



**ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE**

**FAKULTA DOPRAVNÍ**

Ústav letecké dopravy

**Textová analýza nestrukturovaných  
závadových dat v letecké údržbě**

Diplomová práce

Bc. Tomáš Vojtěch

Vedoucí práce: Ing. Andrej Lališ, Ph.D., Ing. Martin Kála

Praha 2021



**K621** ..... **Ústav letecké dopravy**

## **ZADÁNÍ DIPLOMOVÉ PRÁCE** (PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

**Bc. Tomáš Vojtěch**

Kód studijního programu a studijní obor studenta:

**N 3710 – PL – Provoz a řízení letecké dopravy**

Název tématu (česky): **Textová analýza nestrukturovaných závadových dat v letecké údržbě**

Název tématu (anglicky): Text analysis of non-structured data from aircraft maintenance findings

### **Zásady pro vypracování**

Při zpracování diplomové práce se řiďte následujícími pokyny:

- Cíl práce: Návrh pro textovou analýzu a vyhodnocení nestrukturovaných dat za účelem identifikace neočekávaných nálezů v údržbě dopravních letadel
- Analyzujte současné postupy pro záznam nálezů v procesu údržby dopravních letadel
- Identifikujte a analyzujte dostupné slovníky pro klasifikaci nálezů z údržby
- Proveďte analýzu a vyhodnocení dostupných dat z letecké údržbové organizace
- Vytvořte slovník pro provádění textové analýzy a navrhněte způsob vyhodnocení nestrukturovaných dat s ohledem na nálezy v letecké údržbě
- Vytvořené řešení ověřte a vyhodnoťte



- Rozsah grafických prací: dle pokynů vedoucího diplomové práce
- Rozsah průvodní zprávy: minimálně 55 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)
- Seznam odborné literatury: R. Doc Palmer. Maintenance Planning and Scheduling Handbook, McGraw Hill Professional, 1999.  
Kutner, M. H., Nachtsheim, Ch. J., Neter, J. and Wasserman, W. Applied Linear Statistical Models, McGraw-Hill Companies, 2005.

Vedoucí diplomové práce: **Ing. Andrej Lališ, Ph.D.**  
**Ing. Martin Kála**

Datum zadání diplomové práce: **17. července 2020**  
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání diplomové práce: **17. května 2021**  
a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia  
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia

doc. Ing. Jakub Kraus, Ph.D.  
vedoucí  
Ústavu letecké dopravy



doc. Ing. Pavel Hrubeš, Ph.D.  
děkan fakulty

Potvrzuji převzetí zadání diplomové práce.

Bc. Tomáš Vojtěch  
Jméno a podpis studenta

V Praze dne.....17. července 2020

## Poděkování

Na tomto místě bych rád poděkoval všem, kteří mi poskytli podklady pro vypracování této závěrečné práce. Zvláště bych rád poděkoval panu Ing. Andrejovi Lališovi, Ph.D. a Ing. Martinovi Kálovi za odborné vedení a konzultování diplomové práce a za rady, které mi poskytovali po celou dobu její tvorby. Dále také patří můj vděk společnosti CSAT, která mi umožnila přístup k mnoha důležitým informacím a materiálům, které mi pomohly při psaní této práce. V neposlední řadě je mou povinností také poděkovat svým rodičům a blízkým za morální a materiální podporu, které se mi dostávalo po celou dobu studia.

## Prohlášení

Nemám závažný důvod proti užívání tohoto školního díla ve smyslu § 60 Zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon).

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 17.května 2021



Tomáš Vojtěch

Autor: Bc. Tomáš Vojtěch  
Název: Textová analýza nestrukturovaných závadových dat v letecké údržbě  
Instituce: České vysoké učení technické v Praze, Fakulta dopravní  
Obor: Provoz a řízení letecké dopravy  
Rok: 2021

### **Abstrakt**

Údržba letadel na technické základně v sobě zahrnuje plánované, ale i neplánované opravy. Při provádění oprav dochází k podrobným prohlídkám letadel, které jsou předem objednané zákazníkem. Při těchto prohlídkách je možné objevit jakýkoliv další nevyhovující stav letadla z hlediska zachování letové způsobilosti, který je nutný napravit. Tyto neplánované závady jsou nazývány jako nálezy závad a jsou zavedeny do systému údržbové organizace, které jsou povinny vést záznamy o činnostech provedených na letadlech. Popis nálezu závad vytvářejí mechanici při provádění údržby a většinou mají podobu nestrukturovaného textu. Tato práce se zabývá možností využití automatické analýzy těchto nálezů závad za pomoci softwarového nástroje umožňující textovou analýzu. Pro potřeby této analýzy byl manuálně vytvořen definiční slovník pojmů, který obsahuje komponenty a závady objevující se v záznamech údržby. Tento slovník byl následně použit při automatické textové analýze, kde výstupem je soubor obsahující nalezené pojmy v jednotlivých záznamech, ze kterého může být sestavena statistika nejčastěji se vyskytujících komponentů a závad. Je předpokladem, že tuto statistiku bude možné využít při optimalizaci plánování údržby a snadnějšímu odhalování neplánovaných závad na letadle.

### **Klíčová slova**

anotace pojmů, komponenty a závady, nálezy závad, nástroje textové analýzy, neplánovaná údržba, plánování údržby, slovník pojmů údržby letadel, textová analýza, údržba letadel, záznamy údržby

Author: Bc. Tomáš Vojtěch  
Title: Text analysis of non-structured data from aircraft maintenance findings  
Institution: Czech technical university in Prague, Faculty of Transportation  
Sciences  
Study Program: Air Traffic Management  
Year: 2021

### **Abstract**

Aircraft maintenance at the technical base includes both scheduled and unscheduled repairs. When performing repairs, there are detailed inspections of aircraft, which are ordered in advance by the customer. During these inspections, it is possible to detect any other unsatisfactory condition of the aircraft in terms of continuing airworthiness, which needs to be corrected. These unplanned failures are called findings and are entered into the maintenance organization's system, which is required to keep records of activities performed on the aircraft. Finding descriptions are created by mechanics during maintenance and are usually in the form of unstructured text. This diploma thesis is about the possibility of using automatic text analysis of these findings using a dedicated software tool. For the purposes of this analysis, a definition dictionary was manually created, which contains components and failures appearing in maintenance records. This dictionary was then used in automatic text analysis, where the output is a file containing the terms found in the individual records, from which statistics of the most common components and failures can be compiled. It is assumed that this statistic can be used for maintenance planning optimization and easier detection of unplanned aircraft failures.

### **Keywords**

aircraft maintenance, aircraft maintenance dictionary, components and failures, maintenance findings, maintenance planning, maintenance records, text analysis, text analysis tools, unscheduled maintenance, words annotation

## Obsah

Seznam obrázků .....	8
Seznam tabulek.....	9
Seznam grafů .....	9
Seznam použitých zkratk.....	10
1. Úvod.....	12
2. Údržba letadel .....	14
2.1 Nařízení Komise (EU) č. 1321/2014 .....	15
2.1.1 Part M – Požadavky na zachování letové způsobilosti.....	16
2.1.2 Part 145 - Oprávnění organizace k údržbě .....	16
2.1.3 Part 66 - Osvědčující personál.....	17
2.1.4 Part 147 - Požadavky na výcvikové organizace.....	17
2.2 Společnost Czech Airlines Technics .....	17
3. Proces údržby letadla a záznamu nálezu závad .....	20
3.1 Technologické karty.....	20
3.2 Software využívaný údržbovými organizacemi .....	21
3.2.1 AMOS.....	21
3.3 Pracovní příkaz .....	22
3.4 Vznik záznamu nálezu závady.....	25
4. Analýza existujících slovníků .....	26
4.1 Klasifikace nálezů závad z údržby .....	26
4.1.1 Norma ATA 100.....	27
4.1.2 Průvodní technická dokumentace .....	28
4.2 Textová analýza a její využití .....	29
4.2.1 Dostupná vědecká literatura .....	30
4.3 Limitace současného stavu .....	31
5. Nástroje umožňující textovou analýzu .....	33
5.1 TermIt.....	33
5.2 Brat .....	34
5.3 Gate .....	35
5.4 Zhodnocení a volba vhodného nástroje.....	36
6. Zpracování dostupných dat .....	37
6.1 Dostupná data od společnosti CSAT .....	37
6.2 Filtrace a úprava dat.....	38
6.3 Práce s nástrojem TermIt .....	39
6.3.1 Uživatelské prostředí TermIt.....	39

6.3.2 Slovníky a pojmy .....	40
6.3.3 Import dat do nástroje.....	41
6.4 Anotace .....	42
6.4.1 Podmínky pro vytváření pojmů .....	43
7. Slovník komponent a závad.....	45
7.1 Vznik slovníku .....	45
7.2 Komponenty .....	46
7.3 Závady .....	47
7.4 Použitelnost a rozsah zpracování .....	48
8. Provedení a vyhodnocení automatické textové analýzy.....	51
8.1 Spuštění automatické textové analýzy.....	51
8.2 Formy výstupu dokončené textové analýzy .....	52
8.2.1 Zobrazení seznamu výskytu pojmů.....	52
8.2.2 Vizualní reprezentace dat .....	52
8.2.3 Export z databáze nástroje TermIt.....	53
8.3 Ověření správnosti textové analýzy testovací množiny dat .....	55
8.3.1 Porovnání manuální anotace s výsledkem textové analýzy .....	55
8.3.2 Zhodnocení úspěšnosti automatické textové analýzy .....	57
8.4 Ověření textové analýzy na nezávislé sadě dat .....	60
8.4.1 Zhodnocení úspěšnosti automatické textové analýzy .....	60
8.4.2 Porovnání výsledků obou textových analýz .....	63
8.5 Limitace současného stavu automatické textové analýzy .....	64
8.5.1 Označení nesprávné komponenty nebo závady .....	64
8.5.2 Označení více komponent a závad v jednom záznamu .....	65
8.5.3 Uživatelská přívětivost nástroje .....	66
8.6 Možnosti vyhodnocení výstupu automatické textové analýzy .....	67
8.6.1 Tabulkový editor – kontingenční graf .....	67
8.6.2 Databázové nástroje – relační databáze a dotazy .....	69
9. Diskuse .....	71
10. Závěr .....	75
Seznam příloh .....	77
Zdroje.....	78



## Seznam obrázků

Obrázek 1 - CSAT Hangár F.....	18
Obrázek 2 - Vznik MPD .....	20
Obrázek 3 - Software AMOS [12].....	22
Obrázek 4 – AMOS - CSAT Workorder .....	24
Obrázek 5 - Číslování normy ATA 100.....	27
Obrázek 6 - Přenosné zařízení při údržbě [18].....	32
Obrázek 7 - Nástroj TermIt.....	34
Obrázek 8 - Nástroj Brat [20] .....	34
Obrázek 9 - Nástroj Gate [22].....	35
Obrázek 10 - Filtrovaná a upravená data .....	38
Obrázek 11 – Úvodní obrazovka nástroje TermIt.....	39
Obrázek 12 - Vybraný slovník v nástroji TermIt.....	41
Obrázek 13 - Anotace pojmů v dokumentu .....	43
Obrázek 14 - Slovník komponent a závad v nástroji TermIt .....	46
Obrázek 15 - Pojem Component v nástroji TermIt .....	47
Obrázek 16 - Pojem Failure v nástroji TermIt .....	48
Obrázek 17 - Slovník komponent a závad.....	49
Obrázek 18 - Spuštění automatické textové analýzy.....	51
Obrázek 19 - Výskyt pojmů v dokumentu .....	52
Obrázek 20 - Označení nalezených pojmů .....	53
Obrázek 21 - Výstupní soubor z databáze .....	54
Obrázek 22 - Kontingenční tabulka technologických karet (Komponenty a závady).....	69
Obrázek 23 – Dotaz v databázovém nástroji MS Access .....	70

## Seznam tabulek

Tabulka 1 - Příklady pojmů v nástroji TermIt.....	44
Tabulka 2 - Správně nalezené komponenty a závady.....	56
Tabulka 3 - Nesprávně nalezené komponenty a závady .....	56
Tabulka 4 - Nerelevantní záznamy .....	57
Tabulka 5 - Úspěšnost automatické textové analýzy – testovací množina .....	58
Tabulka 6 - Úspěšnost automatické textové analýzy – nezávislá sada dat .....	61
Tabulka 7 - Porovnání výsledků obou textových analýz .....	63
Tabulka 8 - Nesprávné označení pojmu "Missing" .....	65
Tabulka 9 - Označení více komponent a závad v jednom záznamu .....	65

## Seznam grafů

Graf 1 - Rozdělení záznamů dle výsledku textové analýzy – testovací množina dat .....	59
Graf 2 - Hladina úspěšnosti označování pojmů – testovací množina dat.....	59
Graf 3 - Rozdělení záznamů dle výsledku textové analýzy – nezávislá sada dat.....	62
Graf 4 - Hladina úspěšnosti označování pojmů – nezávislá sada dat.....	62

## Seznam použitých zkratk

AD	Airworthiness Directive
AMM	Aircraft Maintenance Manual
AMO	Approved Maintenance Organisation
ATA	Air Transport Association
AWL	Airworthiness Limitations
CAMO	Continuing Airworthiness Management Organisation
CDL	Configuration Deviation List
CMM	Component Maintenance Manual
CMR	Certification Maintenance Requirements
CSAT	Czech Airline Technics
DDG	Dispatch Deviation Guide
EASA	European Aviation Safety Agency
FIM	Fault Isolation Manual
ICAO	International Civil Aviation Organization
IPC	Illustrated Part Catalog
ITEM	Illustrated Tool & Equipment Manual
MPD	Maintenance Planning Document
MRBR	Maintenance Review Board Report
MRO	Maintenance, Repair, Overhaul
MTO	Maintenance Training Organization
NDTM	Non-destructive Testing Manual
NLP	Natural Language Processing
OAMP	Operator Approved Maintenance Program
OHM	Overhaul Manual
SB	Service Bulletin
SL	Service Letter

SRM	Structural Repair Manual
SSM	System Schematic Manual
TC	Task Card
WDM	Wiring Diagram Manual
WO	Workorder

# 1. Úvod

Pravidelná údržba letadel je jedním z nejdůležitějších procesů k zajištění vysoké úrovně bezpečnosti provozu letecké dopravy jako známe dnes. Výrobce stanovené kontrolní prohlídky a intervaly představují závaznou podmínku pro zachování letové způsobilosti a zajištění spolehlivosti a dlouhé životnosti letadel. Vzhledem k jejich nezanedbatelné pořizovací ceně je předpokladem, že letadla budou udržována v provozu schopném stavu déle, než jako známe například u osobních automobilů, aby se jejich provoz dlouhodobě a ekonomicky vyplatil.

Údržba letadel se rozděluje na dva druhy podle rozsahu a množství údržbových prací. Jedná se o traťovou údržbu a o údržbu na technické základně, také označovanou jako těžká údržba. Při traťové údržbě letadlo prochází pouze jednoduchými kontrolami, které nevyžadují rozsáhlejší zásah do jeho systému a konstrukce. Naopak při těžké údržbě letadla už dochází k jeho částečnému, někdy až úplnému rozebrání, aby byl zajištěn neomezený přístup ke kontrole jeho strukturální integrity. Tento typ údržby představuje provádění rozsáhlejších a časově náročnějších údržbových prací s podrobnou kontrolou jednotlivých systémů a vybavení.

