data pages from phase 1 are in this upper drop-down. Cluster center pages from phase 2 are in 'Model Pages' list. You can Add new model pages from the drop-down to the list, or Delete existing ones.

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Model Pages: Fuel oil 1, Fuel oil 2, Air flow 3, Fuel oil 4, Air flow 5, SFOC 6

Selected Model Page: Name: Fuel oil 1; Doc 0, Pg 79

Model Pages are pages that contain data relevant to you.
After you have defined a number of Model Pages, the program tries
to find more of similar pages, which appear in the lower drop-down.

To rename of an existing Model Page, select Model Page in the list, enter new name for it, and press 'Rename'.
To add a new Model Page, select a new page in the upper drop-down, enter Name for it, and press 'Add'.

Lines to Save: For each type of Model page, usually only a few lines contain the data we want to save. We recognize such lines by a text snippet in that line (usually text in header column).
Besides such line itself, we can save some lines before and after it, by entering Offsets to line index: -1 = previous line, 1 = next line. Separate offsets with spaces. Max 3 recognized lines.

Text in Line 1: Specific fuel oil consumptior; Offsets to Line 1: -4 -3 0; Line 2: Text; Line2: Offsets; Line 3: Text; Line 3: Offsets

After you have added enough pages to 'Model Pages' above, press button 'Refresh', and the drop-down below is filled with matching data pages. When you have checked that the desired data pages are indeed in the drop-down, press button 'Results', and a window showing key results is opened. Parameter 'Max Hist Diff' affects page matches.

Max Hist Diff (>0, <~1) small: less pages: 0.5

You may wish to eliminate some pages from the results. Select result page from the drop-down below, and check 'Exclude This Page...'. If you intend to exclude most of the result pages, button 'Exclude All' checks it for all pages. Exclude This Page from Results

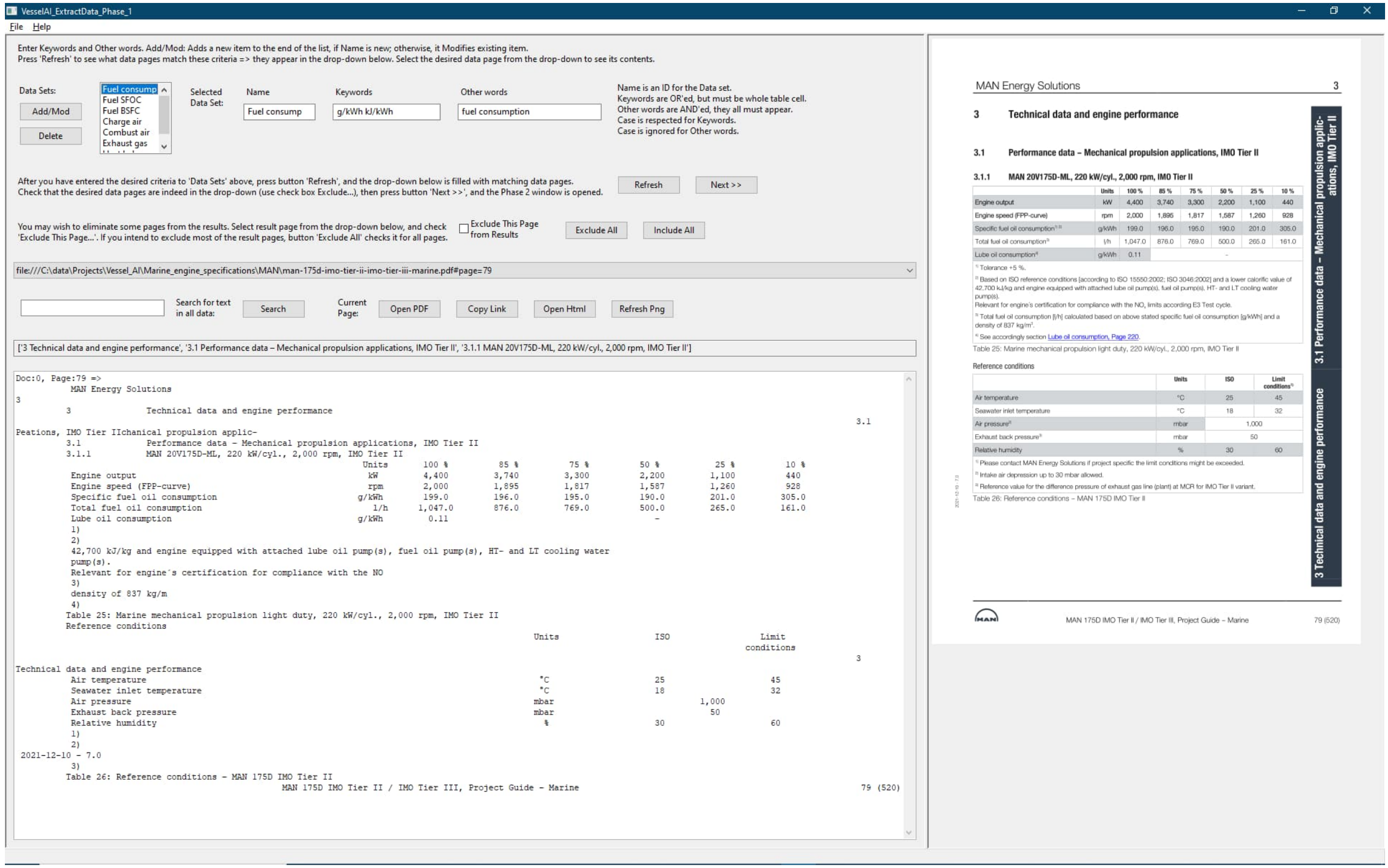
file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=182

['3 Technical data and engine performance', '3.1 Performance data – Mechanical propulsion applications, IMO Tier II', '3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II']

```
Doc:0, Page:79 =>
          MAN Energy Solutions                                                                                        3
         3              Technical data and engine performance
                                                                                                           3.1 Peations,
IMO Tier IIchanical propulsion applic-
         3.1            Performance data - Mechanical propulsion applications, IMO Tier II
         3.1.1          MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II
                                                      Units      100 %       85 %       75 %       50 %       25 %       10 %
          Engine output                                kW        4,400      3,740      3,300      2,200      1,100       440
          Engine speed (FPP-curve)                     rpm       2,000      1,895      1,817      1,587      1,260       928
          Specific fuel oil consumption               g/kWh      199.0      196.0      195.0      190.0      201.0      305.0
          Total fuel oil consumption                   l/h     1,047.0      876.0      769.0      500.0      265.0      161.0
          Lube oil consumption                        g/kWh       0.11                            -
          1)
          2)
          42,700 kJ/kg and engine equipped with attached lube oil pump(s), fuel oil pump(s), HT- and LT cooling water
          pump(s).
          Relevant for engine's certification for compliance with the NO
          3)
          density of 837 kg/m
          4)
         Table 25: Marine mechanical propulsion light duty, 220 kW/cyl., 2,000 rpm, IMO Tier II
         Reference conditions
                                                                       Units            ISO            Limit
                                                                                                    conditions
                                                                                                           3 Technical
data and engine performance
```

Implementation of the workflow (2/4): Phase 1

- The user enters a number of Data Sets, each contains Name (=ID for DS), Keywords, Other words; the list of defined Data Sets is shown
- The Data Set definitions are saved for next use. At first use: reasonable default values
- The algorithm of this program is executed with button ‘Refresh’, and the found files/pages are shown in a drop-down, and the selected page can be seen as an image and as text
- User can exclude some of the result pages from results
- Phase 2 program is started with button ‘Next >>’



Enter Keywords and Other words. Add/Mod: Adds a new item to the end of the list, if Name is new; otherwise, it Modifies existing item.
Press 'Refresh' to see what data pages match these criteria => they appear in the drop-down below. Select the desired data page from the drop-down to see its contents.

