Extract Data from a PDF File using C# and .net core

Problem Description

Commercial pilots receive a flight planning document in PDF format. One PDF can contain information about one or more flights but each flight data is formatted in the same way. The data is presented in different chapters and each chapter contains data for each flight, i.e. flights are repeated in the chapters. The files have many more pages, which are irrelevant for this task and should be ignored.

The solution should be able to identify certain fields and retrieve the data from them. The result should be a list of flight data, with the data for each flight collected in an object.

Task

Design and develop a module that meets the requirements below. Since this is an assignment to show your working style, focus on showing how you would approach the problem, architect the solution, organize the code, and implement enough to show your coding skills. It is okay to leave parts unfinished as long as you have solved the "how".

Feel free to define the data structures as you see fit for the solution. You can then explain to the team why you went in that direction.

Requirements

The fields to be extracted are pre-defined, i.e. the solution only has to retrieve pre-defined information and does not receive any input from the user as to what data needs to be retrieved. However, it should be easy to extend the solution by adding new fields to the definition.

The PDF files are generated by external systems. They can be changed without notice, corrupted or rearranged files can be delivered. Errors must be handled appropriately and made available to the caller of the code for further processing, providing enough data to understand the issue and be able to inform the user about it.

The code must be structured in a way that allows it to be extended, integrated with other systems, and adequately tested.

A sample console application should be provided to demonstrate the usage. However, the functionality should be prepared for integration into a larger project and structured accordingly. Plan any interfaces accordingly.

The solution should be written in the latest version of .net core.

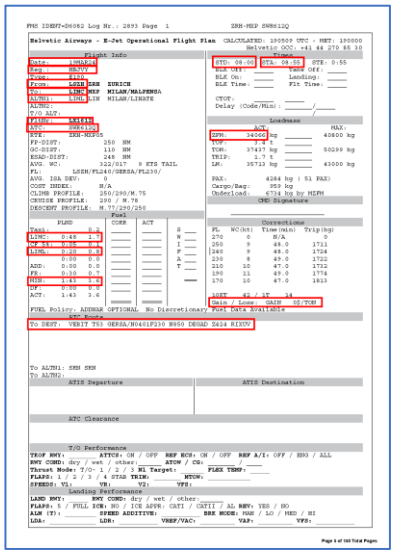
Sample Data

A sample file for development is attached. It contains data for four flights.

Below is an excerpt from the provided sample file showing which fields from the OFP and Crew Briefing chapters need to be extracted for the first flight in the file. It is sufficient to extract only exactly these fields for each flight in the file.

Operational Flight Plan

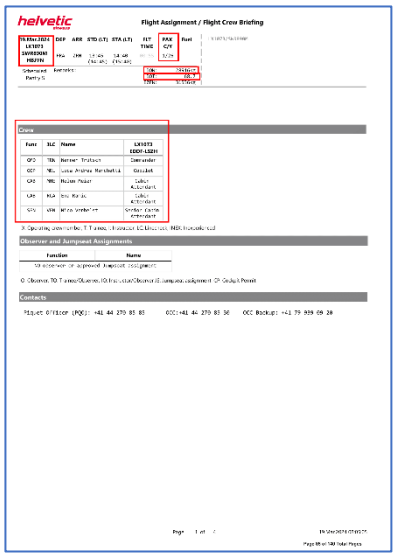
This is the OFP page for the first flight of the sample file. A copy with a better resolution and marked fields can be found at the end of this document.



- Date: 2024-03-19
- Aircraft registration: HBJVY
- Route: from LSZH to LIMC
- Alternate airdrome 1: LIML
- Alternate airdrome 2: none
- Flight number: LX1612
- ATC call sign: SWR612Q
- Departure time: 08:00
- Arrival time: 08:55
- Zero fuel mass (ZFM): 34066
- Time to destination: 0:48
- Fuel to destination: 1.7
- Time to alternate: 0:20
- Fuel to alternate: 0.8
- Minimum fuel required: 3.6
- Route first and last navigation point: VEBIT – RIXUV
- Gain/loss: +0 (positive number for gain, negative number for loss)

Crew Briefing

This is the crew briefing page for the first flight of the sample file. A copy with a better resolution and marked fields can be found at the end of this document.



- Number of passengers in business (C) class: 1
- Number of passengers in economy (Y) class: 26
- Dry operating weight (DOW): 28916
- Dry operating index: 68.7
- Crew and functions (list, number may vary):
  - Werner Trütsch, CMD
  - Luca Andrea Marchetti, COP
  - Helen Meier, CAB
  - Ena Ramic, CAB
  - Nico Veheist, SEN

Tools and Resources to Use

Consider this a normal task from your job. This means that any tools and resources can be used during development. Since this is not currently part of a larger project, feel free to choose any libraries you know and feel comfortable with for IoC, unit testing, mocking, etc.

If you decide to use a library and have a choice between two, one free and one paid, but both can do what you need, choose the free one. If only the paid one has the needed functionality, ask the team if using it is an option.

Be a Team Member, Be Active, Make a Difference

We at capzlog.aero are your team! Use the team as your resource. In this exercise we also play the role of product owner and customer.

If you find that something is unclear and you cannot make good assumptions, ask for clarification.

If you have questions or suggestions, do not hesitate to ask or express them - be part of the team and expect the team to support you.