Údržbu letadel výhradně provádějí schválené údržbové organizace s příslušnými povoleními pro údržbu daného typu letadel. Musí disponovat vyhovujícími prostory, kvalifikovaným personálem a patřičným vybavením. Organizace pravidelně procházejí kontrolními audity, zda splňují všechny požadavky a náležitosti, aby nedošlo k odebrání jejich oprávnění.

Na základě mezinárodních úmluv, standardů a doporučení mezinárodní organizace pro civilní letectví ICAO a dalších vládních nařízení jsou všechny údržbové organizace povinny nepřetržitě vést záznamy o veškeré provedené údržbě a vykonaných pracích na letadle.

Tato diplomová práce je zaměřena na analýzu údržbových záznamů, respektive zabývá se návrhem možnosti jejich zpracování za pomoci automatické textové analýzy. Práce není cílena na analýzu všech údržbových záznamů, ale pouze na záznamy obsahující nálezy neplánovaných závad na letadle při provádění údržby na technické základně. V každém záznamu lze nalézt technikem vyplněný nestrukturovaný text popisující nevyhovující stav nebo závadu na letadle objevenou při prohlídce. V záznamech je možné objevit popis dané komponenty a její závady. Tyto pojmy jsou klíčové pro analýzu záznamů a mohou sloužit například k vyhodnocení nejčastěji se objevujících závad. Práce cílí na návrh možnosti extrahovat z údržbových záznamů tyto informace pomocí strojové textové analýzy, jelikož manuální zpracování by bylo příliš zdlouhavé a nepraktické.

Bezpečnost a spolehlivost letecké dopravy, ke které patří právě i údržba letecké techniky, je pro mě velmi zajímavé téma. Vzhledem k tomuto zájmu a naskytnuté příležitosti věnovat se problematice týkající se údržby letadel a možnosti do tohoto procesu hlouběji nahlédnout došlo ke vzniku tématu této práce. Téma by nemohlo vzniknout bez úzké spolupráce mezi údržbovou organizací CSAT a Ústavem letecké dopravy Fakulty dopravní ČVUT v Praze. Údržbová organizace poskytla svá data, se kterými bylo možné během tvorby této závěrečné práce pracovat a následně provést návrh možnosti jejich zpracování pomocí textové analýzy.

Jedním z cílů této práce je analyzovat současné postupy vzniku záznamů z údržby v údržbové organizaci, seznámit se s obsahem těchto záznamů, použitou terminologií a formu zápisu. V práci je popsán návrh možnosti využití automatické textové analýzy nestrukturované části záznamů, tedy záznamů v podobě volně psaného textu a jejich vyhodnocení. Jelikož se jedná o strojové zpracování dat bylo třeba vybrat vhodný softwarový nástroj umožňující tuto funkci.

Dalším dílčím úkolem bylo vytvoření definičního slovníku pojmů, respektive výrazů, které mají sloužit vybranému nástroji k provádění automatické textové analýzy. Tento slovník slouží jako zdrojový soubor pro vyhledávání a označování pojmů ve vybraných textových datech. Proto došlo k sestavení slovníku s názvem „Slovník komponent a závad“. Ten obsahuje pojmy představující jednotlivé komponenty letadel a jejich závady, které se objevily ve vybraném vzorku dat a byly manuálně extrahovány a zařazeny do slovníku. Na základě vytvořeného slovníku Komponentů a závad byla následně provedena automatická textová analýza všech dostupných údržbových záznamů a její vyhodnocení.

Závěrem a výstupem této závěrečné práce je shrnutí návrhu využití automatické textové analýzy pro vyhodnocování údržbových dat obsahující nálezy závad, dále pak volně dostupný Slovník komponent a závad sestavený za účelem použití v nástrojích umožňující textovou analýzu, a nakonec vyhodnocení vlastní provedené automatické textové analýzy dostupných údržbových záznamů z hlediska její přesnosti a využitelnosti.

## 2. Údržba letadel

Pojem údržba letadel představuje soubor činností vykonaných na letadle, resp. letadlovém výrobku, kterým je drak, motor a vrtule, k zachování jeho letové způsobilosti a možnosti ho dále používat. Jedná se o provedení příslušných kontrol jednotlivých součástí a komponentů letadla a jejich následných oprav nebo výměn v předepsaných intervalech, aby nedošlo k nekontrolovanému opotřebení, šíření závady, či nadužívání, které by mohlo vést k selhání a potenciálnímu vzniku nebezpečné situace při provozu.

Intervaly a příslušné úkony nalezneme v dokumentu zvaném Maintenance Planning Document (MPD). Ten je vytvořen z dokumentu MRB Report (Maintenance Review Board Report), který je schválen výrobcem letadla a je s ním dodáván jako součást jeho průvodní technické dokumentace. MPD mimo jiné obsahuje další nezbytné úkony a části schvalované certifikačními úřady jako je např. EASA.

Všechny opravy se musí řídit manuálem pro údržbu daného typu letadla. Tím je Aircraft Maintenance Manual (AMM). Ten však neobsahuje všechny potřebné informace k provedení oprav, ale odkazuje se na další manuály vydávané výrobcem jako je např. Overhaul Manual (OHM) nebo Structural Repair Manual (SRM), Component Maintenance Manual (CMM) a další.

Údržbu letadel rozdělujeme do dvou následujících kategorií. Jedná se o traťovou údržbu, (Line Maintenance) a těžkou údržbu (Heavy Maintenance).

Traťová údržba zahrnuje provádění jednodušších kontrol a úkonů na letadle, v době mezi jednotlivými naplánovanými lety, kde povoláný mechanik prohlédne celé letadlo dle příslušného check-listu a provede dané úkony. Kontrola se provádí zpravidla na letištní stojánce během odbavování letadla, aby nedošlo k narušení provozu. Patří do ní před/po letová prohlídka letadla, kontrola podvozku a kol včetně brzd, kontrola a doplnění provozních kapalin, odstraňování snadno opravitelných závad a další. Pokud je nalezena rozsáhlejší závada je letadlo přetaženo do hangáru údržbové organizace, kde je možné provést podrobnější prohlídku. [1] [7]

Při těžké údržbě letadla dochází k jeho podrobnější prohlídce, rozsáhlejším opravám a modifikacím, kdy jsou z letadla vymontovány i jednotlivé letadlové celky, aby byl zajištěn dostatečný přístup ke všem částem letadla a jeho komponentům. Na základě počtu nalétaných hodin nebo počtu cyklů letadla udávaný výrobcem dělíme prohlídky na A, B, C nebo D check, od kterých se pak odvíjí rozsah prováděných prací na letadle. Nejnižší A check zahrnuje kontrolu a test klíčových systémů letadla, lubrikaci pohyblivých částí, výměnu filtrů nebo přítomnost a funkčnost nouzového vybavení. Naopak vyšší C nebo D check představuje téměř

kompletní rozebrání letadla a vyjmutí velkých letadlových celků, odkrytí draku a křídel ke kontrole vnitřní konstrukce, rozsáhlé opravy a testy letadlových systémů a další. [1] [7]

Pro potřeby této diplomové práce se již nadále budeme zabývat pouze těžkou údržbou letadel a její problematikou, protože traťová údržba zahrnuje většinou úkoly a odstraňování závad lehčího charakteru, na které tato práce není zaměřena.

## 2.1 Nařízení Komise (EU) č. 1321/2014

Požadavky na zachování letové způsobilosti letadel a letadlových výrobků jsou legislativně zakotveny v Nařízení Komise (EU) č. 1321/2014<sup>1</sup>. Vztahuje se na letadla a letadlové celky, které jsou zapsány v leteckém rejstříku členského státu nebo třetí země v případě, kdy členský stát pro toto letadlo vykonává dozor, a předepisuje společné požadavky a postupy technického a provozního charakteru pro zajištění zachování letové způsobilosti. V nařízení lze najít specifikace a požadavky na organizace řídící zachování letové způsobilosti, ale i na organizace přímo vykonávající údržbu. Zároveň jsou zde uvedeny požadavky na výcvikové organizace poskytující odborný výcvik personálu pro práci v letecké údržbě, ze kterého také plyne kvalifikace, rozdělení a povinnosti osvědčujícího personálu. [2]

Nařízení obsahuje několik příloh, kde každá z nich se věnuje jedné výše zmíněné oblasti: Jedná se o tyto 4 přílohy, které budou níže podrobněji rozepsány:

- Part M – Požadavky na zachování letové způsobilosti
- Part 145 – Oprávnění organizace k údržbě
- Part 66 – Osvědčující personál
- Part 147 – Požadavky na výcvikové organizace

Nařízení Komise (EU) č. 1321/2014 bylo změněno Nařízeními Komise (EU) č. 2019/1383 a 2019/1384. Změna se týká především všeobecného letectví (General Aviation), kde byly upraveny stávající požadavky týkající se zachování letové způsobilosti především pro tento druh letectví. Byly vydány tyto 3 nové přílohy:

- Part ML – pravidla pro zachování letové způsobilosti jiných než složitých motorových letadel (létařících zařízení)
- Part CAMO – požadavky na organizace k řízení zachování letové způsobilosti spadající dříve do Part M podkapitoly G
- Part CAO – umožňuje zároveň řídit zachování letové způsobilost, tak i provádět samotnou údržbu [3]

---

<sup>1</sup> Nařízení Komise (EU) č. 1321/2014 ze dne 26. listopadu 2014 o zachování letové způsobilosti letadel a leteckých výrobků, letadlových částí a zařízení a schvalování organizací a personálu zapojených do těchto úkolů



Tyto části se však netýkají problematiky údržby složitých motorových letadel a jsou zde uvedeny pouze pro úplnost.

### **2.1.1 Part M – Požadavky na zachování letové způsobilosti**

V části M jsou uvedeny požadavky na organizace řídící zachování letové způsobilosti. Těmito organizacemi jsou CAMO (Continuing Airworthiness Management Organisation), které vytvářejí a udržují plán údržby, vedou záznamy o letadle a provedené údržbě, řídí příkazy k zachování letové způsobilosti, provedení modifikací a další. Nařízení stanovuje odpovědnost organizace, úkoly k zachování letové způsobilosti, normy pro provádění údržby, upřesňuje požadavky na odborný personál a pracovní prostory, ale také provozně technické záležitosti, jako je hlášení událostí nebo osvědčení o uvolnění letadla do provozu. [2]

### **2.1.2 Part 145 - Oprávnění organizace k údržbě**

Část 145 předepisuje technické požadavky pro organizace fyzicky provádějící údržbu letadel. Jedná se o organizace AMO (Approved Maintenance Organisation) a mají oprávnění k provádění této činnosti. Nařízení stanovuje požadavky na osvědčující personál, provozní prostory a vybavení, dále pak rozsah a podmínky oprávnění, včetně práv organizace. Každá z těchto organizací musí mít zpracovaný výklad, ve kterém se nachází popis a rozsah prováděných prací, který je schvalován příslušným úřadem. Výklad musí obsahovat organizační schéma, povinnosti odpovědných osob, seznam osvědčujícího a podpůrného personálu, seznam provozovatelů, pro které bude organizace údržbu provádět a další. Organizace může provádět údržbu všech typů letadel, na která získá oprávnění nebo sjednávat údržbu u jiné organizace, ale také sama po dokončení údržby uvolňovat letadla do provozu [2]

Mimo jiné, je v této části nařízení stanoveno, že údržbové organizace musí vést veškeré záznamy o provedené údržbě. Tyto záznamy musí uchovávat po stanovenou dobu, která v současnosti činí 3 roky a po tuto dobu musí zajistit, aby nedošlo k jejich poškození, pozměnění nebo odcizení. V případě potřeby se organizace může záznamy o provedené údržbě prokázat, že byly správně provedeny veškeré úkony, než bylo letadlu vydáno osvědčení o uvolnění do provozu. I po ukončení činnosti organizace je stanoveno, že záznamy z údržby se musejí uchovávat ještě po dobu 2 let. [2]

### 2.1.3 Part 66 - Osvědčující personál

V části 66 nalezneme technické požadavky na osvědčující personál v letecké údržbě. Jsou zde stanoveny podmínky udělení průkazu způsobilosti k údržbě letadel, předpoklady a požadavky kladené na praxi a znalost personálu, práva a omezení nebo také podmínky zachování platnosti průkazu. Rozlišujeme 4 kategorie průkazů:

- A osvědčující mechanik – traťová údržba
- B1 osvědčující mechanik – drak, motor, systémy
- B2 osvědčující mechanik – avionika
- B3 osvědčující mechanik – letouny s pístovým motorem, MTOW < 2 000 kg
- C osvědčující mechanik – uvolnění letadla do provozu

Dle skupiny letadel pak dále rozlišujeme letadlo s turbínovým nebo pístovým motorem, nebo zda se jedná o letoun nebo vrtulník. [2]

Další upřesňující kategorie průkazů jsou definovány v doplňujícím nařízení upravující současné nařízení č. 1321/2014 v částech 66 a 147. Mimo jiné také obsahuje AMC a GM (Přijatelné způsoby průkazu a Poradenský materiál). Tento dokument je možné vyhledat na stránkách Úřadu pro civilní letectví. [8]

### 2.1.4 Part 147 - Požadavky na výcvikové organizace

Část 147 je zaměřená na požadavky na výcvikové organizace zajišťující odborný výcvik osvědčujícího personálu v letecké údržbě. Těmito organizacemi jsou MTO (Maintenance Training Organization) a jejich hlavní činností je zajištění teoretického a praktického výcviku dle osnov uvedených v části 66. Nařízení stanovuje technické požadavky na výukové prostory, vybavení a studijní materiály, dále také nutnost vést záznamy o instruktorech a examinátorech, práva organizace, ale také i výcvikové postupy a záznamy o vykonaných zkouškách. Organizace musí mít zpracovaný výklad, který obsahuje organizační schéma, odpovědné osoby a zkoušející instruktory, dále pak seznam poskytovaných kurzů a organizací, se kterými má výcviková organizace uzavřenou smlouvu ohledně výcviku. [2]

## 2.2 Společnost Czech Airlines Technics

Czech Airlines Technics (dále jen CSAT) je oprávněnou údržbovou organizací AMO dle Nařízení Komise (EU) č. 1321/2014 části 145. Byla založena v roce 2010 a v té době se jednalo o dceřinou společnost Českých aerolinií. V současnosti je však po sloučení v roce 2018 vlastníkem CSAT společnost Letiště Praha, a. s., která se stala jediným akcionářem. [4]

Společnost CSAT sídlí na letišti Václava Havla v Praze v prostorách Hangáru F. Má dlouholetou zkušenost s hangárovou údržbou proudových, ale i turbovrtulových letadel. Poskytuje služby v oblasti traťové údržby a údržby na technické základně, údržbu a opravy podvozků a údržbu komponentů u vybraných letadel typu Boeing 737, Airbus A320 Family a ATR. Ve společnosti je zaměstnáno více než 800 zaměstnanců, od kvalifikovaných mechaniků, přes letecké inženýry až po administrativní a podpůrný personál. [4]

Mezi hlavními zákazníky těžké údržby nalezneme společnosti České aerolinie a. s., Finnair, Transavia Airlines, Austrian Airlines, Jet2.com, Smartwings a. s. a Neos. [5]

CSAT má na letišti k dispozici dva hangáry, Hangár F (Obrázek 1) a Hangár S. Hangár F slouží především pro provádění údržby na technické základně, tj. těžké údržby, ale je možné zde provádět i kontroly údržby traťové. Do hangáru lze umístit vedle sebe až šest letadel typu B737, A320 Family nebo ATR, na kterých lze souběžně provádět údržbu. Ve zbývajících prostorech se dále nachází sklady, klempířská a kompozitová dílna, lakovna, nebo laboratoř nedestruktivního testování a další prostory pro provádění činností, na které má společnost oprávnění. Hangár jako takový a jeho ostatní prostory zároveň slouží jako kancelářské zázemí společnosti, ale i jako zázemí pro mechaniky a ostatní pomocný personál. Novější Hangár S stojí naproti Hangáru F a byl dostavěn v roce 2018. Je využíván výhradně pro traťovou údržbu a je zde k dispozici jedno stání pro letadlo typu B737, A320 Family nebo ATR. [4]



Obrázek 1 - CSAT Hangár F

V následujícím odstavci jsou uvedeny typy letadel, u kterých je společnost CSAT oprávněna k vykonávání údržby.