Data Sets: Fuel consump, Fuel SFOC, Fuel BSFC, Charge air, Combust air, Exhaust gas

Selected Data Set:	Name	Keywords	Other words
	Fuel consump	g/kWh kJ/kWh	fuel consumption

Name is an ID for the Data set.
Keywords are OR'ed, but must be whole table cell.
Other words are AND'ed, they all must appear.
Case is respected for Keywords.
Case is ignored for Other words.

After you have entered the desired criteria to 'Data Sets' above, press button 'Refresh', and the drop-down below is filled with matching data pages.
Check that the desired data pages are indeed in the drop-down (use check box Exclude...), then press button 'Next >>', and the Phase 2 window is opened.

You may wish to eliminate some pages from the results. Select result page from the drop-down below, and check 'Exclude This Page...'. If you intend to exclude most of the result pages, button 'Exclude All' checks it for all pages.

[ ] Exclude This Page from Results

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Search for text in all data:

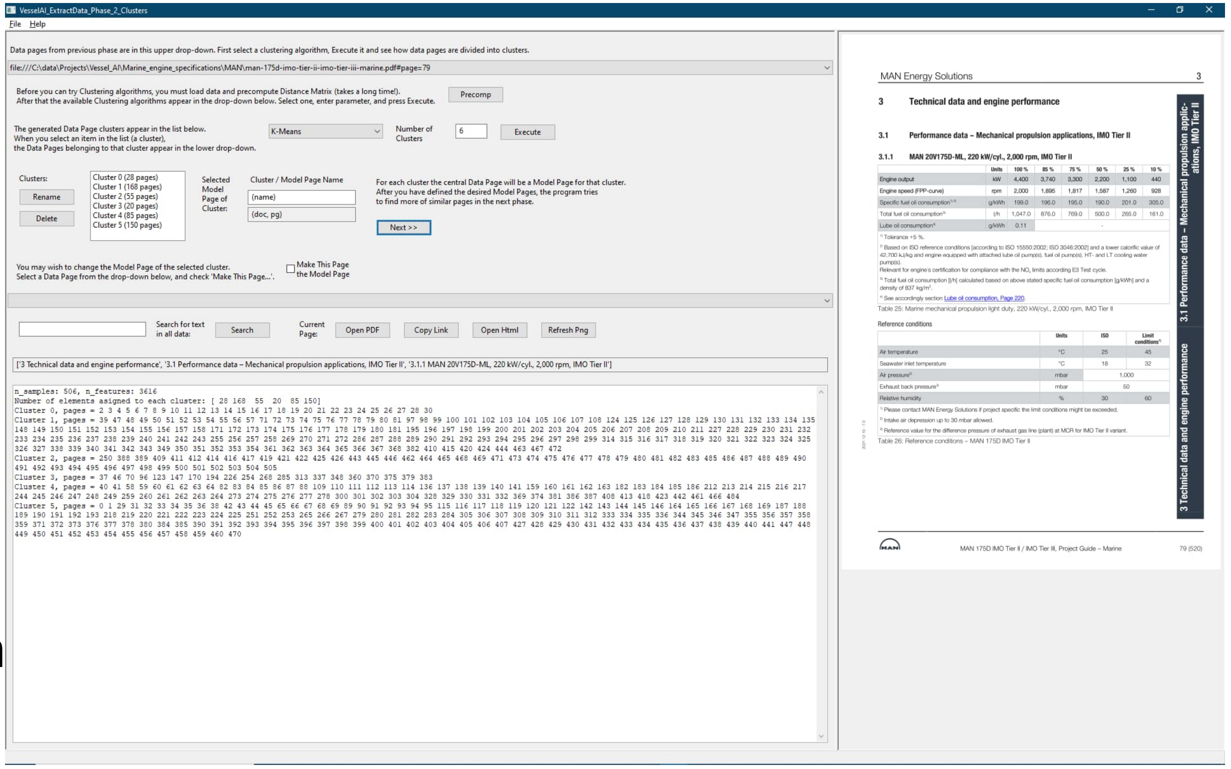
Current Page:

['3 Technical data and engine performance', '3.1 Performance data – Mechanical propulsion applications, IMO Tier II', '3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II']

```
Doc:0, Page:79 =>
          MAN Energy Solutions
3
         3              Technical data and engine performance
                                                                                                     3.1
Peations, IMO Tier IIchanical propulsion applic-
         3.1            Performance data - Mechanical propulsion applications, IMO Tier II
         3.1.1          MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II
                                                    Units      100 %       85 %       75 %       50 %       25 %       10 %
          Engine output                              kW        4,400      3,740      3,300      2,200      1,100       440
          Engine speed (FPP-curve)                   rpm       2,000      1,895      1,817      1,587      1,260       928
          Specific fuel oil consumption            g/kWh       199.0      196.0      195.0      190.0      201.0      305.0
          Total fuel oil consumption                 l/h     1,047.0      876.0      769.0      500.0      265.0      161.0
          Lube oil consumption                     g/kWh        0.11                              -
          1)
          2)
          42,700 kJ/kg and engine equipped with attached lube oil pump(s), fuel oil pump(s), HT- and LT cooling water
          pump(s).
          Relevant for engine´s certification for compliance with the NO
          3)
          density of 837 kg/m
          4)
         Table 25: Marine mechanical propulsion light duty, 220 kW/cyl., 2,000 rpm, IMO Tier II
         Reference conditions
                                                                    Units             ISO             Limit
                                                                                                   conditions
                                                                                                             3
Technical data and engine performance
          Air temperature                                            °C               25              45
          Seawater inlet temperature                                 °C               18              32
          Air pressure                                              mbar                   1,000
          Exhaust back pressure                                     mbar                     50
          Relative humidity                                          %                30              60
          1)
          2)
 2021-12-10 - 7.0
          3)
         Table 26: Reference conditions - MAN 175D IMO Tier II
                                     MAN 175D IMO Tier II / IMO Tier III, Project Guide - Marine                     79 (520)
```

Implementation of the workflow (3/4): Phase 2

- The program tries to group Phase 1 result pages into clusters
- The central page of each cluster becomes a Model Page for Phase 3
- At first two algorithms for page clustering were available: **K-means** and **DBSCAN**; both from scikit learn (a.k.a. sklearn) program library:
  - https://scikit-learn.org/stable/auto_examples/text/plot_document_clustering.html#sphx-glr-auto-examples-text-plot-document-clustering-py
  - (“Clustering text documents using k-means”)
  - https://programmer.group/61755e726f9b3.html
  - (“Self defined distance measurement method of clustering algorithm in scikit learn Library”)
  - Later **AffinityPropagation** and **SpectralClustering** algorithms from scikit learn were also added!
- A lengthy Precomputation step is necessary before the 4 algorithms can be tried, but clustering itself is fast


VesselAI_ExtractData_Phase_2_Clusters
File
Help
Precomp
K-Means
Number of Clusters
6
Execute
Clusters:
Rename
Delete
Selected Model Page of Cluster:
Cluster / Model Page Name
(name)
(doc, pg)
Next >>
Make This Page the Model Page
Search for text in all data:
Search
Current Page:
Open PDF
Copy Link
Open Html
Refresh Png

Data pages from previous phase are in this upper drop-down. First select a clustering algorithm, Execute it and see how data pages are divided into clusters.

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Before you can try Clustering algorithms, you must load data and precompute Distance Matrix (takes a long time!).
After that the available Clustering algorithms appear in the drop-down below. Select one, enter parameter, and press Execute.

The generated Data Page clusters appear in the list below.
When you select an item in the list (a cluster),
the Data Pages belonging to that cluster appear in the lower drop-down.