Readable Copies of File

Operational Flight Plan

FMS IDENT=D6082 Log Nr.: 2893 Page 1 ZRH-MXP SWR612Q

Helvetic Airways - E-Jet Operational Flight Plan CALCULATED: 190509 UTC - MET: 190000
Helvetic OCC: +41 44 270 85 30

Flight Info

Date: 19MAR24
Reg.: HBJVY
Type: E190
From: **LSZH ZRH ZURICH**
To: **LIMC MXP MILAN/MALPENSA**
ALTN1: LIML LIN MILAN/LINATE
ALTN2:
T/O ALT:
FltNr: **LX1612**
ATC: SWR612Q
RTE: ZRH-MXP05
FP-DIST: 250 NM
GC-DIST: 110 NM
ESAD-DIST: 248 NM
AVG. WC: 322/017 9 KTS TAIL
FL: LSZH/FL240/GERSA/FL230/
AVG. ISA DEV: 0
COST INDEX: N/A
CLIMB PROFILE: 250/290/M.75
CRUISE PROFILE: 290 / M.78
DESCENT PROFILE: M.77/290/250

Times

STD: 08:00 STA: 08:55 STE: 0:55
BLK Off: _____ Take Off: _____
BLK On: _____ Landing: _____
BLK Time: _____ Flt Time: _____

CTOT: _____ _____
Delay (Code/Min): _____/_____
/

Loadmass

	ACT:		MAX:
ZFM:	34066 kg	_____	40800 kg
TOF:	3.4 t	_____	
TOM:	37437 kg	_____	50299 kg
TRIP:	1.7 t	_____	
LM:	35713 kg	_____	43000 kg

PAX: 4284 kg (51 PAX)
Cargo/Bag: 959 kg
Underload: 6734 kg by MZFM

CMD Signature

_____

Fuel

	PLND		CORR	ACT	
Taxi:		0.2	_____	_____	S ___
LIMC:	0:48	1.7	_____	_____	W ___
CF 5%:	0:05	0.1	_____	_____	I ___
LIML:	0:20	0.8	_____	_____	F ___
	0:00	0.0	_____	_____	A ___
ADD:	0:00	0.0	_____	_____	T ___
FR:	0:30	0.7	_____	_____	
MIN:	1:43	3.6	_____	_____	====
DF:	0:00	0.0	_____	_____	
ACT:	1:43	3.6	====	====	

Corrections

FL	WC(kt)	Time(min)	Trip(kg)
270	0	N/A	0
250	9	48.0	1711
240	9	48.0	1724
230	8	49.0	1722
210	10	47.0	1732
190	11	49.0	1774
170	10	47.0	1813

10KT 42 / 1T 14
Gain / Loss: GAIN 0$/TON

FUEL Policy: ADDNAR OPTIONAL No Discretionary Fuel Data Available

ATC Route

To DEST: VEBIT T53 GERSA/N0401F230 N850 DEGAD Z424 RIXUV

To ALTN1: SRN SRN
To ALTN2:

ATIS Departure

ATIS Destination

ATC Clearance

T/O Performance

TKOF RWY: _____ **ATTCS:** ON / OFF **REF ECS:** ON / OFF **REF A/I:** OFF / ENG / ALL
RWY COND: dry / wet / other:_____ **ATOW / CG:** _____ / ____
Thrust Mode: T/O- 1 / 2 / 3 **N1 Target:** _____ **FLEX TEMP:** _____
FLAPS: 1 / 2 / 3 / 4 STAB **TRIM:** _____ **MTOW:** _____
SPEEDS: V1: **VR:** **V2:** **VFS:**

Landing Performance

LAND RWY: _____ **RWY COND:** dry / wet / other:_____
FLAPS: 5 / FULL **ICE:** NO / ICE APPR: CATI / CATII / AL **REV:** YES / NO
ALM (T): _____ **SPEED ADDITIVE:** _____ **BRK MODE:** MAN / LO / MED / HI
LDA: _____ **LDR:** _____ **VREF/VAC:** _____ **VAP:** _____ **VFS:** _____

Crew Briefing

helvetic airways

Flight Assignment / Flight Crew Briefing

19.Mar.2024 LX1073 SWR890M HBJVN	DEP	ARR	STD (LT)	STA (LT)	FLT TIME	PAX C/Y	Fuel
	FRA	ZRH	13:45 (14:45)	14:40 (15:40)	00:55	1/26	

LX1073/SWR890M

Scheduled
Pantry S

Remarks:

DOW: 28916kg
DOI: 68.7
EZFW: 31536kg

Crew

Func	3LC	Name	LX1073 EDDF-LSZH
CMD	TRW	Werner Trütsch	Commander
COP	MRL	Luca Andrea Marchetti	Copilot
CAB	MHE	Helen Meier	Cabin Attendant
CAB	RCA	Ena Ramic	Cabin Attendant
SEN	VEN	Nico Verhelst	Senior Cabin Attendant

X: Operating crewmember, T: Trainee, I: Instructor, LC: Linecheck, INEX: Inexperienced

Observer and Jumpseat Assignments

Function	Name
NO observer or approved Jumpseat assignment	

O: Observer, TO: Trainee/Observer, IO: Instructor/Observer JS: Jumpseat assignment, CP: Cockpit Permit

Contacts

Piquet Officer (PQO): +41 44 270 85 85 OCC:+41 44 270 85 30 OCC Backup: +41 79 939 09 20