Provádění těžké údržby těchto letadel:

- B737-300/400/500 (CFM)
- B737-600/700/800/900 (CFM)
- A318/A319/A320/A321
- ATR42/72

Provádění traťové údržby těchto letadel:

- B737 300/400/500/600/700/  
800/900 (CFM)
- B737-8/9 MAX
- B757-200/300 (RB)
- B767-200/300/400 (CF6/PW4000)
- B777-200/300 (GE90/RR)
- B787 Dreamliner
- A320F (CFM/V2500)
- A330-200/300 (PW4000)
- ATR 42/72 (PW)
- EMBRAER 170/190

Dále společnost poskytuje údržbu letadlových podvozků Boeingu B737 SG/NG a podvozkových dílů Airbus A320 a Embraer, údržbu komponentů, strukturální opravy draku, ale také i nedestruktivní testování a odhalování poruch (Non-Destructive Testing). Společnost se také zabývá prodejem materiálu či půjčováním náradí, nebo také nabízí podporu provozovatelům pro zachování letové způsobilosti z pohledu dlouholeté praxe v údržbě letadel. Společnost se neustále rozvíjí a snaží se poskytovat širší pole služeb svým zákazníkům. [6]

### 3. Proces údržby letadla a záznamu nálezu závad

V této části diplomové práce je popsán postup, jakým dochází ke vzniku záznamů nálezů závad v údržbě letadel, jejich nalezení, zaznamenání, zpracování a zavedení do systému a další kroky a úkony, které vedou k jejich vyřešení.

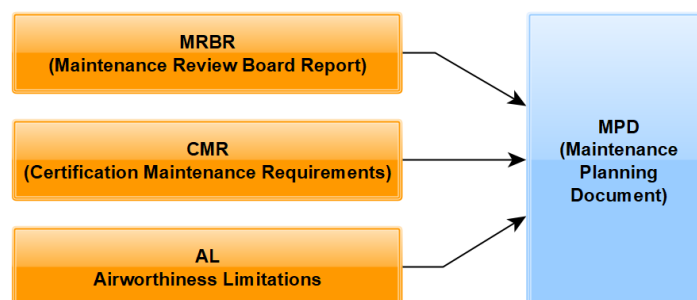
Proces údržby letadla je koncipován a rozdělen na řadu jednotlivých kontrol, dle úkonů a intervalů, které vydává výrobce letadla. Provozovatel si z údajů výrobce a podmínek certifikačního úřadu sestaví nebo nechá sestavit plán údržby, který je vázán na daný typ letadla nebo letadlový park. Podle tohoto plánu, který musí být schválen národním leteckým úřadem, provádí vlastní údržbu letadel nebo si ji sjedná u libovolné údržbové organizace s příslušným oprávněním.

Plán údržby provozovatele obsahuje jednotlivé technologické karty (viz kapitola 3.1), které udávají, jaké úkony mají být provedeny a v jakém intervalu. Provozovatel si sám seskupuje karty s podobnými intervaly a vznikne sada, která je najednou vykonána při jedné údržbě letadla na technické základně, bez nutnosti odstavovat letadlo z provozu kvůli jedné této kartě.

#### 3.1 Technologické karty

Technologické karty označované jako Task Cards nebo Job Cards obsahují úkony, které mají být provedeny při údržbě nebo opravě součásti letadla nebo jeho komponentů. Nalezneme zde v jakém intervalu mají být provedeny, podrobný popis, náhledy a odkazy na další technickou dokumentaci o letadle nebo také požadavky na nářadí, spotřební materiál a časovou náročnost prací.

Technologické karty vydává výrobce letadla, které si následně spravuje provozovatel letadla a vkládá do svého plánu údržby. Plán vzniká na základě Maintenance Review Board Report (MRB Report) dokumentu, který obsahuje úkoly plánované údržby (Tasks) a frekvence jejich provádění. Následně jsou k MRB Reportu přidruženy další části schvalované certifikačními úřady jako je např. EASA. Jedná se o Airworthiness Limitations (AWL), Certification



Obrázek 2 - Vznik MPD

Maintenance Requirements (CMR) a doporučení výrobce letadla. Tímto vznikne dokument zvaný Maintenance Planning Document (MPD) (Obrázek 2). [10]

Následně pak dochází ke vzniku Operator Approved Maintenance Program (OAMP). Jedná se o schválený program údržby, který vychází z MPD a je připravený na míru zákazníkovi, resp. provozovateli letadla. Mimo jiné v sobě zahrnuje instrukce k zachování letové způsobilosti (AD), servisní bulletiny (SB), servisní dopisy (SL), ale také i požadavky ze strany provozovatele.

V letecké údržbě také dochází ke sdružování úkolů údržby do několika jednotlivých balíčků prací (Maintenance packages), které obsahují logicky seskupené technologické karty s podobnými kontrolními intervaly. Tímto se provádí optimalizace údržby a ekonomiky provozu letadel. Práce obsažené v balíku jsou provedeny najednou při větší kontrole letadla jako je C Check nebo D Check, aby se letadlo nemuselo pokaždé odstavovat při nutnosti provedení menšího počtu technologických karet nebo jiných úkonů. [10]

## **3.2 Software využívaný údržbovými organizacemi**

K zajištění přehlednosti a jednoduchosti plánování údržby letadel je možné využít celou řadu dostupných softwarových aplikací, které byly pro tento účel vyvinuty. Software umožňuje komplexní plánování a časovou osu celé údržby letadla, rozvržení dostupných technických a personálních zdrojů, přidělování jednotlivých úkolů a vedení záznamů o provedené údržbě.

Jedná se například o tyto softwarové aplikace a jejich vývojáře:

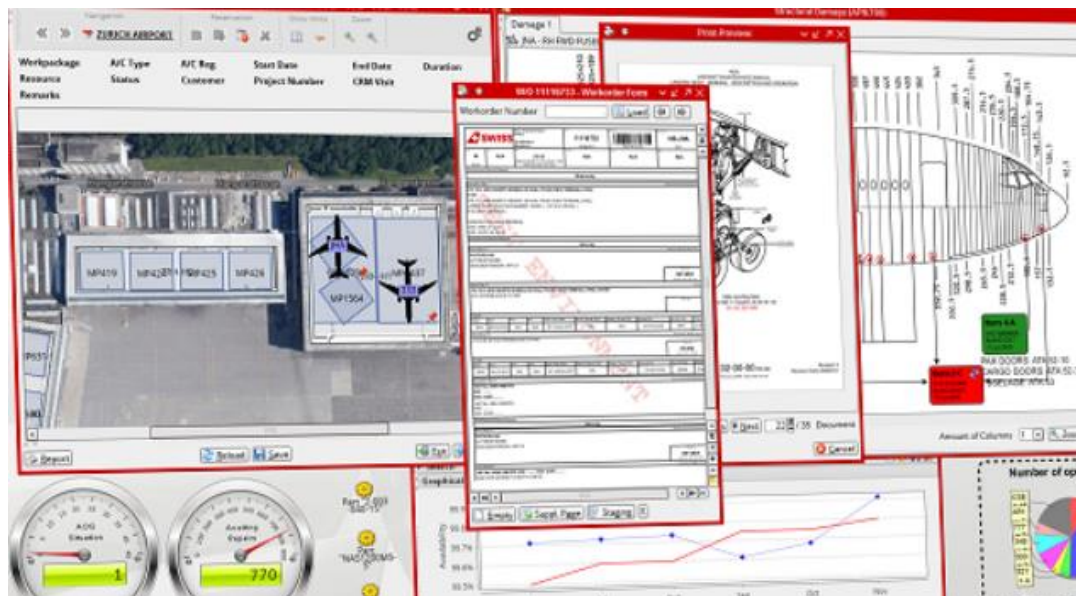
- AMOS – Swiss AviationSoftware
- OASES – Communications Software
- ENVISON – Rusanda
- RAAS – Aviation InterTec Services Inc.
- a další [12]

Software AMOS je využíván společností Czech Airlines Technics a z tohoto důvodu je níže podrobněji popsán.

### **3.2.1 AMOS**

Jedná se o softwarovou platformu určenou pro desktopové počítače vyvinutou společností Swiss AviationSoftware, která slouží jako podpora řízení procesu údržby letadel. Umožňuje organizaci a rozvržení plánované údržby, technického chodu procesu oprav a logistiku prací, jako je např. přidělování dostupných zdrojů a kapacit, zajištění náhradních dílů i spotřebního

materiálu, ale také i vedení záznamů provedených oprav (Obrázek 3). Zároveň je zajištěn i soulad se složitým systémem leteckých předpisů a nařízeních.



Obrázek 3 - Software AMOS [12]

Funkce a rozsah softwaru AMOS, který je v současné době k dispozici, pokrývá požadavky leteckých společností všech velikostí, údržbových organizací pracujících pro určité letecké společnosti, ale i nezávislých poskytovatelů služeb MRO (Maintenance, Repair, Overhaul). Software je vhodný pro všechny typy letadel včetně rotorových a letadel nové generace.

Jádro softwaru obsahuje několik uživatelských rozhraní, jak pro Boeing, tak i Airbus a další. Software dále disponuje řadou modulů podle toho, které si zákazník objednal. Jedná se například o moduly Production, Planning, Engineering, Human Resources nebo také Financial Management.

Shrneme-li výše uvedené, tak je AMOS schopen optimalizovat provozní a finanční efektivitu v odděleních údržby, inženýrství a logistiky moderních leteckých společností a poskytovatelů služeb MRO. [12]

### 3.3 Pracovní příkaz

V údržbě letadlové techniky je možné se setkat s velkým počtem dokumentů a formulářů, které jsou potřeba k řádnému provádění údržby a zajištění její organizovanosti a přehlednosti. Jedním z těchto dokumentů je Workorder (Obrázek 4). Jedná se o pracovní příkaz, který nařizuje provedení určitého typu úkonu na letadle nebo letadlovém celku.

Tyto příkazy mohou vznikat z důvodu dodržení limitů k zachování letové způsobilosti, vydaných příkazů zachování letové způsobilosti a další, nebo také na základě zjištění nálezů



závad na letadle při provádění podrobných prohlídek, které si objednal zákazník u údržbové organizace. Tyto prohlídky se nazývají „Maintenance checks“. Technik údržby letadel si k sobě vezme tzv. technologickou kartu, která v sobě obsahuje úkony, které musí být provedeny a místa, která musí být zkontrolována. Při zjištění jakékoliv nesrovnalosti či nevyhovujícího stavu, mechanik vše pečlivě zaznamená jako nález závady, resp. závad, které následně zadá do systému údržbové organizace. Systém pak automaticky vytvoří z tohoto nálezu nový pracovní příkaz, který přesně udává, co na letadle musí být provedeno nebo opraveno.

Každý Workorder má své specifické číslo a čárový kód pro snadnou identifikaci letadla. Je v něm uvedeno registrační označení letadla, kterého se pracovní příkaz týká, dále pak jméno zákazníka nebo také termíny splnění. Pokud se jedná o letadlový celek nebo součást, tak je zde uvedeno číslo dílu, popis a jeho sériové číslo. Je zde i označeno z jakého původu pracovní příkaz vzniknul. Workordery mohou vznikat na základě:

- Hlášení pilota (Pilot Report)
- Hlášení posádky (Crew Report)
- Nálezu při údržbě (Maintenance finding)
- Plánovaných prací (Scheduled)

Hlášení pilotů a posádek vzniká na základě projevení se závady během provozu letadla, plánovaný pracovní příkaz (Scheduled) vychází z plánovaných prací a údržby na letadle a nález při údržbě je neplánovanou závadou objevenou při provádění plánované údržby letadla.

Důležitou položkou je odkaz na technologickou kartu, při které došlo k nálezu na letadle a kapitola ATA, která označuje, o jakou část letadla nebo systém se jedná. V poli Description pak nalezneme podrobný popis závady, kterou je potřeba odstranit nebo opravit. Po ukončení prací se do pole Action zaznamená to, co bylo provedeno, vyměněno nebo opraveno. Pokud byl nalezen další nevyhovující stav ostatních součástí opravované komponenty, tak je to v tomto poli také zaznamenáno a následně je pro tento nález vytvořený nový Workorder. Za vše zodpovídá příslušný mechanik, který svou práci stvrzuje podpisem.

Ve Workorderu můžeme také nalézt tabulku s vyžadovanými součástkami, které jsou potřeba pro provedení opravy nebo také počet hodin strávených prováděním daného pracovního příkazu.