Cluster 0 (28 pages)
Cluster 1 (168 pages)
Cluster 2 (55 pages)
Cluster 3 (20 pages)
Cluster 4 (85 pages)
Cluster 5 (150 pages)

For each cluster the central Data Page will be a Model Page for that cluster.
After you have defined the desired Model Pages, the program tries
to find more of similar pages in the next phase.

You may wish to change the Model Page of the selected cluster.
Select a Data Page from the drop-down below, and check 'Make This Page...'.

['3 Technical data and engine performance', '3.1 Performance data – Mechanical propulsion applications, IMO Tier II', '3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II']

```
n_samples: 506, n_features: 3616
Number of elements asigned to each cluster: [ 28 168  55  20  85 150]
Cluster 0, pages = 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 30
Cluster 1, pages = 39 47 48 49 50 51 52 53 54 55 56 57 71 72 73 74 75 76 77 78 79 80 81 97 98 99 100 101 102 103 104 105 106 107 108 124 125 126 127 128 129 130 131 132 133 134 135
148 149 150 151 152 153 154 155 156 157 158 171 172 173 174 175 176 177 178 179 180 181 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 227 228 229 230 231 232
233 234 235 236 237 238 239 240 241 242 243 255 256 257 258 269 270 271 272 286 287 288 289 290 291 292 293 294 295 296 297 298 299 314 315 316 317 318 319 320 321 322 323 324 325
326 327 338 339 340 341 342 343 349 350 351 352 353 354 361 362 363 364 365 366 367 368 382 410 415 420 424 444 463 467 472
Cluster 2, pages = 250 388 389 409 411 412 414 416 417 419 421 422 425 426 443 445 446 462 464 465 468 469 471 473 474 475 476 477 478 479 480 481 482 483 485 486 487 488 489 490
491 492 493 494 495 496 497 498 499 500 501 502 503 504 505
Cluster 3, pages = 37 46 70 96 123 147 170 194 226 254 268 285 313 337 348 360 370 375 379 383
Cluster 4, pages = 40 41 58 59 60 61 62 63 64 82 83 84 85 86 87 88 109 110 111 112 113 114 136 137 138 139 140 141 159 160 161 162 163 182 183 184 185 186 212 213 214 215 216 217
244 245 246 247 248 249 259 260 261 262 263 264 273 274 275 276 277 278 300 301 302 303 304 328 329 330 331 332 369 374 381 386 387 408 413 418 423 442 461 466 484
Cluster 5, pages = 0 1 29 31 32 33 34 35 36 38 42 43 44 45 65 66 67 68 69 89 90 91 92 93 94 95 115 116 117 118 119 120 121 122 142 143 144 145 146 164 165 166 167 168 169 187 188
189 190 191 192 193 218 219 220 221 222 223 224 225 251 252 253 265 266 267 279 280 281 282 283 284 305 306 307 308 309 310 311 312 333 334 335 336 344 345 346 347 355 356 357 358
359 371 372 373 376 377 378 380 384 385 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 447 448
449 450 451 452 453 454 455 456 457 458 459 460 470
```

MAN Energy Solutions 3

3 Technical data and engine performance

3.1 Performance data – Mechanical propulsion applications, IMO Tier II

3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II

	Units	100 %	85 %	75 %	50 %	25 %	10 %
Engine output	kW	4,400	3,740	3,300	2,200	1,100	440
Engine speed (FPP-curve)	rpm	2,000	1,895	1,817	1,587	1,260	928
Specific fuel oil consumption[1][2]	g/kWh	199.0	196.0	195.0	190.0	201.0	305.0
Total fuel oil consumption[3]	l/h	1,047.0	876.0	769.0	500.0	265.0	161.0
Lube oil consumption[4]	g/kWh	0.11	-				

[1] Tolerance +5 %.

[2] Based on ISO reference conditions [according to ISO 15550:2002; ISO 3046:2002] and a lower calorific value of 42,700 kJ/kg and engine equipped with attached lube oil pump(s), fuel oil pump(s), HT- and LT cooling water pump(s).
Relevant for engine's certification for compliance with the NO_x limits according E3 Test cycle.

[3] Total fuel oil consumption [l/h] calculated based on above stated specific fuel oil consumption [g/kWh] and a density of 837 kg/m³.

[4] See accordingly section Lube oil consumption, Page 220.

Table 25: Marine mechanical propulsion light duty, 220 kW/cyl., 2,000 rpm, IMO Tier II

Reference conditions

	Units	ISO	Limit conditions[1]
Air temperature	°C	25	45
Seawater inlet temperature	°C	18	32
Air pressure[2]	mbar	1,000	
Exhaust back pressure[3]	mbar	50	
Relative humidity	%	30	60

[1] Please contact MAN Energy Solutions if project specific the limit conditions might be exceeded.

[2] Intake air depression up to 30 mbar allowed.

[3] Reference value for the difference pressure of exhaust gas line (plant) at MCR for IMO Tier II variant.

Table 26: Reference conditions – MAN 175D IMO Tier II

3.1 Performance data – Mechanical propulsion applications, IMO Tier II

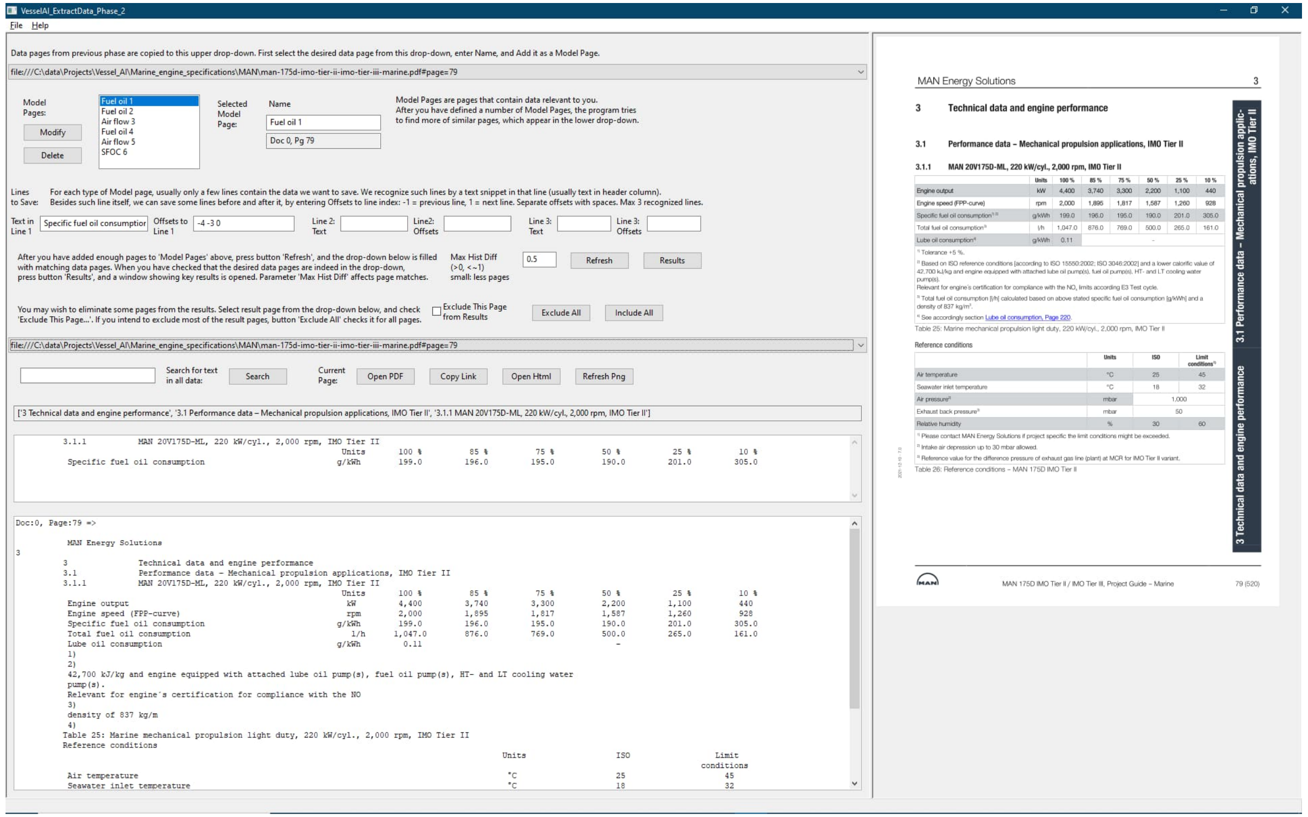
3 Technical data and engine performance

2021-12-10 - 7.0

MAN 175D IMO Tier II / IMO Tier III, Project Guide – Marine 79 (520)

Implementation of the workflow (4/4): Phase 3

- The user selects a number Phase 2 result pages as Model Pages, each contains Name (=ID for MP) and rules for lines of relevant data; the list of defined Model Pages is shown
- The contents of this list comes from Phase 2 (no data: reasonable default contents)
- For each Model Page, the user must define the lines where relevant data values are found
- The algorithm of this program is executed with button 'Refresh', and the found files/pages are shown in a drop-down, and the selected page can be seen as an image and as text. Data in relevant lines shown separately
- User can exclude some of the result pages from results
- Result data (=relevant lines) are shown with button 'Results'. New Notepad window is opened



Data pages from previous phase are copied to this upper drop-down. First select the desired data page from this drop-down, enter Name, and Add it as a Model Page.

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Model Pages: Fuel oil 1, Fuel oil 2, Air flow 3, Fuel oil 4, Air flow 5, SFOC 6

Selected Model Page: Name: Fuel oil 1; Doc 0, Pg 79

Model Pages are pages that contain data relevant to you.
After you have defined a number of Model Pages, the program tries to find more of similar pages, which appear in the lower drop-down.

Lines to Save: For each type of Model page, usually only a few lines contain the data we want to save. We recognize such lines by a text snippet in that line (usually text in header column).
Besides such line itself, we can save some lines before and after it, by entering Offsets to line index: -1 = previous line, 1 = next line. Separate offsets with spaces. Max 3 recognized lines.

Text in Line 1	Offsets to Line 1	Line 2: Text	Line2: Offsets	Line 3: Text	Line 3: Offsets
Specific fuel oil consumptior	-4 -3 0				

After you have added enough pages to 'Model Pages' above, press button 'Refresh', and the drop-down below is filled with matching data pages. When you have checked that the desired data pages are indeed in the drop-down, press button 'Results', and a window showing key results is opened. Parameter 'Max Hist Diff' affects page matches.

Max Hist Diff (>0, <~1) small: less pages: 0.5

You may wish to eliminate some pages from the results. Select result page from the drop-down below, and check 'Exclude This Page...'. If you intend to exclude most of the result pages, button 'Exclude All' checks it for all pages.

☐ Exclude This Page from Results

file:///C:\data\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

Search for text in all data:

Current Page:

['3 Technical data and engine performance', '3.1 Performance data – Mechanical propulsion applications, IMO Tier II', '3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II']

```
         3.1.1           MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II
                                                             Units       100 %        85 %        75 %         50 %        25 %        10 %
          Specific fuel oil consumption                     g/kWh       199.0       196.0       195.0        190.0       201.0       305.0
```

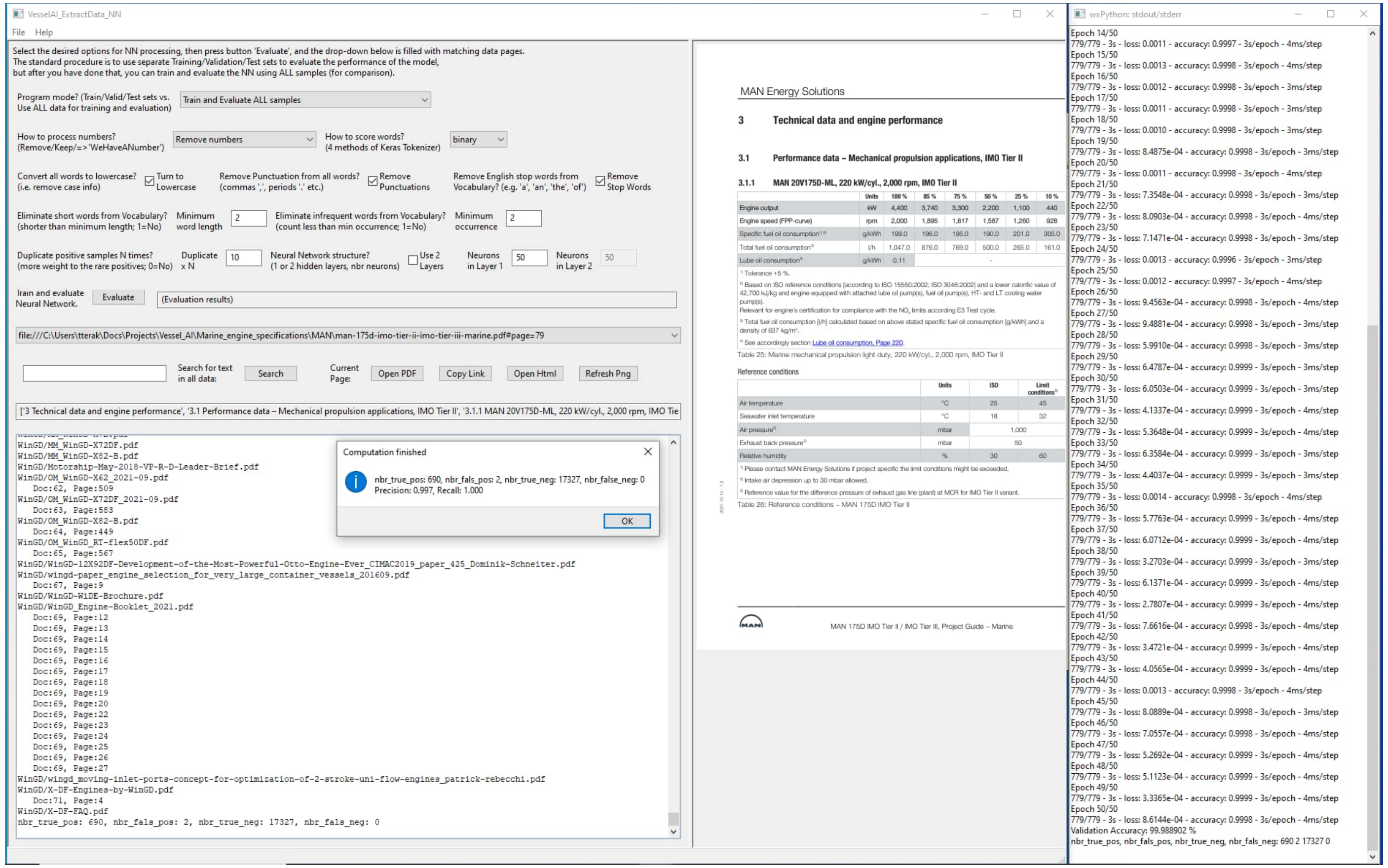
```
Doc:0, Page:79 =>

          MAN Energy Solutions
3
         3               Technical data and engine performance
         3.1             Performance data - Mechanical propulsion applications, IMO Tier II
         3.1.1           MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II
                                                             Units       100 %        85 %        75 %         50 %        25 %        10 %
          Engine output                                      kW         4,400       3,740       3,300        2,200       1,100        440
          Engine speed (FPP-curve)                           rpm        2,000       1,895       1,817        1,587       1,260        928
          Specific fuel oil consumption                     g/kWh       199.0       196.0       195.0        190.0       201.0       305.0
          Total fuel oil consumption                          l/h      1,047.0      876.0       769.0        500.0       265.0       161.0
          Lube oil consumption                              g/kWh       0.11                                 -
          1)
          2)
          42,700 kJ/kg and engine equipped with attached lube oil pump(s), fuel oil pump(s), HT- and LT cooling water
          pump(s).
          Relevant for engine´s certification for compliance with the NO
          3)
          density of 837 kg/m
          4)
         Table 25: Marine mechanical propulsion light duty, 220 kW/cyl., 2,000 rpm, IMO Tier II
         Reference conditions
                                                                                     Units              ISO                Limit
                                                                                                                         conditions
          Air temperature                                                            °C                 25                   45
          Seawater inlet temperature                                                 °C                 18                   32
```