<b>CZECH AIRLINES TECHNICS</b> Czech Airlines Technics Jana Kasparska 1/1069 160 00 Praha 6 Czech Republic		<b>WORKORDER</b>		A/C Registration	Barcode 	ATA Chapter 25-30						
<input type="checkbox"/> Component:	Part	Description	Part	Station / P/N Code								
<input checked="" type="checkbox"/> Customer:		Project:	Type/Origin									
<input checked="" type="checkbox"/> Finding ref. Taskcard:	253000-002	Due At:	Date	Time	FAC	FAC						
Description of complaint or planned work						Type						
1 WORKSTEP ADDED BY ADAMICKI ON 09.MAY.2018, 08:10 FINDING (NRC) TASKCARD 253000-002 (1.0) / ITEM 1 GALLEY G5 THE PART FROM LATCH ASSY IN BOX 529 WAS FOUND MISSING. DESCRIPTION SIGN						<input type="radio"/> Pilot Report <input type="radio"/> Cabin Report <input checked="" type="radio"/> Maintenance D. <input type="radio"/> Scheduled <input type="radio"/> Feedback R. <input type="radio"/> Briefing C. <input type="radio"/> Robbery						
						Issue Sign / Stamp						
Station PRG Date 09.May.2018 Time (UTC) 08:10 Flight #												
Action / Work Performed						Performed	Inspected					
ACTION PERFORMED BY ON 15.MAY.2018, 07:20 FINDING (NRC) TASKCARD 253000-002 (1.0) / ITEM 1 GALLEY G5 NEW LATCH HAS BEEN INSTALLED. AMM:25-33-50 REV:66 PERFORMED SIGN												
<input type="checkbox"/> PFI MM <input type="checkbox"/> sign/3LC <input type="checkbox"/> Daily Check MM <input type="checkbox"/> sign/3LC												
Continued on <input type="radio"/> Next Workorder / Log Page No: <input type="radio"/> Staging <input type="radio"/> Supplement Page (same number) Sign:												
Component Changes				References		Used Manhours						
Label Number	Seq.	Partnumber	Serialnumber	Type	Description	Pers. ID	MHRS	Date	Type			
	ON			MET	253000-002 (1.0) / ITE		00:45	15.05.18	M-INT			
	OFF			REQ	253000-002 (1.0) / ITE							
	ON			WP	OH-LVA/H-18 C13							
	OFF											
	ON											
	OFF											
Oil Uplift (In Quarts)	Engine	#1	#2	IDG/CSD	#1	#2	APU	Hydr	#AY	#BG	#CB	Apu Counter
												HRS: CYC:
Transfer												
<input type="checkbox"/> DD (HIL) <input type="checkbox"/> MEL Ref. MEL CAT: <input type="radio"/> A <input type="radio"/> B <input type="radio"/> C <input type="radio"/> D Downgrades <input type="checkbox"/> Prev. Approach CAT: <input type="checkbox"/> No Autoland <input type="checkbox"/> No RVSM <input type="checkbox"/> ETOPS (Minutes) <input type="checkbox"/> APU No Bleed <input type="checkbox"/> APU No Elec.												
NO	Station	Date	Time	Days	HRS	CYCL	Check	Reason	Sign / Stamp			
										Release Transfer		
										Release Transfer		
										Release Transfer		
<input type="checkbox"/> AOG Risk <input type="checkbox"/> Occurrence <input type="checkbox"/> OPS Consequence <input type="checkbox"/> Watch Item												
Work Performed, Workorder Closed										Sign / Stamp		
Station	Date	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time	Time
PRG	15.May.2018	07:22										
Certifying Company <input checked="" type="radio"/> CZ.145.0067 <input type="radio"/>										Closing Sign		<input type="checkbox"/> IVDI <input checked="" type="checkbox"/> N/A
Certifies that the work specified, except as otherwise specified, was carried out in accordance with PART 145 and in respect to that work the Aircraft / Aircraft component is considered ready for release to service												
M-R Planning		Part Request										
Skill / Shop	Est. MHRS	Partnumber	Description		IPC Reference	QTY						
MECH-INT	01:00	216534-6	LATCH ASSY			1						
Resources												
<input type="checkbox"/> Hangar <input type="checkbox"/> Dock <input type="checkbox"/> Jacks <input type="checkbox"/> PWR Run <input type="checkbox"/> Test Flight <input type="checkbox"/> Mass & Balance <input checked="" type="checkbox"/> Est. Gt. 01:00 HR <input type="checkbox"/> Idle Run <input type="checkbox"/> Other:												
Statistics / Assessment												
<input type="checkbox"/> Incident <input type="checkbox"/> Warranty <input type="checkbox"/> Accident <input type="checkbox"/> Insurance												

Obrázek 4 – AMOS - CSAT Workorder

### 3.4 Vznik záznamu nálezu závady

Při provádění samotné údržby je příslušnému technikovi údržby letadel vždy přidělena jedna z těchto technologických karet. Tu si vezme k ruce a dozví se z ní, které části letadla se týká, dle číslování ATA (viz kapitola 4.1.1), a jaký systém nebo součást mají být zkontrolovány. Technologická karta obsahuje příslušné pokyny kontroly a odkazy na další technickou dokumentaci k letadlu jako je např. AMM (Aircraft Maintenance Manual). Bývá zde i uveden potřebný materiál, nářadí nebo také časová náročnost práce.

Během provádění příslušné technologické karty může mechanik objevit u jakékoliv zkoumané části poškození nebo nevyhovující stav. Jedná se tedy o nález závady, který musí být řádně označen a zaznamenán. Následně je mechanikovou povinností zavést tyto nálezy závad do systému údržbové organizace. Zpravidla připadá jeden záznam jednomu nálezu závady. Technik zde uvede při jaké technologické kartě k nálezu závady došlo, číslování dle ATA normy, registrační označení letadla, popis nálezu závady, případně i požadovaný materiál nebo komponentu. Vše je stvrzeno jeho podpisem. Některé náležitosti jsou automaticky načteny pomocí čárového kódu, který usnadňuje proces vzniku záznamů nálezů.

Následně jsou, na základě vzniklých nálezů závad na letadle, systémem vygenerovány tzv. pracovní příkazy (viz kapitola 3.3), které přímo nařizují provedení příslušné opravné práce nebo nápravy závady. Technik si tento pracovní příkaz vezme k ruce, dozví se z něj potřebné informace, mimo jiné i popis závady (pole Description), provede vlastní opravu či návrh řešení a vykonanou činnost zaznamená, tentokrát do pole Action Performed. Na základě dalších nálezů nebo potřebných prací může být pracovní příkaz rozšiřován nebo je uzavřen.

## 4. Analýza existujících slovníků

V této kapitole diplomové práce je pojednáno o stávajícím způsobu vzniku záznamu nálezu závady v údržbě letadel, resp. původu používané terminologie. Hlavním zdrojem používaných výrazů jsou dokumenty poskytované výrobcem letadla. Tím je průvodní technická dokumentace, která v sobě obsahuje veškeré potřebné informace o provozu a údržbě letadla daného typu.

Tato dokumentace je rozdělena na větší množství specificky zaměřených manuálů, právě na základě složitosti letadla a jeho systémů. Manuály jsou systematicky rozděleny dle mezinárodně uznávané normy ATA 100, o které je pojednáno níže v této části práce, stejně tak jako o průvodní technické dokumentaci.

Veškerá technická dokumentace a manuály, které jsou dostupné technikům údržby letadel, slouží jako zdroj terminologie, kterou mechanici používají k záznamu nálezů na letadle. V manuálu jsou obsaženy přesné popisy a názvy součástí opravovaných celků letadel. Při opravách mají technici manuály stále k dispozici a řídí se dle uvedených postupů a pokynů. Záznamy nálezů závad by pak měly vznikat na základě termínů obsažených právě v těchto manuálech.

Je možné říci, že stejné či podobné nálezy závad by měly být zaznamenány stejnými výrazy. To však není v každém případě pravda, protože každý technik může volit jinou formulaci slov či synonyma a nalezenou závadu popsat trochu jinak, může použít zkráceniny nebo udělat gramatickou chybu. Kombinace výše zmíněných faktorů však přináší komplikace z hlediska automatické analýzy záznamů, jelikož by byl potřeba univerzální slovník, který by pokryl tyto variace.

### 4.1 Klasifikace nálezů závad z údržby

Klasifikace nálezů z údržby letadel je pro tuto diplomovou práci zcela podstatná. Na základě ní může vzniknout univerzální slovník pojmů, který lze použít pro automatickou textovou analýzu záznamů z údržby.

Nálezy lze klasifikovat do dvou různých kategorií. Jedná se o komponentu (Component) a o závadu (Failure). V záznamech nálezů je vždy uveden jak popis komponenty, u které mechanik objevil nevyhovující stav, tak i popis tohoto stavu, resp. selhání nebo poruchy. Pokud není přímo uveden komponent, tak je uvedena lokace nálezu závady a popis poškození.

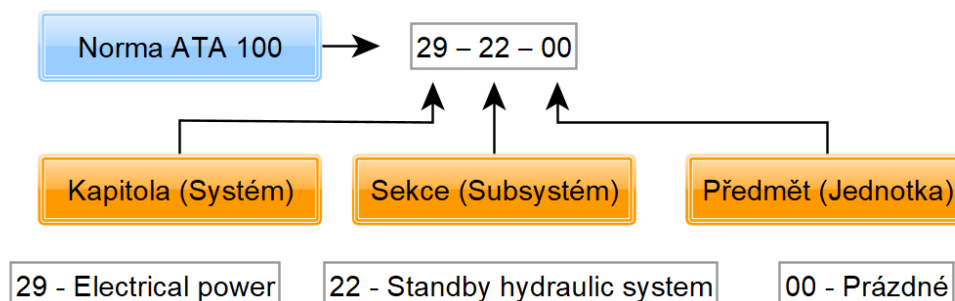
Text nálezů obsahuje také další doplňující informace, jako je např. počet poškozených komponent, které však nejsou pro potřeby této práce důležité, protože hledáme pouze

komponentu a její poškození. Kombinace komponenty a její závady tvoří nejdůležitější základ pro potřeby textové analýzy a obsáhnou zvolenou míru podrobnosti.

#### 4.1.1 Norma ATA 100

Jedná se o normu, kterou vydala v roce 1956 americká asociace pro leteckou dopravu ATA (Air Transport Association). Norma je mezinárodně uznávaná a byla již od té doby nespočetněkrát aktualizována. Udává číselné rozdělení a klasifikaci všech systémů a subsystémů v letadle. Dle tohoto číslování je rozdělena průvodní technická dokumentace a lze se podle ní snadno orientovat. Tato norma je univerzální a stejné označení dle normy ATA 100 označuje stejnou část nebo systém u různých výrobců letadla. To zajišťuje přehlednost a nezaměnitelnost systémů a komponentů. [9]

Označení normou ATA 100 (Obrázek 5) se skládá z číselného označení dle tabulkové předlohy. První dvě čísla udávají oblast letadla nebo systém, kterou lze zúžit oblast zájmu. Druhým dvojčíslím upřesňuje specifickou podoblast či subsystém zvoleného systému a třetí dvojčíslí v případě použití ještě více upřesňuje specifickou komponentu nebo součást, která je předmětem zájmu a je třeba ji zkontrolovat nebo opravit. Níže je provedeno vysvětlení na příkladu. [9]



Obrázek 5 - Číslování normy ATA 100

Normu ATA 100 lze rozdělit do skupin:

- Letadlo – obecně ATA 0–20
  - např. ATA 4 – Airworthiness limitation, ATA 11 – Placards and markings
- Systémy letadla ATA 21–50
  - např. ATA 27 – Flight control, ATA 32 – Landing gear
- Konstrukce letadla ATA 51–57
  - např. ATA 53 – Fuselage, ATA 57 - Wings
- Vrtule ATA 60–67
  - např. ATA 61 – Propellers / Propulsors

- Pohonná jednotka 71–85
  - Např. ATA 76 – Engine control, ATA 79 - Oil

### 4.1.2 Průvodní technická dokumentace

Jedná se o řadu dokumentů, které výrobce letadla dodává zákazníkovi, resp. provozovateli letadla, aby mu bylo umožněno zajistit zachování letové způsobilosti. Tato dokumentace je následně poskytována údržbové organizaci provozovatele nebo organizaci, u které má provozovatel sjednanou smlouvu.

Základním dokumentem k údržbě letadla je Aircraft Maintenance Manual (AMM). Ten obsahuje informace k provedení údržby a oprav letadla nebo jeho komponentů, které jsou na něm nainstalovány. U složitých motorových letadel by byl však tento dokument příliš rozsáhlý a nepřehledný, proto neobsahuje všechny potřebné informace. K tomu slouží další technická dokumentace, na kterou se AMM může odkazovat. Některá dokumentace je dodávána spolu s letadlem a jiná zase k dispozici u poskytovatele služeb třetích stran. [13]

Jedná se o následující dokumenty, které nemusí být všechny vždy dodávány výrobcem letadla: [13]

- Overhaul Manual (OHM)
- Structural Repair Manual (SRM)
- Wiring Diagram Manual (WDM)
- Fault Isolation Manual (FIM)
- System Schematic Manual (SSM)
- Non-destructive Testing Manual (NDTM)
- Illustrated Tool & Equipment Manual (ITEM)
- Component Maintenance Manual (CMM)
- Illustrated Part Catalog (IPC)
- Dispatch Deviation Guide (DDG)
- Configuration Deviation List (CDL)

Všechnu tuto dokumentaci musí mít mechanici k dispozici a jsou povinni ji využívat a řídit se dle ní při provádění údržby na letadle. Technici pracují s technologickými kartami, které se na dokumentaci odkazují a zde si vyhledají příslušnou část, kterou dostali za úkol zkontrolovat nebo opravit a řídí se dle uvedeného postupu. Po provedení příslušných úkonů jsou povinni vše pečlivě zaznamenat a nahrát do systému organizace pro zajištění vedení záznamů o provedené údržbě.

Jelikož v dokumentaci nalezneme přesný popis a název jednotlivých komponentů, tak se jedná o nejčastější zdroj terminologie používané k vytváření záznamů z údržby. Záznamy však mohou být zcela odlišné i v případě, že se jedná o stejnou závadu či nález, protože každý technik ji může popsat trochu jinými slovy (synonyma, zkráceniny apod.) nebo nepoužije úplně celý název komponenty, které se porucha týká.

## 4.2 Textová analýza a její využití

Textová analýza, jinými slovy označovaná jako „text mining“ nebo také textová analytika je proces, který umožňuje systematické zkoumání textových dat s cílem odhalit souvislosti a trendy v těchto datech a následné zavedení strategických postojů nebo opatření dle zjištěných výsledků. Textová analýza může být prováděna manuálně nebo automaticky systémovým nástrojem, který proces značně zjednoduší.

Manuální provádění analýzy často znamená ruční procházení záznamů organizace, např. excelovských tabulek, jednou nebo více zodpovědnými osobami, což může být při velkém rozsahu dat časově náročné, neefektivní a nepřesné. Proto se organizace častěji obrací na automatické možnosti zpracování textových dat pomocí nejrůznějších softwarů k tomu určených. Tyto softwary využívají strojového učení a algoritmů pro přirozené zpracování jazyka („NLP – natural language processing“) k nalezení významu a pochopení ohromného množství textových dat. [14]

V nejjednodušší podobě textová analýza může vypadat jako jednoduchý nástroj sloužící k prostému počítání výskytu slov v objemu zkoumaných dat. Pokročilejší metody textové analýzy však v sobě zahrnují nejenom tuto základní funkci, ale i analýzu vztahů a souvislostí mezi nalezenými výskyty, a to umožňuje přiřadit jednotlivým pojmům určitou váhu. Z hlediska metody využití k automatické analýze textu využíváme tyto dva přístupy: [14]

- specifikování individuálních slovníků (slov nebo frází) a zpozorovaných vzorců – často zpracováno manuálně a následně rozšiřováno, podle toho, co hledáme
- samostatné zpracování počítačem, který nalezne výskyty a vztahy za nás – podle úrovně analýzy může být provedeno automaticky s minimálním externím vstupem

Využití textové analýzy je vhodné, pokud máme velké množství nestrukturovaných dat, které nemají sami o sobě žádný význam nebo nejsou kategorizovaná. Tyto data v sobě obsahují informaci, která může být v jistém ohledu podstatná nebo nás může zajímat. Data proto musí být analyzována pro potřebu hlubšího pochopení významu a spojitosti dostupných záznamů organizace a následné možnosti změny přístupu při vykonávání její činnosti. Aplikování textové analýzy dat je možné kdekoliv, kde máme k dispozici jakákoliv textová data a má za úkol pomoci snadno těmto datům kvantitativně i kvalitativně rozumět. Je možné dohledat

využití analýzy v oblastech podpory, zpětné vazby nebo hodnocení zákazníků, marketingu, sociálních medií v podobě analýzy příspěvků, ochrany před spamem apod. [15]

V případě více technické aplikace pro potřeby této práce nebo podobným účelům bylo možné dohledat několik článků, které se touto problematikou zabývají a jsou popsány v další kapitole.