Notes

- Please note that when you open the window of the next phase (button ‘Next >>’), a new window is opened, but the original window remains open behind this new window; however, because the new window uses the same Python/wxPython environment as the original one, the original window remains inactive until the new window is closed
- Lengthy operations (e.g. ‘Precomp’ in Phase 2, ‘Refresh’ in Phases 1 and 3) are executed in the main thread, so the UI is frozen until the operation is completed. Progress is shown in the bottom subwindow, but this window is not always updated properly; especially if you move the focus into another window, the UI becomes frozen until the computation ends
- The programs have been tested in Windows 11, Anaconda Python; in other environments (e.g. Linux) there may be problems…

Neural network program

- Ground truth of relevant pages by human expert: 690 pages were deemed relevant
- NN program: many settings were tried, results of best one
  - Precision = 0.998, Recall = 0.979, Accuracy = 0.999
- NN produced much better results than either Phase 1 or Phase 3
- However, for new data (e.g. new marine engine manufacturer) both NN and Phase 3 are likely to fail
- But Phase 1 could still work: SI units (our keywords) and technical terms (our other words) are the same



Select the desired options for NN processing, then press button 'Evaluate', and the drop-down below is filled with matching data pages.
The standard procedure is to use separate Training/Validation/Test sets to evaluate the performance of the model,
but after you have done that, you can train and evaluate the NN using ALL samples (for comparison).

Program mode? (Train/Valid/Test sets vs. Use ALL data for training and evaluation): Train and Evaluate ALL samples

How to process numbers? (Remove/Keep/=>'WeHaveANumber'): Remove numbers

How to score words? (4 methods of Keras Tokenizer): binary

Convert all words to lowercase? (i.e. remove case info): [✓] Turn to Lowercase

Remove Punctuation from all words? (commas ',', periods '.' etc.): [✓] Remove Punctuations

Remove English stop words from Vocabulary? (e.g. 'a', 'an', 'the', 'of'): [✓] Remove Stop Words

Eliminate short words from Vocabulary? (shorter than minimum length; 1=No): Minimum word length 2

Eliminate infrequent words from Vocabulary? (count less than min occurrence; 1=No): Minimum occurrence 2

Duplicate positive samples N times? (more weight to the rare positives; 0=No): Duplicate x N 10

Neural Network structure? (1 or 2 hidden layers, nbr neurons): [ ] Use 2 Layers; Neurons in Layer 1: 50; Neurons in Layer 2: 50

Train and evaluate Neural Network. Evaluate (Evaluation results)

file:///C:\Users\tterak\Docs\Projects\Vessel_AI\Marine_engine_specifications\MAN\man-175d-imo-tier-ii-imo-tier-iii-marine.pdf#page=79

['3 Technical data and engine performance', '3.1 Performance data – Mechanical propulsion applications, IMO Tier II', '3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tie

```
WinGD/MM_WinGD-X72DF.pdf
WinGD/MM_WinGD-X82-B.pdf
WinGD/Motorship-May-2018-VP-R-D-Leader-Brief.pdf
WinGD/OM_WinGD-X62_2021-09.pdf
   Doc:62, Page:509
WinGD/OM_WinGD-X72DF_2021-09.pdf
   Doc:63, Page:583
WinGD/OM_WinGD-X82-B.pdf
   Doc:64, Page:449
WinGD/OM_WinGD_RT-flex50DF.pdf
   Doc:65, Page:567
WinGD/WinGD-12X92DF-Development-of-the-Most-Powerful-Otto-Engine-Ever_CIMAC2019_paper_425_Dominik-Schneiter.pdf
WinGD/wingd-paper_engine_selection_for_very_large_container_vessels_201609.pdf
   Doc:67, Page:9
WinGD/WinGD-WiDE-Brochure.pdf
WinGD/WinGD_Engine-Booklet_2021.pdf
   Doc:69, Page:12
   Doc:69, Page:13
   Doc:69, Page:14
   Doc:69, Page:15
   Doc:69, Page:16
   Doc:69, Page:17
   Doc:69, Page:18
   Doc:69, Page:19
   Doc:69, Page:20
   Doc:69, Page:22
   Doc:69, Page:23
   Doc:69, Page:24
   Doc:69, Page:25
   Doc:69, Page:26
   Doc:69, Page:27
WinGD/wingd_moving-inlet-ports-concept-for-optimization-of-2-stroke-uni-flow-engines_patrick-rebecchi.pdf
WinGD/X-DF-Engines-by-WinGD.pdf
   Doc:71, Page:4
WinGD/X-DF-FAQ.pdf
nbr_true_pos: 690, nbr_fals_pos: 2, nbr_true_neg: 17327, nbr_fals_neg: 0
```

Computation finished

nbr_true_pos: 690, nbr_fals_pos: 2, nbr_true_neg: 17327, nbr_false_neg: 0
Precision: 0.997, Recall: 1.000

MAN Energy Solutions

3 Technical data and engine performance

3.1 Performance data – Mechanical propulsion applications, IMO Tier II

3.1.1 MAN 20V175D-ML, 220 kW/cyl., 2,000 rpm, IMO Tier II

	Units	100 %	85 %	75 %	50 %	25 %	10 %
Engine output	kW	4,400	3,740	3,300	2,200	1,100	440
Engine speed (FPP-curve)	rpm	2,000	1,895	1,817	1,587	1,260	928
Specific fuel oil consumption[1] [2]	g/kWh	199.0	196.0	195.0	190.0	201.0	305.0
Total fuel oil consumption[3]	l/h	1,047.0	876.0	769.0	500.0	265.0	161.0
Lube oil consumption[4]	g/kWh	0.11	-				

[1] Tolerance +5 %.

[2] Based on ISO reference conditions [according to ISO 15550:2002; ISO 3046:2002] and a lower calorific value of 42,700 kJ/kg and engine equipped with attached lube oil pump(s), fuel oil pump(s), HT- and LT cooling water pump(s).
Relevant for engine's certification for compliance with the NO_x limits according E3 Test cycle.

[3] Total fuel oil consumption [l/h] calculated based on above stated specific fuel oil consumption [g/kWh] and a density of 837 kg/m³.

[4] See accordingly section Lube oil consumption, Page 220.

Table 25: Marine mechanical propulsion light duty, 220 kW/cyl., 2,000 rpm, IMO Tier II

Reference conditions

	Units	ISO	Limit conditions[1]
Air temperature	°C	25	45
Seawater inlet temperature	°C	18	32
Air pressure[2]	mbar	1,000	
Exhaust back pressure[3]	mbar	50	
Relative humidity	%	30	60

[1] Please contact MAN Energy Solutions if project specific the limit conditions might be exceeded.

[2] Intake air depression up to 30 mbar allowed.

[3] Reference value for the difference pressure of exhaust gas line (plant) at MCR for IMO Tier II variant.