#### 4.2.1 Dostupná vědecká literatura

Při zjišťování podkladů využitelnosti textové analýzy nestrukturovaných dat bylo možné dohledat několik vědeckých článků, které se týkají klasifikace nálezů a analýzy záznamů z oblasti průmyslu a údržby („Maintenance“). Jedná se o případové studie textové analýzy záznamů z údržby v oblasti železniční infrastruktury a tankovacích stanic. Nebylo však možné dohledat studii zabývající se možností použití takovéto analýzy přímo v oblasti údržby letadel.

V prvním článku nazvaném „*Natural Language Processing of Maintenance Records Data*“ se můžeme dočíst, že záznamy o údržbě obsahují řadu vyplňovacích textových polí. Jedná se např. o identifikační číslo záznamu, informace týkající se systému, subsystému a komponent, provedenou činnost, příčinu poruchy nebo nápravné opatření. Některá pole jsou systémem pevně daná, jiná jsou zase vyplněna ručně personálem. Tyto textové vstupy mohou obsahovat jakákoliv slova v libovolném množství a tvoří tak nestrukturovaný text. Všechny tyto dostupné informace spolu tvoří podklady pro možnou analýzu záznamů údržby. [16]

Možnost získat relevantní informace závisí na kvalitě hrubých dat a způsobu jejich zpracování. Pro monitorování výkonosti a nákladů využívají organizace počítačových analýz a jiných metod hlavně tam, kde je důraz kladen na bezpečnost provozu jako je např. výroba, doprava, letectví nebo jaderný průmysl. Nedostatek relevantních dat a informací je jeden z hlavních problémů v rozhodovacím procesu údržby. Dostupnost správných informací příslušnému personálu v požadované kvalitě a ve správný čas je zcela zásadní. [16]

K provedení textové analýzy je potřeba mít k dispozici určitý dokument či slovník, který definuje nebo jinak kategorizuje významné pojmy a slouží jako zdroj analýzy a strojového učení. Výstupem analýzy je pak statistika zadaných výrazů, resp. výskytu definovaných slov v záznamech údržby, v textovém poli, které je zadáváno příslušnou obsluhou. [16]

Závěrem článku je, že pomocí textové analýzy je možné zlepšit schopnost identifikace poruch, jelikož do té doby nebyly k dispozici statistická pozorování, které mohou mít velký vliv na střední dobu mezi poruchami a údržbovými časy. Máme tak k dispozici více relevantní data do systému spolehlivosti, který slouží pro odhad údržbových intervalů.

Druhý článek s názvem „*Clustering and Classification of Maintenance Logs using Text Data Mining*“ se věnuje případové studii o textové analýze záznamů údržby přečerpávacích stanic přehrady. Jejím cílem je klasifikace dat plánované i neplánované údržby, která slouží jako podklad pro samotnou analýzu dat. V jistém ohledu se opět jedná o jakýsi slovník určující objekty našeho zájmu, ve kterém jsou uvedeny výrazy, jejich dělení a klasifikace nebo také vazba na ostatní termíny. [17]

Po spuštění analýzy nejprve dochází k tomu, že každému výrazu (termínu) je spočtena a přidělena určitá váha, která hraje roli na počtu výskytu v datech a také na klasifikaci určené slovníkem. Následně pak systém automaticky vytváří skupiny („clustery“), kde jsou shlukovány výrazy s podobným významem, váhou a vazbou určenou právě definičním slovníkem. Výstupem analýzy je opět statistika nejčastěji se objevujících výrazů (komponentů, úkonů, závad apod.), kterou je možné využít ke změně přístupu k údržbě, optimalizaci jejího procesu nebo možnosti plánování intervalů oprav. [17]

Po zhodnocení obou případových studií lze konstatovat, že k provedení jakékoliv textové analýzy dat je zapotřebí mít dostupný referenční dokument nebo slovník, který definuje slova, výrazy nebo fráze, které nás zajímají. Může taktéž obsahovat vzájemné vazby mezi těmito „termíny“ a jejich klasifikaci.

Jedním z dílčích úkolů této práce je vytvoření definičního slovníku z prostředí údržby letadel pro potřeby textové analýzy záznamů údržbových dat z tohoto odvětví.

### **4.3 Limitace současného stavu**

Při provádění údržby letadel a vedení příslušné dokumentace o opravách lze nalézt několik omezení, které mají především vliv na vznik a formu záznamů nálezů závad. Jak už bylo řečeno dříve, záznam nálezu závady vzniká na základě prohlídky letadla, resp. letadlové části nebo systému letadla, kterou provede mechanik dle příslušných pokynů uvedených v technologické kartě a další technické dokumentaci.

Pokud při prohlídce nebo zkoumání komponenty mechanik nalezne poškození či jiný nevyhovující stav, musí dojít k jeho řádnému zaznamenání. Proto se mechanik vydá na příslušné místo, kde má přístup do systému údržbové organizace a nalezenou závadu do něj zaeviduje. Nevýhodou můžeme nalézt v tom, že mechanik neprovádí prohlídku s přenosným počítačovým zařízením, např. notebookem nebo tabletem (Obrázek 6). Takto by mohly být nálezy závad zaznamenány do systému téměř okamžitě, a to bezprostředně po jejich objevení, nikoliv však až po dokončení prohlídky dle příslušného checklistu. Mohla by se tím eliminovat možnost chyby při vytváření záznamu nálezu závady např. v podobě záměny, nepřesnosti a popisu závady, pokud by jich mechanik našel při dané prohlídce větší počet.



Dalším omezením a nevýhodou, která se týká popisu nalezené závady na letadle, je skutečnost, že každý mechanik je individuální osobností s rozdílnou jazykovou úrovní a schopností se vyjadřovat. Mechanici musí být schopni komunikovat písemnou i slovní formou daným úředním jazykem v oblasti civilního letectví, kterým je v Evropě angličtina. Při vytváření záznamu nálezu závady se může pravděpodobně stát, že stejnou závadu nalezenou na letadle mohou dva různí mechanici popsat každý trochu jinými slovy, resp. synonymy, která v závěru budou znamenat totéž. Záznamu nálezu závady budou zcela pravděpodobně rozumět oba mechanici, a tím pádem z pohledu manuální práce se záznamy to nečiní žádný problém. Ten přichází až s automatickým zpracováním těchto záznamů.

Největší limitací analýzy záznamů z prostředí údržby letadel je vysoká obtížnost manuálního zpracování a extrakce potřebných informací a omezené používání nástrojů k vyhodnocení nestrukturovaných dat. V údržbě nalezneme velké množství těchto záznamů a údržbové organizace provádějí jejich analýzu pouze v omezené míře.

Při automatickém zpracování dat je zapotřebí, aby byl počítač, resp. nástroj na zpracování dat schopen rozlišovat jednotlivá slova, jejich význam a případné vazby jako jsou např. synonyma, zkratky, používané názvy, klasifikace nebo jiné označení apod. K tomu, aby bylo možné provádět takovéto automatické analýzy musí být k dispozici definiční slovník pojmů objevujících se v záznamech, na které je v plánu aplikovat tuto metodu.

Další nevýhodu, kterou je potřeba uvést, v návaznosti na předchozí odstavec, je neexistence veřejně dostupného definičního slovníku z prostředí údržby letadel. Je pravděpodobné, že pokud údržbové organizace využívají možnosti textových analýz, tak si spravují vlastní definiční slovníky pojmů, které mají jen pro vlastní použití.



Obrázek 6 - Přenosné zařízení při údržbě [18]

## 5. Nástroje umožňující textovou analýzu

V této části práce jsou rozebrány dostupné možnosti softwarových nástrojů, pomocí kterých je možné provádět textovou analýzu souborů a dokumentů. V případě této práce se jedná o analýzu záznamů z údržby letadel týkajících se neplánovaných nálezů závad při provádění prohlídek na technické základně. Tyto záznamy jsou v podobě nestrukturovaného textu a je třeba nalézt vhodný nástroj pro jeho klasifikaci a vyhodnocení.

Při zjišťování dostupných možností bylo objeveno několik nástrojů, které funkci textové analýzy umožňují. Jejich popis a možnosti využití jsou popsány v následujících kapitolách. Na závěr je dostupné porovnání a volba vhodného nástroje.

Vybrané nástroje umožňující textovou analýzu:

- Termlt
- Brat
- Gate

### 5.1 Termlt

Termlt je snadno použitelný online webový nástroj vyvinutý Fakultou elektrotechnickou ČVUT v Praze, který slouží jako přehledný správce uživatelem vytvořených slovníků a pro následnou editaci vybraných pojmů. Termlt umožňuje svým uživatelům definovat vlastní pojmy (termy), uspořádat je do hierarchií dle požadované struktury a použít je pro anotaci zvolených textových dokumentů a jejich vyhledávání. U každého pojmu lze uvést jeho základní informace, které ho specifikují. Tím je název pojmu (label), definice, zdroj definice, další doplňující údaje nebo poznámky. Pojmy mohou být následně dále klasifikovány pomocí nadřazených pojmů nebo dalších zjednodušených slovníků, které umožňují vytvoření požadované hierarchické struktury. [19]

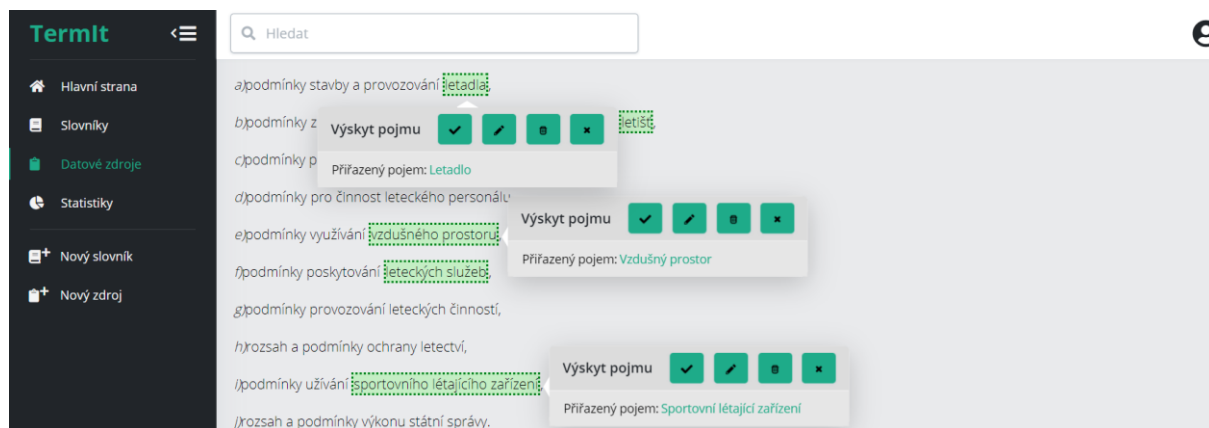
Hlavní funkcí nástroje Termlt je:

- Správa slovníků (Vocabulary management)
- Anotace zdrojů a vyhledávání (Resource annotation and search)

Při použití nástroje Termlt nejdříve dochází k vytvoření uživatelem požadovaného slovníku, do kterého manuálně zadá vybrané pojmy, které jsou předmětem jeho zájmu. Následně je pak ke slovníku připojen dokument, u kterého má uživatel v úmyslu provést textovou analýzu. Po jejím spuštění nástroj začne analyzovat dokument na základě konceptu uvedeného v přidruženém slovníku tak, aby našel zmínky v textu, které se odkazují na konkrétní pojmy ve slovníku a poskytuje vazby mezi nimi. Nalezené pojmy jsou následně v textu zvýrazněny

na základě přiřazeného ohodnocení výskytu a uživatel má možnost navrhovaný výskyt pojmu schválit nebo změnit na jiný termín (Obrázek 7). [19]

Výstupem z nástroje je pak statistika výskytu pojmů v analyzovaných dokumentech, které jsou uvedeny ve vytvořeném slovníku, a který byl použit k provedení textové analýzy. TermIt je pod neustálým vývojem a jsou do něj postupně importovány nové funkce a opravy chyb. [19]

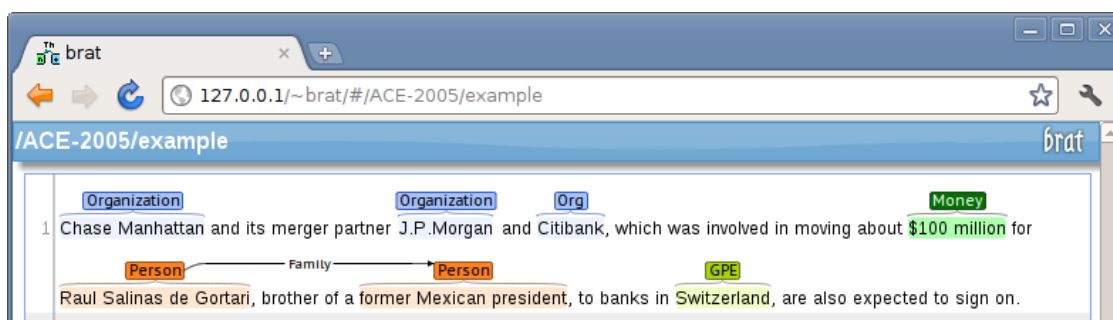


Obrázek 7 - Nástroj TermIt

## 5.2 Brat

Brat („brat rapid annotation tool“) je online webový nástroj, který umožňuje provádět anotaci zvolených textových souborů nebo dokumentů. Slouží pro přidávání „poznámek“ k existujícím pojmům v textu, které jsou pro uživatele svým způsobem důležité, a stanovení jejich významu nebo vazby na jiné pojmy. Brat je navržen především pro strukturované anotace, kde předpokladem je, že pojmy, kterými chceme text kategorizovat nemají libovolnou podobu, ale pevně danou formu, resp. specifickou strukturu dle předem daných podmínek. Ta je klíčová, aby mohly být textové soubory automaticky zpracovávány a interpretovány za pomoci počítače. [20]

Jednotlivým výrazům v textu lze přiřadit uživatelem vytvořené kategorie, kterými dojde k označení zvolených anotovaných pojmů, o které má uživatel zájem. Následně je pak umožněno dále vytvářet vazby mezi těmito kategoriemi a utvářet tak celkovou anotační strukturu textového dokumentu pomocí nadřazených pojmů a vazeb (Obrázek 8). [20]



Obrázek 8 - Nástroj Brat [20]