Table 26: Reference conditions – MAN 175D IMO Tier II

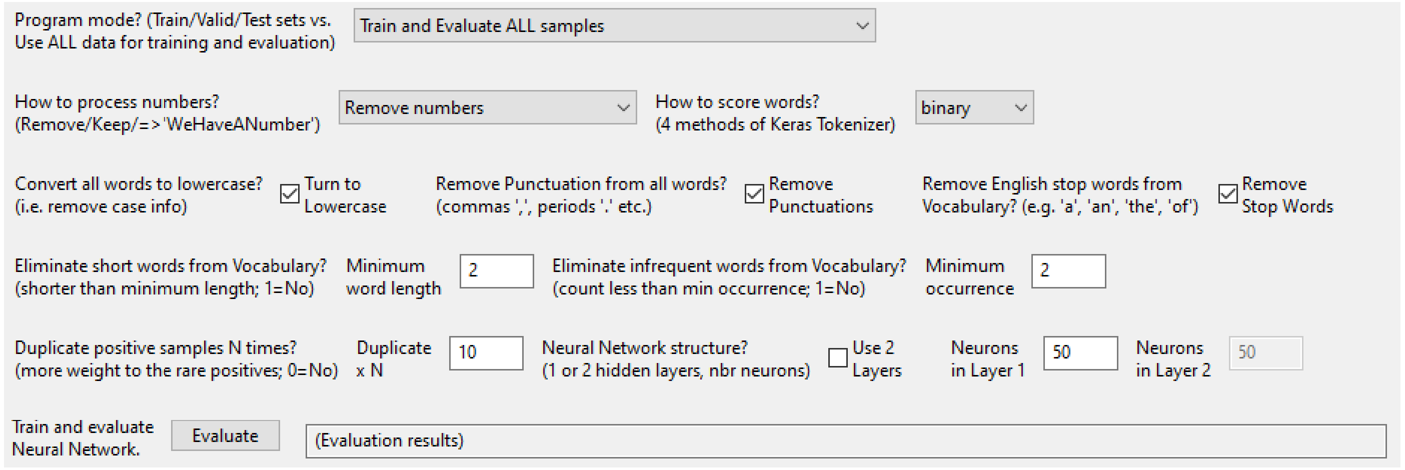
2021-12-10 - 7.0

MAN 175D IMO Tier II / IMO Tier III, Project Guide – Marine

```
Epoch 14/50
779/779 - 3s - loss: 0.0011 - accuracy: 0.9997 - 3s/epoch - 4ms/step
Epoch 15/50
779/779 - 3s - loss: 0.0013 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 16/50
779/779 - 3s - loss: 0.0012 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 17/50
779/779 - 3s - loss: 0.0011 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 18/50
779/779 - 3s - loss: 0.0010 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 19/50
779/779 - 3s - loss: 8.4875e-04 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 20/50
779/779 - 3s - loss: 0.0011 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 21/50
779/779 - 3s - loss: 7.3548e-04 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 22/50
779/779 - 3s - loss: 8.0903e-04 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 23/50
779/779 - 3s - loss: 7.1471e-04 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 24/50
779/779 - 3s - loss: 0.0013 - accuracy: 0.9996 - 3s/epoch - 3ms/step
Epoch 25/50
779/779 - 3s - loss: 0.0012 - accuracy: 0.9997 - 3s/epoch - 4ms/step
Epoch 26/50
779/779 - 3s - loss: 9.4563e-04 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 27/50
779/779 - 3s - loss: 9.4881e-04 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 28/50
779/779 - 3s - loss: 5.9910e-04 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 29/50
779/779 - 3s - loss: 6.4787e-04 - accuracy: 0.9999 - 3s/epoch - 3ms/step
Epoch 30/50
779/779 - 3s - loss: 6.0503e-04 - accuracy: 0.9999 - 3s/epoch - 3ms/step
Epoch 31/50
779/779 - 3s - loss: 4.1337e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 32/50
779/779 - 3s - loss: 5.3648e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 33/50
779/779 - 3s - loss: 6.3584e-04 - accuracy: 0.9999 - 3s/epoch - 3ms/step
Epoch 34/50
779/779 - 3s - loss: 4.4037e-04 - accuracy: 0.9999 - 3s/epoch - 3ms/step
Epoch 35/50
779/779 - 3s - loss: 0.0014 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 36/50
779/779 - 3s - loss: 5.7763e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 37/50
779/779 - 3s - loss: 6.0712e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 38/50
779/779 - 3s - loss: 3.2703e-04 - accuracy: 0.9999 - 3s/epoch - 3ms/step
Epoch 39/50
779/779 - 3s - loss: 6.1371e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 40/50
779/779 - 3s - loss: 2.7807e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 41/50
779/779 - 3s - loss: 7.6616e-04 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 42/50
779/779 - 3s - loss: 3.4721e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 43/50
779/779 - 3s - loss: 4.0565e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 44/50
779/779 - 3s - loss: 0.0013 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 45/50
779/779 - 3s - loss: 8.0889e-04 - accuracy: 0.9998 - 3s/epoch - 3ms/step
Epoch 46/50
779/779 - 3s - loss: 7.0557e-04 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Epoch 47/50
779/779 - 3s - loss: 5.2692e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 48/50
779/779 - 3s - loss: 5.1123e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 49/50
779/779 - 3s - loss: 3.3365e-04 - accuracy: 0.9999 - 3s/epoch - 4ms/step
Epoch 50/50
779/779 - 3s - loss: 8.6144e-04 - accuracy: 0.9998 - 3s/epoch - 4ms/step
Validation Accuracy: 99.988902 %
nbr_true_pos, nbr_fals_pos, nbr_true_neg, nbr_fals_neg: 690 2 17327 0
```

Neural network program, notes

- The NN program uses a rather simple NN, only one or two hidden layers, and Bag of Words model, ref. https://machinelearningmastery.com/deep-learning-bag-of-words-model-sentiment-analysis/
- There are many options to reduce the amount of text that must be processed; see check boxes and numerical entries in the screen dump below
- Program mode tells whether to divide the 18 019 input PDF pages randomly into Training (60 % of pages), Validation (20 %), and Test (20 %) samples, or to use all pages as Training, Validation, and Test samples. The previous mode ("Train/Valid/Test sets") is recommended, it is the standard method
- For number processing there are 3 alternatives: The words that are numbers can be removed, kept, or replaced with a fixed string 'WeHaveANumber'



Program mode? (Train/Valid/Test sets vs. Use ALL data for training and evaluation): Train and Evaluate ALL samples

How to process numbers? (Remove/Keep/=>'WeHaveANumber'): Remove numbers

How to score words? (4 methods of Keras Tokenizer): binary

Convert all words to lowercase? (i.e. remove case info) [✓ Turn to Lowercase]

Remove Punctuation from all words? (commas ',', periods '.' etc.) [✓ Remove Punctuations]

Remove English stop words from Vocabulary? (e.g. 'a', 'an', 'the', 'of') [✓ Remove Stop Words]

Eliminate short words from Vocabulary? (shorter than minimum length; 1=No) Minimum word length: 2

Eliminate infrequent words from Vocabulary? (count less than min occurrence; 1=No) Minimum occurrence: 2

Duplicate positive samples N times? (more weight to the rare positives; 0=No) Duplicate x N: 10

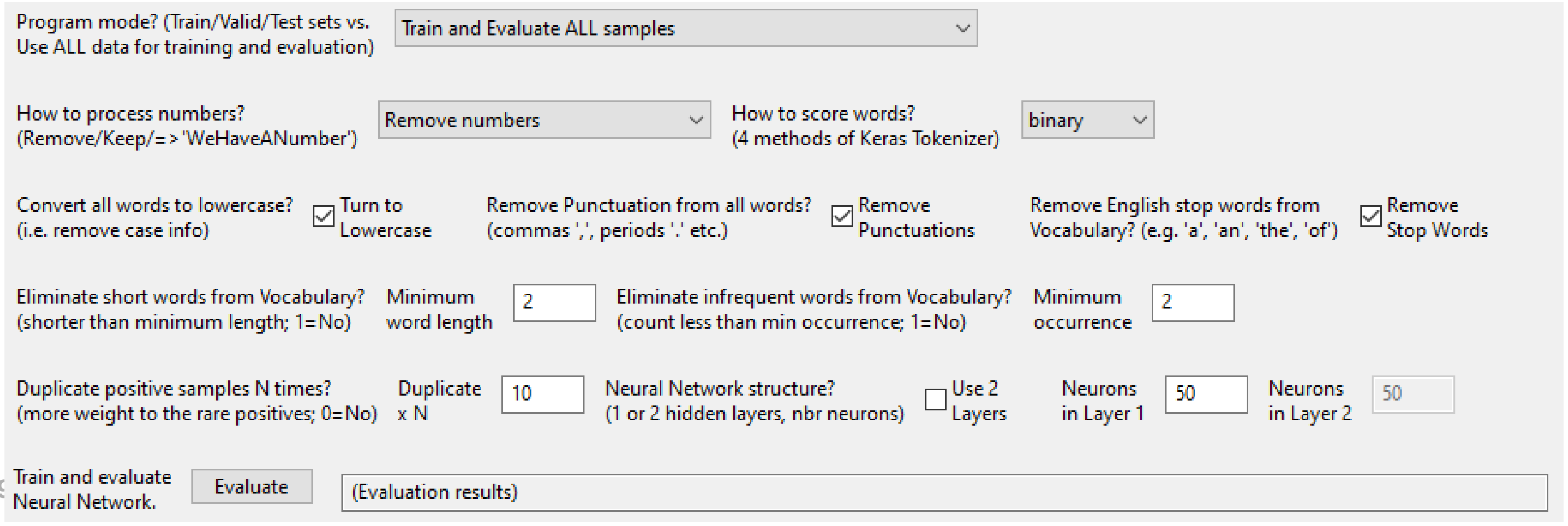
Neural Network structure? (1 or 2 hidden layers, nbr neurons) ☐ Use 2 Layers; Neurons in Layer 1: 50; Neurons in Layer 2: 50

Train and evaluate Neural Network. Evaluate

(Evaluation results)

Neural network program, notes

- Keras Tokenizer has 4 different modes, when text entries are converted into number vectors (vector size = vocabulary size): Modes ‘binary, ‘count’, ‘tfidf’ and ‘freq’
- Usually data sets have an equal number of positive and negative entries. However, in our data there are much less relevant/positive (=Ground Truth) pages than irrelevant/negative pages. We have only 690 relevant pages of 18 019 pages, less than 4 %. The simplest way to handle this imbalanced data set is to duplicate the positive samples in the Training set
- Different numbers of hidden neurons can be tried, as well as 1 or 2 hidden layers



Program mode? (Train/Valid/Test sets vs. Use ALL data for training and evaluation): Train and Evaluate ALL samples

How to process numbers? (Remove/Keep/=> 'WeHaveANumber'): Remove numbers
How to score words? (4 methods of Keras Tokenizer): binary

Convert all words to lowercase? (i.e. remove case info): ✓ Turn to Lowercase
Remove Punctuation from all words? (commas ',', periods '.' etc.): ✓ Remove Punctuations
Remove English stop words from Vocabulary? (e.g. 'a', 'an', 'the', 'of'): ✓ Remove Stop Words

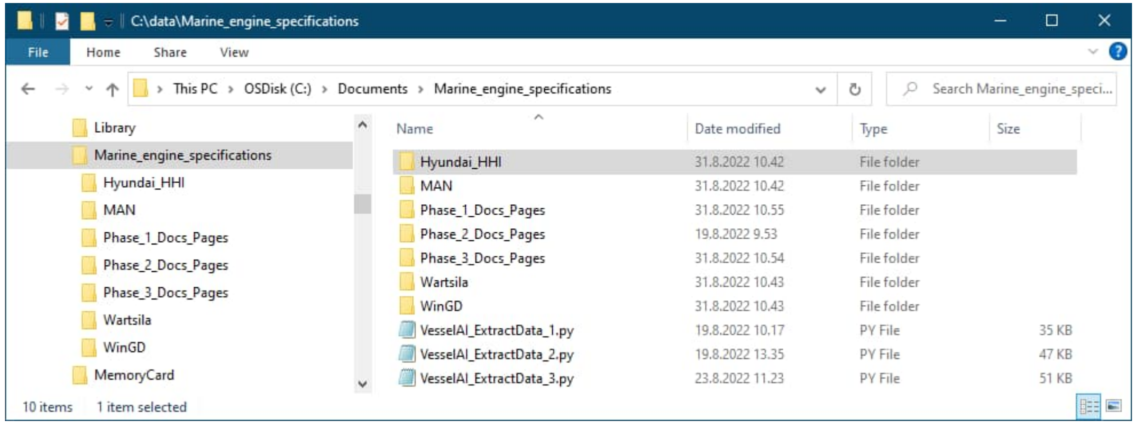
Eliminate short words from Vocabulary? (shorter than minimum length; 1=No): Minimum word length 2
Eliminate infrequent words from Vocabulary? (count less than min occurrence; 1=No): Minimum occurrence 2

Duplicate positive samples N times? (more weight to the rare positives; 0=No): Duplicate x N 10
Neural Network structure? (1 or 2 hidden layers, nbr neurons): ☐ Use 2 Layers; Neurons in Layer 1: 50; Neurons in Layer 2: 50

Train and evaluate Neural Network. Evaluate
(Evaluation results)

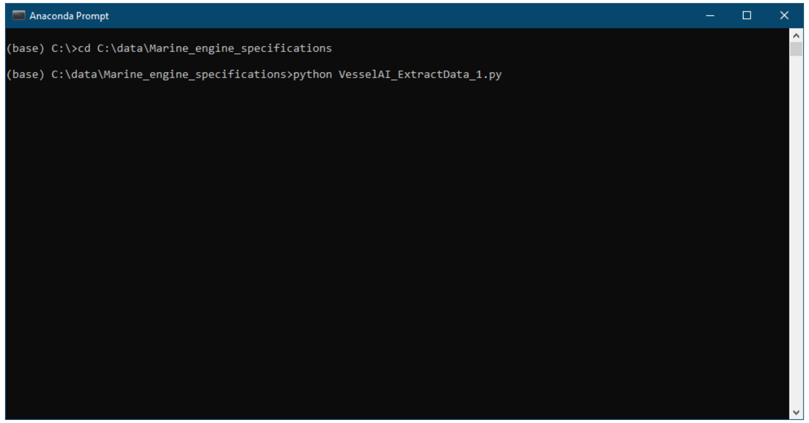
Installation

- The *.zip file contains a folder ‘Marine_engine_specifications’
- This folder contains 4 Python files:
  - VesselAI_ExtractData_1.py
  - VesselAI_ExtractData_2.py
  - VesselAI_ExtractData_3.py
  - VesselAI_ExtractData_NN.py
- …and also subfolders for PDF files (Hyundai_HHI, MAN, Wartsila, WinGD) as well as empty subfolders where result data will be written (Phase_?_Docs_Pages)
- The PDF files must be downloaded from Internet, see slides 5-9



C:\data\Marine_engine_specifications
File
Home
Share
View
This PC › OSDisk (C:) › Documents › Marine_engine_specifications
Search Marine_engine_speci...
Library
Marine_engine_specifications
Hyundai_HHI
MAN
Phase_1_Docs_Pages
Phase_2_Docs_Pages
Phase_3_Docs_Pages
Wartsila
WinGD
MemoryCard
10 items
1 item selected