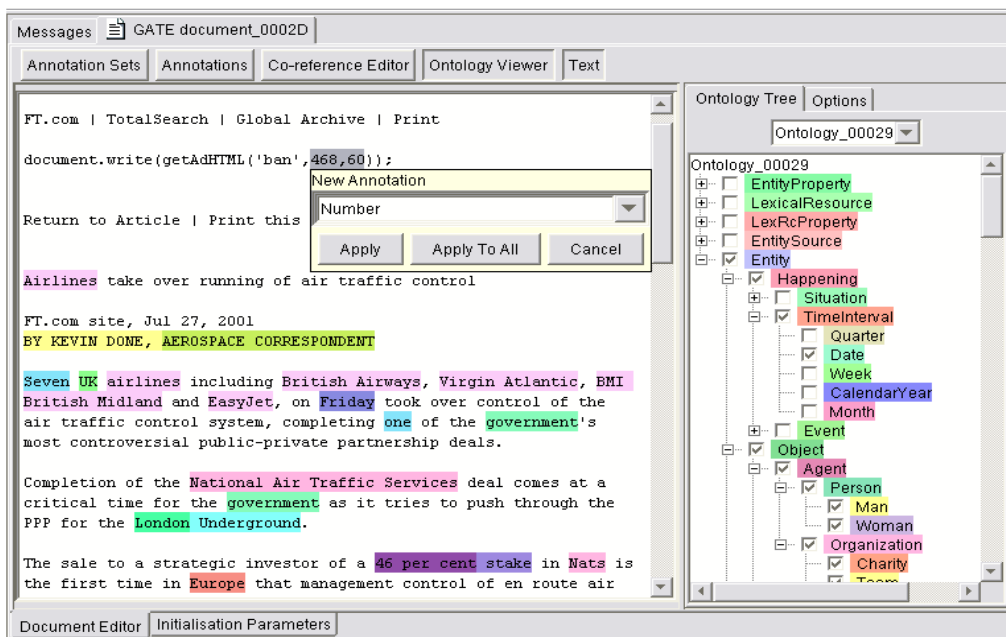
Použité kategorie anotace, vybraný typ a omezení týkající se jejich použití, jsou plně přizpůsobitelné dle požadavků uživatele, a tím je tak umožněno použití nástroje Brat na téměř jakoukoliv formu anotace textu. V nástroji je také implementováno několik funkcí, které spoléhají na techniku přirozeného zpracování jazyka (NLP). [20]

### 5.3 Gate

Gate („general architecture for text engineering“) je volně dostupný softwarový nástroj, který je možné aplikovat téměř na jakoukoliv problematiku týkající se zpracovávání textu. Už více než 15 let se aktivně používá pro všechny typy výpočetních úloh zahrnující analýzu a zpracování přirozeného jazyka. Gate je možné použít k textové analýze jakéhokoliv textu, všech struktur a velikostí. Nástroj má rozsáhlou komunitu vývojářů, uživatelů a vědců, a je využíván nejrůznějšími organizacemi, výzkumnými laboratořemi a univerzitami po celém světě. [21]

Nástroj je složen z několika jednotlivých rozhraní, kde každé má jinou funkci a účel použití. Jedná se např. o GATE Teamware, GATE Developer, GATE Mimir a další. Podle potřeby uživatele jsou nainstalována příslušná „GATE“ rozhraní a následně dochází k jejich aplikování na dostupná textová data, které jsou potřeba analyzovat. [21]

GATE Teamware je snadno použitelná webová platforma pro tvorbu a správu anotací (Obrázek 9). GATE Developer je vývojové prostředí, které poskytuje bohatou sadu grafických interaktivních nástrojů pro zpracování přirozeného jazyka a napomáhá vytváření složitých struktur, vizualizací a zpracování výsledků. [21]



Obrázek 9 - Nástroj Gate [22]

## 5.4 Zhodnocení a volba vhodného nástroje

Všechny výše uvedené softwarové nástroje jsou svým způsobem unikátní a každý má své výhody i nevýhody. Záleží pak na uživateli, který z nich si vybere a poslouží mu k požadovanému účelu. Krátký popis jednotlivých nástrojů však nedokáže vystihnout všechny jejich funkce, a proto je zde nutné uvést jednoduché zhodnocení.

Hlavní funkcí všech tří představených nástrojů umožňující provádění textové analýzy je tvorba anotací pojmů (termů) obsažených v textu a jejich ontologické struktury. Dochází tak ke vzniku definičních slovníků, podle kterých je prováděna samotná textová analýza souborů a dokumentů. Následně pak dochází k vizuální reprezentaci dat, v podobě označení výskytů pojmů dle definičního slovníku. Výstupem z analýzy je ve většině případů statistika počtu vyskytujících se pojmů v analyzovaném dokumentu, respektující uživatelem zvolené atributy a zvolené vazby na jiné pojmy dle definované ontologické struktury.

Nástroj Brat představuje komplikace v podobě jeho zprovoznění a aplikace. Navíc umožňuje funkci vytváření vazeb mezi jednotlivými termíny, kterou však v této práci není potřeba využít. Naopak nástroj Gate je uživatelsky náročný, a ne snadno pochopitelný nástroj. Anotace termínů v něm trvá násobně déle než v Termlt nebo Bratu.

Pro potřeby této práce se jeví jako nejvhodnější nástroj Termlt. Disponuje jednoduchým a uživatelsky přívětivým webovým rozhraním, které je samo vysvětlující a umožňuje uživateli se v něm rychle a snadno orientovat. Mezi jeho hlavní funkce patří snadné provádění anotace pojmů přímo v importovaných dokumentech, klasifikace vybraných pojmů, správu slovníků a ontologických struktur a v neposlední řadě také provedení automatické textové analýzy.

Ostatní uvedené nástroje disponují daleko rozsáhlejšími funkcemi, které však přesahují zamýšlený rozsah této práce. Z toho důvodu a výše uvedených vlastností byl vybrán nástroj Termlt. Při psaní této práce navíc byla možná komunikace přímo s tvůrci tohoto nástroje, kteří ho upravovali pro potřebu jeho použití na údržbové záznamy.

## 6. Zpracování dostupných dat

Tato kapitola je zaměřena na metodiku práce s daty z letecké údržby, jejich rozbor, úpravu a další zpracování pomocí nástroje umožňující automatickou textovou analýzu. V první řadě je důležité se s daty seznámit, zjistit co obsahují a které údaje jsou z hlediska této práce nejvíce zajímavé. Za tímto krokem následuje filtrace a úprava dat podle zvolených atributů.

Níže v této kapitole je představeno prostředí softwarového nástroje TermIt, který umožňuje anotaci pojmů, tvorbu slovníků a textovou analýzu. Je potřeba popsat postup používání tohoto programu, podmínky vytváření pojmů a způsob jakým lze data do software importovat.

V samotném závěru je popsána metodika postupu při anotaci dostupných údržbových dat, výběr pojmů a jejich slučování a provedení automatické analýzy.

### 6.1 Dostupná data od společnosti CSAT

Pro potřeby této práce údržbová organizace CSAT poskytla záznamy z prostředí letecké údržby jako podklad pro zjištění možnosti jejich využití při provádění automatické textové analýzy. Vzhledem k důvěrnosti dat nebylo možné je přímo v práci uvést, ani nejsou součástí žádné z její příloh. Veškeré použité názorné ukázky dat mají pouze informativní charakter z hlediska problematiky této práce a byla anonymizovaná tak, aby byla zachována jejich důvěrnost.

Data obsahují veškeré záznamy o provedené údržbě za roky 2017, 2018, 2019 a polovinu roku 2020. Data byla poskytnuta v tabulkové podobě, což umožnilo jednoduchou práci a jejich snadnou editaci v jakémkoliv tabulkovém editoru. V tomto případě však vyhovoval nejvíce tabulkový editor Microsoft Excel, který umožňoval následný export dat ve vhodném formátu.

Každý řádek tabulky představuje jeden záznam z údržby letadla, který je podrobněji a systematicky rozepsán v jednotlivých sloupcích. Ve sloupcích nalezneme registrační značku letadla, specifické číslo pracovního příkazu („Workorder number“), druh a stav záznamu, číslování dle normy ATA 100, odkaz na technologickou kartu, popis pracovního příkazu, popis provedené činnosti a další.

Jelikož nás však zajímají pouze nálezové záznamy závad, tak je nejprve potřeba data filtrovat a upravit pro další použití. O tom je pojednáno v následující kapitole.

## 6.2 Filtrace a úprava dat

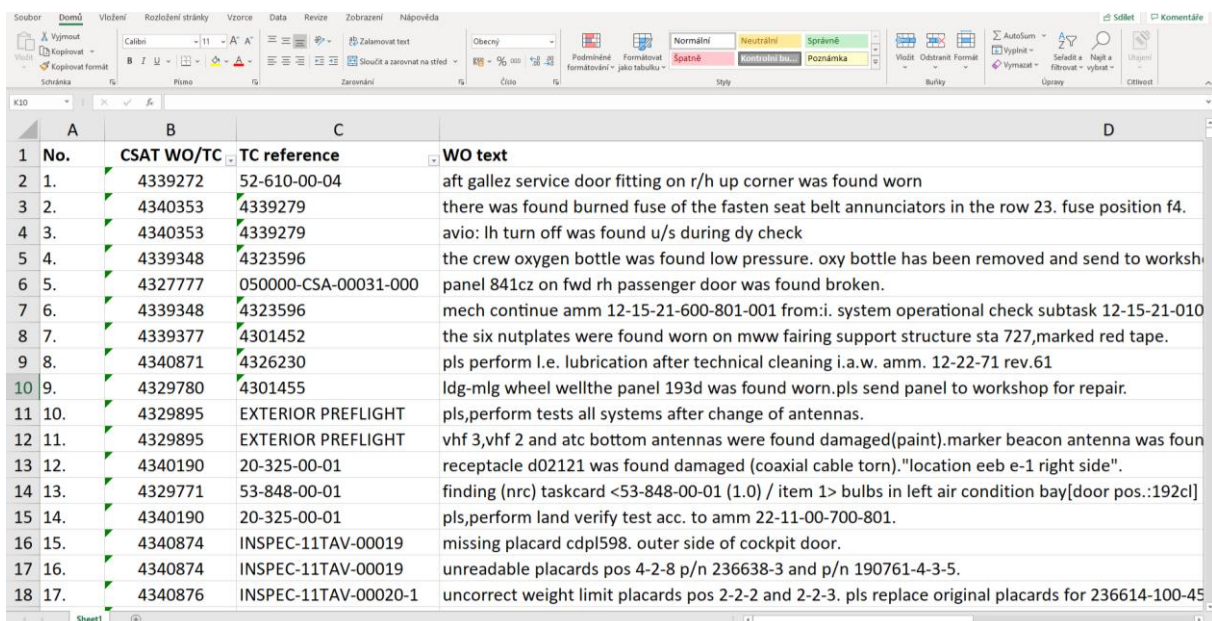
Poskytnutá data v sobě zahrnují celou řadu záznamů o provedené údržbě. Jelikož je práce zaměřená pouze na nálezné záznamy závad, tak je nejdříve potřeba provést filtraci těchto záznamů a jejich následnou úpravu na požadovaný tvar.

Všechny tabulkové editory v sobě většinou zahrnují funkci filtr, pomocí které lze provést jednoduché třídění dat. V tomto případě byl použit na sloupec „type“, který definuje druh záznamu. Z tohoto sloupce byly vybrány pouze záznamy s označením M (Maintenance), které představují neočekávané nálezy závad při údržbě. Data s jiným označením jsou např. záznamy o provedené plánované údržbě (S) nebo vykonané technologické kartě (TC).

Dalším krokem bylo odstranění nepotřebných sloupců s údaji o údržbě. K následujícímu zpracování dat pomocí automatické textové analýzy byly zachovány sloupce CSAT WO/TC, TC reference a WO text.

- CSAT WO/TC – specifické identifikační číslo pracovního příkazu, které se vztahuje právě k jedné provedené práci nebo nálezu závady na letadle
- TC reference – odkaz na technologickou kartu, podle které byla prováděna prohlídka na letadle a byla při ní nalezena závada
- WO text – slovní popis provedené činnosti na letadle nebo objevené závady v podobě nestrukturovaného textu

Výsledná data mají následující podobu (Obrázek 10)



No.	CSAT WO/TC	TC reference	WO text
1.	4339272	52-610-00-04	aft gallez service door fitting on r/h up corner was found worn
2.	4340353	4339279	there was found burned fuse of the fasten seat belt annunciators in the row 23. fuse position f4.
3.	4340353	4339279	avio: lh turn off was found u/s during dy check
4.	4339348	4323596	the crew oxygen bottle was found low pressure. oxy bottle has been removed and send to worksh
5.	4327777	050000-CSA-00031-000	panel 841cz on fwd rh passenger door was found broken.
6.	4339348	4323596	mech continue amm 12-15-21-600-801-001 from:i. system operational check subtask 12-15-21-010
7.	4339377	4301452	the six nutplates were found worn on mww fairing support structure sta 727,marked red tape.
8.	4340871	4326230	pls perform i.e. lubrication after technical cleaning i.a.w. amm. 12-22-71 rev.61
9.	4329780	4301455	ldg-mlg wheel wellthe panel 193d was found worn.pls send panel to workshop for repair.
10.	4329895	EXTERIOR PREFLIGHT	pls,perform tests all systems after change of antennas.
11.	4329895	EXTERIOR PREFLIGHT	vhf 3,vhf 2 and atc bottom antennas were found damaged(paint).marker beacon antenna was foun
12.	4340190	20-325-00-01	receptacle d02121 was found damaged (coaxial cable torn)."location eeb e-1 right side".
13.	4329771	53-848-00-01	finding (nrc) taskcard <53-848-00-01 (1.0) / item 1> bulbs in left air condition bay[door pos.:192cl]
14.	4340190	20-325-00-01	pls,perform land verify test acc. to amm 22-11-00-700-801.
15.	4340874	INSPEC-11TAV-00019	missing placard cdp1598. outer side of cockpit door.
16.	4340874	INSPEC-11TAV-00019	unreadable placards pos 4-2-8 p/n 236638-3 and p/n 190761-4-3-5.
17.	4340876	INSPEC-11TAV-00020-1	uncorrect weight limit placards pos 2-2-2 and 2-2-3. pls replace original placards for 236614-100-45

Obrázek 10 - Filtrovaná a upravená data



## 6.3 Práce s nástrojem Termlt

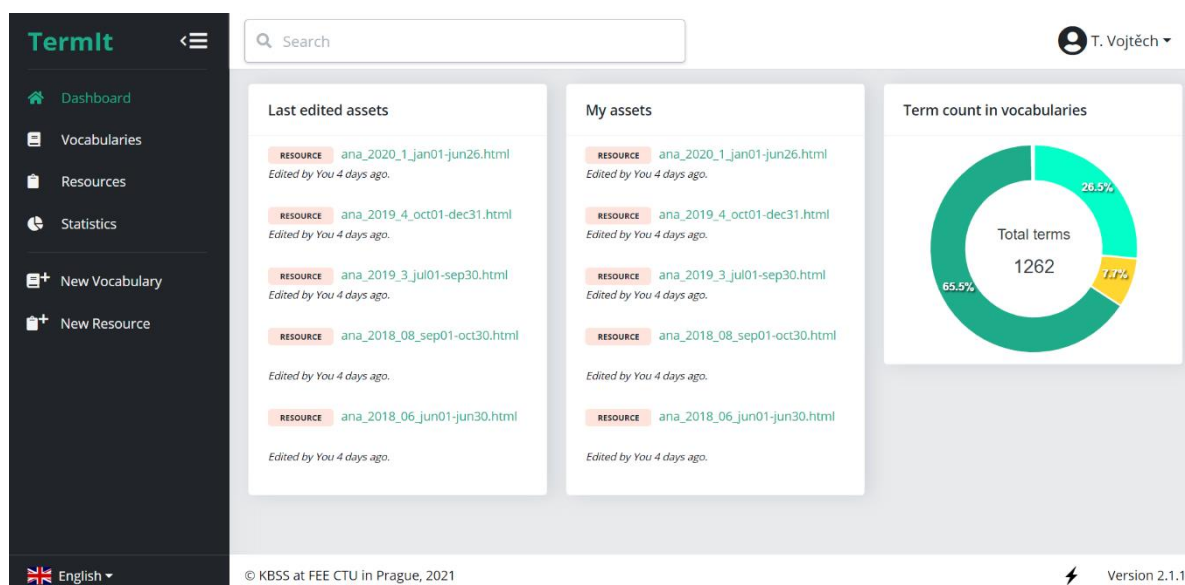
V následujících podkapitolách je přibližena metodika práce s nástrojem Termlt. Je zde popsáno, jakým způsobem je možné se v nástroji orientovat, jak vytvářet slovníky a pojmy a jak s nimi dále pracovat, co je potřeba provést pro úspěšný import souborů a další důležité poznatky týkající se používání nástroje.

### 6.3.1 Uživatelské prostředí Termlt

K přihlášení a práci s nástrojem dochází pomocí běžně dostupných webových prohlížečů na příslušném internetovém odkaze<sup>2</sup>. Po úspěšném přihlášení se uživateli otevře úvodní obrazovka (Obrázek 11), na které je možné si prohlédnout naposledy provedené změny v nástroji, např. nově přidané pojmy nebo zdroje, upravené pojmy apod. Je zde také ukazatel počtu pojmů v jednotlivých slovnících.

Na levé straně nástroje se nachází uživatelské menu, pomocí kterého se lze přesunout do hlavních funkcí programu. Je zde umožněno přímo vytvořit nový slovník nebo datový zdroj, ale hlavně je odtud možné se přesunout na stávající seznam s již vytvořenými Slovníky (Vocabularies) a Zdroji (Resources). Na závěr je vhodné uvést možnost přepnutí nástroje mezi českým a anglickým jazykem.

Nástroj Termlt byl vydán již v několika verzích a je pod neustálým vývojem. Vývojáři pravidelně pracují na řešení a opravě chyb, přidávání nových funkcí a zdokonalování a vyladování nástroje.



Obrázek 11 – Úvodní obrazovka nástroje Termlt

<sup>2</sup> Odkaz na veřejně přístupnou verzi nástroje Termlt - <https://kbss.felk.cvut.cz/termit-dev/#/login>



## 6.3.2 Slovníky a pojmy

Pomocí záložky „Vocabularies“ je možné se přesunout do správce slovníků. Tady je uživateli umožněno vytvářet nové slovníky anebo spravovat ty stávající. Pro vytvoření nového slovníku stačí kliknout na tlačítko „+ New Vocabulary“ (Přidat nový slovník), kde je po uživateli vyžádáno pouze zadání názvu nového slovníku. Dále pak může volitelně přidat krátký popis nebo také ke slovníku připojit jeden nebo více dokumentů, ze kterých bude čerpat termíny. Slovník je následně vytvořen a zobrazí se ve stávajícím přehledu mezi ostatními slovníky. Po jeho rozkliknutí se uživatel přesune do editace, kde se mu zobrazí další funkce, které může ve slovníku provádět.

Hlavním účelem slovníků je v sobě shromažďovat a uchovávat pojmy, jejich význam a definice nebo popis termínů, které se týkají jeho zaměření a pokrývají oblast dané problematiky. Slovníky mohou být např. naučné, překladové, jazykové apod. V případě této práce se jedná o technický slovník pojmů objevujících se v záznamech z letecké údržby.

Po otevření příslušného slovníku se uživateli zobrazí seznam stávajících pojmů, funkce pro vytvoření nového pojmu, nástroje pro editaci slovníku a připojení dokumentů a další (Obrázek 12).

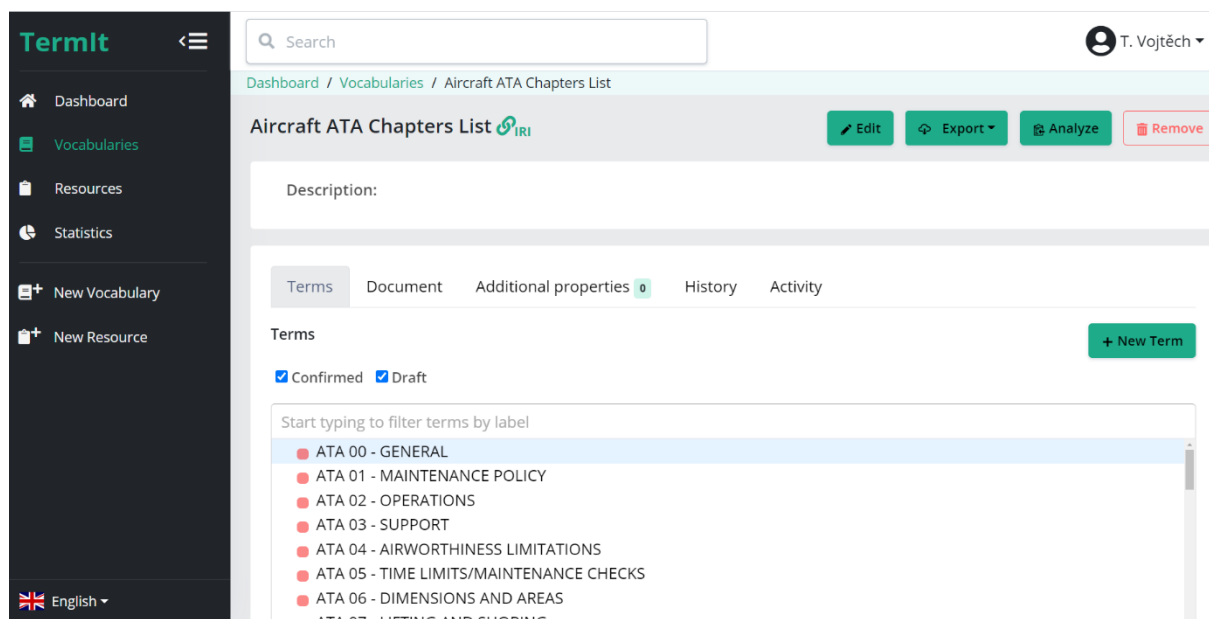
Pro vytvoření nového pojmu je potřeba kliknout na funkci „+ New Term“, která uživatele přesune na stránku pro založení nového termínu. Zde je od uživatele vyžadováno vyplnění příslušných specifikací, které budou pojem definovat. Jedná se především o název pojmu, který je povinný, ostatní položky jsou volitelné.

Přehled nastavitelných položek, kterými lze pojmy definovat a rozdělit:

- Název – preferované označení pojmu, kterým bude ve slovníku jednoznačně identifikován (povinné)
- Synonyma – slova se stejným nebo podobným významem jako je hlavní název pojmu a v určitých případech se používají místo něho
- Definice
  - Text – popisuje daný termín a udává jeho význam
  - Zdroj – odkaz na původ definice
- Doplňující poznámka – text, který neovlivňuje definici pojmu, pouze upřesňuje význam
- Nadřazené pojmy – zachycení vazby na obecnější pojem (hierarchická struktura)
- Typ pojmu – charakter pojmu, kterým lze rozlišit mezi typem (typ události, typ vztahu) a instancí (událost, vztah)
- Vyhledávací texty – slouží zejména pro vyhledávání pojmů, nejsou určeny pro vizuální prezentaci dat jako je hlavní název

Pro potřeby této práce byly použity pouze následující nastavitelné položky:

- Název – hlavní označení komponenty nebo selhání např. bearing („ložisko“), worn („opotřebený“)
- Synonyma – jiná slova používaná pro danou komponentu nebo selhání
- Vyhledávací texty – použito pro nalezené gramatické chyby a překlepy
- Nadřazené pojmy – rozlišení pojmů na komponenty a selhání („Component / Failure“)



Obrázek 12 - Vybraný slovník v nástroji Termlt

### 6.3.3 Import dat do nástroje

Aby bylo možné data v nástroji Termlt použít, je nejdříve potřeba soubory upravit na požadovaný datový formát. V tomto případě se jedná o formát typu html. V kapitole 6.2 Filtrace a úprava dat již bylo popsáno jakým způsobem došlo ke třídění a formátování dat na požadovaný tvar, a proto v následujícím odstavci je popsán převod souborů s těmito daty na soubory, které jsou podporované nástrojem Termlt.

V tabulkovém editoru, ve kterém byla data upravena je možné zvolit funkci „Uložit soubor jako“, kde po nastavení příslušných atribut je soubor s daty exportován do požadovaného formátu. Těmito atributy je např. zvolení přípony souboru .html, export pouze jednoho listu souboru nebo také nastavení kódování znaků, aby byla zajištěna kompatibilita.

Anotační nástroje tohoto typu mají většinou požadovaný přesně daný vstupní formát souboru. V případě nástroje Termlt je to HTML. Export dat byl proveden pomocí Microsoft Excel, který umožňuje export souborů i do dalších datových formátů pro potřeby jiných anotačních nástrojů.

V uživatelském prostředí Termlt se k nahrávání dat používá funkce pro přidání nového zdroje („+ New Resource“). Uživatel je vyzván k zadání názvu nového zdroje nebo datové sady, který bude jednoznačným identifikátorem. Následně pomocí průzkumníka souborů dochází k výběru a nahrání dat z počítače. Pokud je import dat úspěšný, tak se nový zdroj zobrazí v seznamu zdrojů a je s ním možné dále pracovat.

Uživatel je nyní schopen si zobrazit v nástroji všechny importované soubory a může začít provádět jejich anotaci, ze které bude vznikat slovník pojmů v těchto dokumentech. Dalším krokem je možnost spouštění automatické textové analýzy.

## 6.4 Anotace

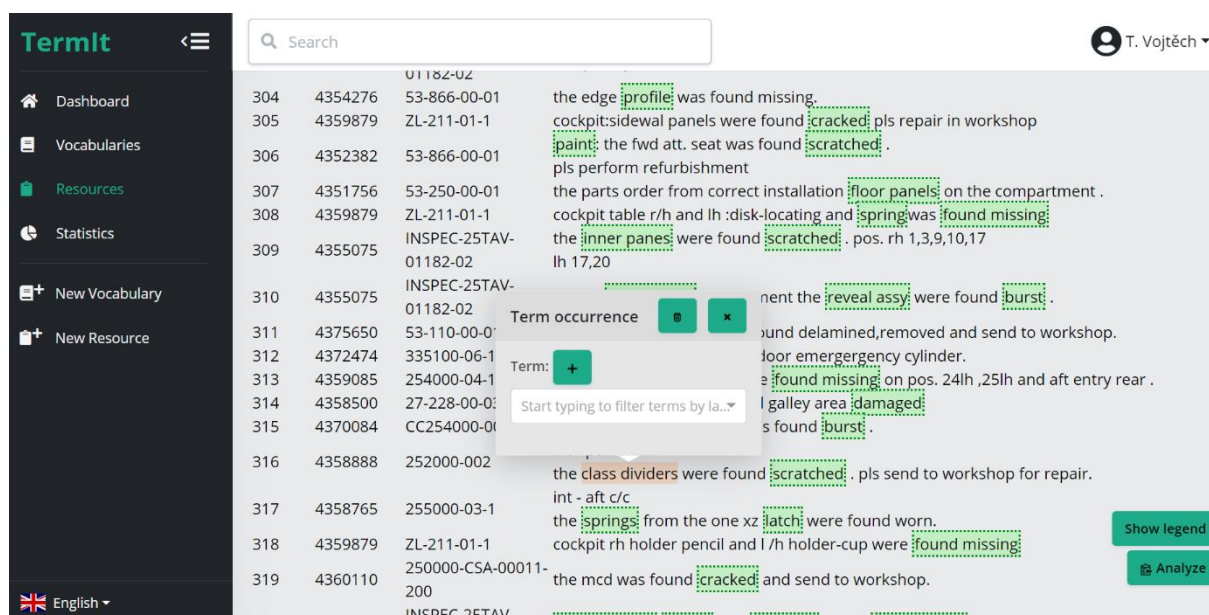
Vytváření pojmů v nástroji Termlt lze provádět dvěma možnými způsoby. Obě varianty představují manuální zadání požadovaného termínu do slovníku, který pak bude použit pro následnou automatickou textovou analýzu. Nové pojmy lze přidávat ve zvoleném slovníku nebo při procházení importovaného dokumentu. Obě varianty mají své výhody i nevýhody.

Případ, kdy jsou pojmy přidávány ve zvoleném slovníku byl již popsán v kapitole 6.3.1 Slovníky a pojmy. Výhodou je, že termíny lze vytvářet i bez přiřazeného dokumentu a uživateli je tak umožněno dopředu si připravit slovník pojmů, dle jeho požadavků. Ten bude možné následně aplikovat na jakýkoliv importovaný dokument, u kterého bude chtít provést automatickou textovou analýzu. Naopak nevýhodou je, že uživatel musí dopředu vědět, jaké výrazy do slovníku zařadit a v jaké podobě se v dokumentech budou vyskytovat.

Druhou možností, jakou lze anotaci provádět, je vytváření pojmů přímo při procházení importovaného dokumentu (Obrázek 13). Uživatel vybere možnost obsah („Content“) a dojde k zobrazení vybraného souboru, ve kterém je možné provádět manuální výběr termínů přímo v textu. Požadované pojmy jsou vybrány a označeny pomocí kurzoru, kde následně dojde k zobrazení tabulky možností. Zde uživatel vybere možnost přidat nový pojem a pokračuje dle zobrazených instrukcí. Název pojmu je automaticky zvolen stejný, jako pojem označený ručně v textu, a v případě potřeby je možné vyplnit i další nastavitelné položky (resp. Synonyma, Vyhledávací texty a Nadřazené pojmy), které byly podrobně popsány v předcházející kapitole.

V rámci této práce byla provedena anotace poměrného množství dostupných dat (přibližně 3500 záznamů), která sloužila jako podkladový zdroj pro vytvoření slovníku. Data byla postupně procházena po jednotlivých záznamech, a přitom byly manuálně označeny specifické komponenty a selhání. Je předpokladem, že jeden záznam bude obsahovat právě

jednu specifickou komponentu a selhání. Pokud se v jednom záznamu objeví více komponent, bude se jednat pouze o upřesnění umístění dané závady.



Obrázek 13 - Anotace pojmů v dokumentu

### 6.4.1 Podmínky pro vytváření pojmů

Je předpokladem, že bude potřeba provádět kontinuální udržování slovníku a kontrolu jeho aktuálnosti. V údržbové organizaci tuto činnost může provádět více osob, a proto byly z tohoto důvodu zavedeny podmínky pro vytváření nových pojmů. Mělo by tím být zajištěno, že bude docházet ke tvorbě jednotvárného a organizovaného slovníku a pojmy budou definovány přesně tak, jak to bude vyhovovat údržbové organizaci s ohledem na funkčnost nástroje TermIt. Je potřeba zajistit, aby všechny pojmy byly vytvořeny stejným způsobem a měly předepsaný tvar.