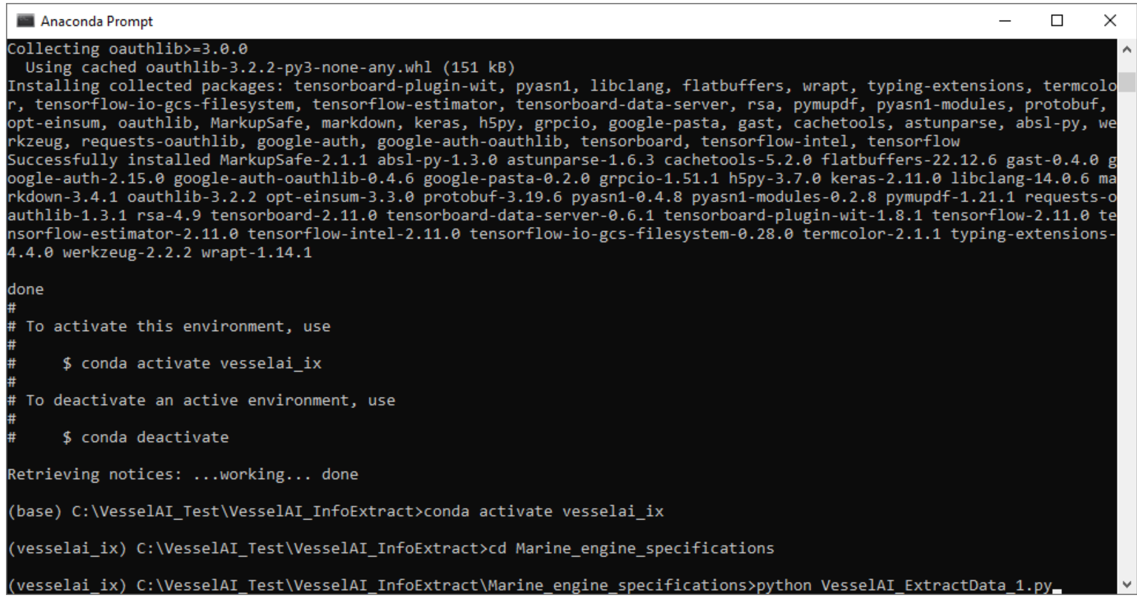
Name	Date modified	Type	Size
Hyundai_HHI	31.8.2022 10.42	File folder	
MAN	31.8.2022 10.42	File folder	
Phase_1_Docs_Pages	31.8.2022 10.55	File folder	
Phase_2_Docs_Pages	19.8.2022 9.53	File folder	
Phase_3_Docs_Pages	31.8.2022 10.54	File folder	
Wartsila	31.8.2022 10.43	File folder	
WinGD	31.8.2022 10.43	File folder	
VesselAI_ExtractData_1.py	19.8.2022 10.17	PY File	35 KB
VesselAI_ExtractData_2.py	19.8.2022 13.35	PY File	47 KB
VesselAI_ExtractData_3.py	23.8.2022 11.23	PY File	51 KB



```
(base) C:\>cd C:\data\Marine_engine_specifications

(base) C:\data\Marine_engine_specifications>python VesselAI_ExtractData_1.py
```

Installation, continued

- Extract the contents of the *.zip file into a local folder, and then go to the folder ‘VesselAI_InfoExtract’ (inside the new folder) with your Python Prompt
- In Anaconda Python environment, enter command:
  - conda env create -f env.yml
- This will create Anaconda environment ‘vesselai_ix’, containing the necessary Python program libraries; then activate this new environment:
  - conda activate vesselai_ix
- Go to subfolder ‘Marine_engine_specifications’
  - cd Marine_engine_specifications
- Remember that subfolders ‘MAN’, ‘Wartsila’, ‘Hyundai_HHI’, and ‘WinGD’ must contain PDF files you have downloaded, as instructed earlier
- Start the Phase 1 program, *_1.py
  - python Vesselai_ExtractData_1.py


Anaconda Prompt

```
Collecting oauthlib>=3.0.0
  Using cached oauthlib-3.2.2-py3-none-any.whl (151 kB)
Installing collected packages: tensorboard-plugin-wit, pyasn1, libclang, flatbuffers, wrapt, typing-extensions, termcolo
r, tensorflow-io-gcs-filesystem, tensorflow-estimator, tensorboard-data-server, rsa, pymupdf, pyasn1-modules, protobuf,
opt-einsum, oauthlib, MarkupSafe, markdown, keras, h5py, grpcio, google-pasta, gast, cachetools, astunparse, absl-py, we
rkzeug, requests-oauthlib, google-auth, google-auth-oauthlib, tensorboard, tensorflow-intel, tensorflow
Successfully installed MarkupSafe-2.1.1 absl-py-1.3.0 astunparse-1.6.3 cachetools-5.2.0 flatbuffers-22.12.6 gast-0.4.0 g
oogle-auth-2.15.0 google-auth-oauthlib-0.4.6 google-pasta-0.2.0 grpcio-1.51.1 h5py-3.7.0 keras-2.11.0 libclang-14.0.6 ma
rkdown-3.4.1 oauthlib-3.2.2 opt-einsum-3.3.0 protobuf-3.19.6 pyasn1-0.4.8 pyasn1-modules-0.2.8 pymupdf-1.21.1 requests-o
authlib-1.3.1 rsa-4.9 tensorboard-2.11.0 tensorboard-data-server-0.6.1 tensorboard-plugin-wit-1.8.1 tensorflow-2.11.0 te
nsorflow-estimator-2.11.0 tensorflow-intel-2.11.0 tensorflow-io-gcs-filesystem-0.28.0 termcolor-2.1.1 typing-extensions-
4.4.0 werkzeug-2.2.2 wrapt-1.14.1

done
#
# To activate this environment, use
#
#     $ conda activate vesselai_ix
#
# To deactivate an active environment, use
#
#     $ conda deactivate

Retrieving notices: ...working... done

(base) C:\VesselAI_Test\VesselAI_InfoExtract>conda activate vesselai_ix

(vesselai_ix) C:\VesselAI_Test\VesselAI_InfoExtract>cd Marine_engine_specifications

(vesselai_ix) C:\VesselAI_Test\VesselAI_InfoExtract\Marine_engine_specifications>python VesselAI_ExtractData_1.py
```

Installation, continued

- If you do not have the Anaconda Python, this will probably fail at first, because you have not installed all the necessary libraries to your Python environment
- But when starting the python program fails, the name of the missing library is shown. Then install this library (pip install <name_of_lib>). Continue until *_1.py starts
- Close this *_1.py immediately. Then try *_2.py and *_3.py. Continue installing missing libraries until all 3 programs start
- Then try to go through the whole workflow with default values: start *_1.py:
  - >python VesselAI_ExtractData_1.py
  - *_Phase_1: Maximize, move vertical slider right, press ‘Refresh’, (wait…), press ‘Next >>’
  - *_Phase_2: Maximize, move vertical slider right, press ‘Precomp’, (wait…), select algorithm ‘DBSCAN’ from the drop-down, press ‘execute’, press ‘Next >>’
  - *_Phase_3: Maximize, move vertical slider right. Next you should enter relevant line definitions for each cluster, the press ‘Refresh’, (wait…), press ‘Results’.
  - => However, because this is tedious (DBSCAN produces 41 clusters!), delete all the files in folder ‘Phase_3_Docs_Pages’, and then start ‘VesselAI_ExtractData_3.py’ directly. And now:
  - *_Phase_3: Maximize, move vertical slider right, press ‘Refresh’, (wait…), press ‘Results’
  - And now Notepad shows file ‘Phase3_key_results.txt’, containing relevant data lines only!