Tyto podmínky jsou důležité i z hlediska efektivního fungování nástroje a provádění automatických textových analýz. Jelikož jsou všechny záznamy z letecké údržby psané anglickým jazykem, tak jsou tyto podmínky pro tuto skutečnost náležitě upraveny.

Nástroj TermIt byl původně vyvinut pro český jazyk, avšak po projednání s vývojáři došlo k jeho menším úpravám, aby ho bylo možné použít i pro anglické výrazy. Výsledkem konzultací jsou následující obecné podmínky pro zadávání pojmů:

#### Název („Label / Preferred Label“)

- hlavní slovo nebo fráze, která bude použita pro identifikaci termínu
- vždy psáno s velkým počátečním písmenem, zbytek malá písmena

### Synonyma („Synonyms / Alternative Label“)

- ostatní slova nebo fráze, které se vztahují k danému termínu
- synonyma, slova se stejným nebo podobným významem používaná pro daný termín
- zkráceniny slov nebo frází (např. Ident. plate, part num., apod)
- zkratky slov (např. MLG – Main Landing Gear, ADM – Air Data Module apod.)
- slova, která jsou psána jinak, ale mají stejný význam (např. Collectible / Collectable)
- nepravidelná množná čísla (např. Person / People)

### Vyhledávací texty („Search strings / Hidden Label“)

- názvy, které by se neměly graficky reprezentovat (nezobrazovat se), ale dále slouží pro funkci vyhledávání pojmu
- překlepy a gramatické chyby (např. worn / wonr, damage / demage, apod.)
- zastaralé výrazy pro daný termín
- vše, co nespadá do předchozích kategorií

V následující tabulce (tabulka 1) jsou uvedeny tři případy pojmů zadaných do nástroje Termlt na základě výše uvedených podmínek.

*Tabulka 1 - Příklady pojmů v nástroji Termlt*

<b>Label (Preferred Label)</b>	<b>Synonym (Alternative Label)</b>	<b>Search String (Hidden Label)</b>
Identification plate	Identification label, ident. plate	name plate
Inoperative	Inop, Faulty, Malfunction	not working, incorrect function
Pressure transfer unit	PTU	-

## 7. Slovník komponent a závad

V této části práce je představen slovník pojmů z prostředí letecké údržby, který byl vytvořen za účelem toho, aby bylo možné provádět automatickou textovou analýzu údržbových záznamů. Jak už bylo zmíněno dříve, definiční slovník pojmů tvoří nezbytný základ pro provádění textových analýz. Tvůrce slovníku je schopen si do něj zařadit jen ty pojmy, které ho z určitého důvodu zajímají a jsou pro něj důležité.

Pro potřebu této práce byl vytvořen slovník, který byl nazván „Slovník komponent a závad“. Název slovníku vystihuje jeho obsah a je možné v něm nalézt jednotlivé komponenty letadel nebo jejich závady, resp. poškození, které bylo nalezeno při inspekci letadel a objevilo se v záznamech o údržbě.

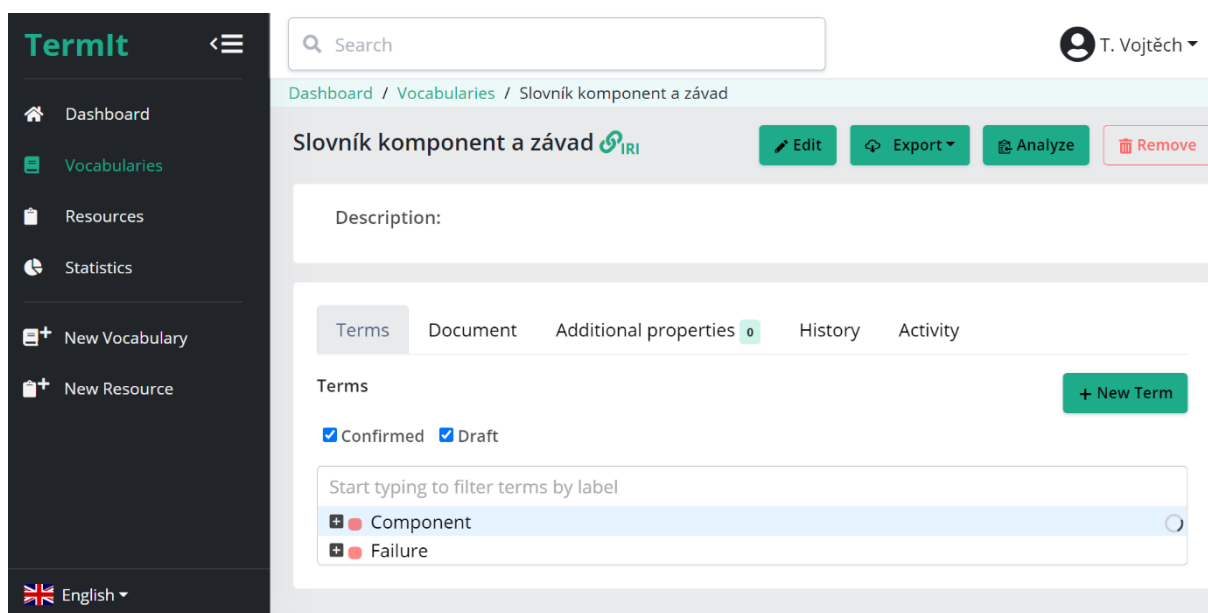
V následujících kapitolách je popsáno, jakým způsobem slovník vzniknul, co je jeho obsahem včetně názorných příkladů a jaká je jeho použitelnost a rozsah zpracování.

### 7.1 Vznik slovníku

Tato kapitola je zaměřena na způsob, jakým byl slovník postupně formulován, jeho rozdělení, organizaci a jaké pojmy je možné v něm nalézt. Postup, jakým je uživatel schopen vytvořit nový slovník a pojmy v nástroji TermIt byl popsán výše, v kapitole 6.3.2 Slovníky a pojmy, a tento proces již nebude znovu popisován.

Slovník komponent a závad vzniknul na základě manuálního procházení dostupných dat z letecké údržby. Tyto data představují záznamy nalezených neplánovaných závad zjištěných při provádění plánované údržby letadel. Poměrná část těchto záznamů byla řádek po řádku manuálně zpracována a byl proveden rozbor jednotlivých zápisů z hlediska vyskytujících se letadlových komponent a příčin jejich selhání. Procházení dat bylo provedeno v nástroji TermIt, který umožňuje provádění anotace pojmů přímo v importovaných dokumentech. To výrazně usnadňuje práci se záznamy a přispívá tak k jednoduché tvorbě referenčního slovníku.

Pro potřebu kategorizace jednotlivých pojmů ve slovníku byly vytvořeny dva hlavní nadřazené pojmy. Jedná se o pojem Komponenta („Component“) a pojem Závada („Failure“) (Obrázek 14). Do těchto dvou kategorií byly při anotaci údržbových záznamů rozřazovány vybrané termíny, a to hned ve fázi vytváření nových pojmů. Nadřazené pojmy od sebe jednotlivé výrazy rozlišují právě dle zvolené kategorie. V prvním případě se pojem Komponenta vztahuje výhradně na letadlové součásti (komponenty) nebo letadlové celky a v případě pojmu Závada se jedná o poruchový stav nebo selhání zjištěné při prohlídce letadla. Podrobněji jsou tyto případy popsány v následujících kapitolách, které se věnují jednotlivým případům.



Obrázek 14 - Slovník komponent a závad v nástroji Termlt

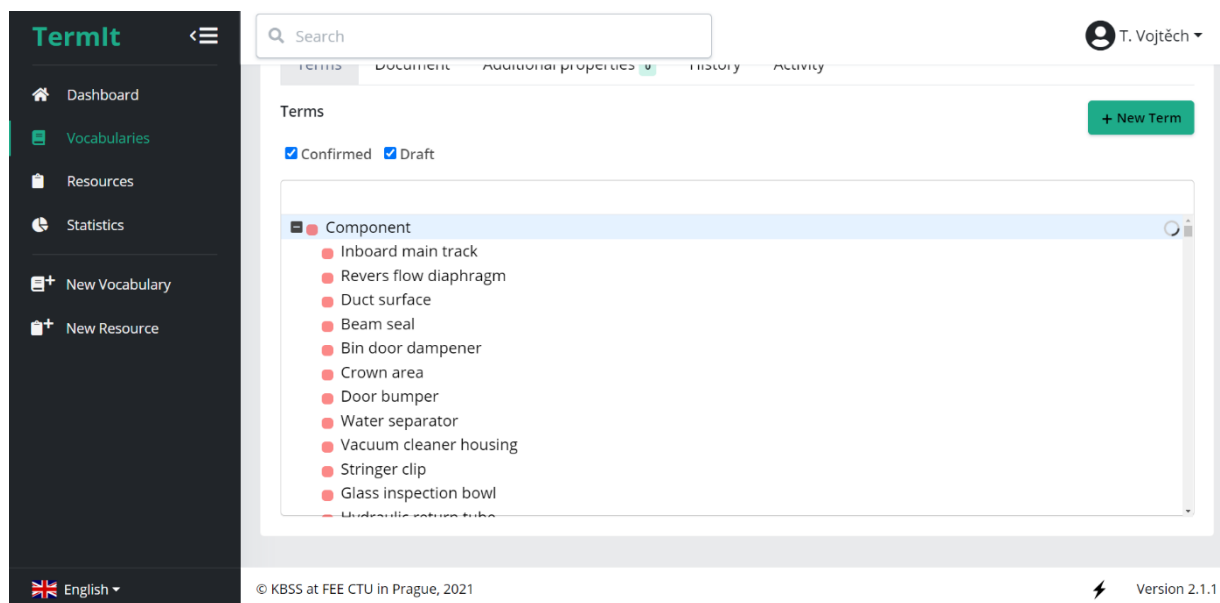
## 7.2 Komponenty

Ve Slovníku komponent a závad je možné nalézt nadřazený pojem komponent („Component“). Komponent, resp. komponenty mohou představovat jednotlivé díly, součásti nebo systémy letadla, anebo letadlové celky. Během údržby procházejí jednotlivé komponenty letadla inspekcí a pokud dojde k nálezů závady, tak jsou zaznamenány do systému údržbové organizace k dalšímu zpracování a naplánování opravy. Ze záznamů nálezů závad byl pomocí nástroje Termlt proveden manuální výběr těch komponent, které byly na letadle nalezeny poškozené nebo s nevyhovujícím stavem. Tyto komponenty představují seznam pojmů ve slovníku a tvoří podklad pro provedení automatické textové analýzy. Umožňují nalezení výskytu dané komponenty ve všech zkoumaných záznamech z údržby, na které bude textová analýza aplikována.

V nástroji Termlt lze pojem „Component“ pomocí kurzoru myši rozkliknout a následně se zobrazí seznam výrazů, které jsou k tomuto pojmu přiřazeny (Obrázek 15). Protože je letadlo komplexní stroj, tak mezi těmito pojmy je možné nalézt velký počet rozdílných letadlových komponent. Lišit se mohou velikostí, funkcí, jestli se jedná o celek nebo jeho součást, anebo zda se jedná jen o vybavení letadla. Protože hledáme pouze poškozené komponenty nalezené při údržbě, tak se v tomto slovníku mohou objevit různorodé termíny. Od jednotlivých šroubů a nýtů, přes nejrůznější těsnění, ložiska, uchycovací prvky, pohyblivé mechanismy až po velké komponenty jako jsou krycí panely, nosníky nebo součásti konstrukce podvozku, pohyblivých ploch apod.

Příklady komponent ve Slovníku komponent a závad:

- Ring sealing („kroužkové těsnění“)
- Control valve („ovládací ventil“)
- Antichafing plate („deska proti prodření“)
- Air intake flap actuator („aktuátor klapky přívodu vzduchu“)
- Access panel („přístupový panel“)
- Shock absorber („tlumič“)
- Vertical stabilizer („směrovka“)



Obrázek 15 - Pojem Component v nástroji Termt

## 7.3 Závady

Obdobně jako bylo popsáno v předchozí kapitole, tak ve Slovníku komponent a závad můžeme nalézt druhý nadřazený pojem, kterým je Závada („Failure“). Pojmy, které jsou pod tímto termínem zařazeny představují zjištěné selhání dané komponenty při provádění inspekční kontroly letadla. Tato selhání se stejně jako příslušná komponenta objeví v záznamech o údržbě, které vyplnil příslušný mechanik na základě jeho zjištění. Technik zaznamená technický stav komponenty, její poruchu nebo poškození. Na dané komponentě může být například zjištěno, že došlo k nálezu trhliny, koroze, opotřebení mimo limit a další.

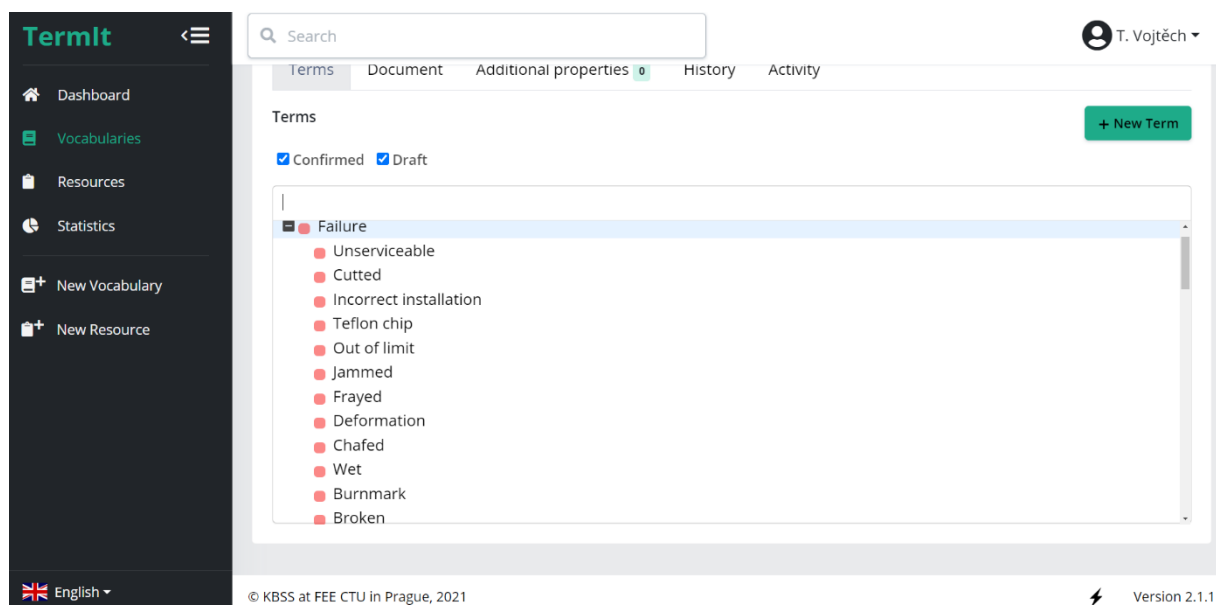
Z použitých výrazů, které se objevily v záznamech, byly vybrány právě ty pojmy označující závady a následně byly zařazeny do této části slovníku. Pojmy tvoří podklad pro automatickou textovou analýzu z hlediska identifikace nejčastěji se vyskytujících závad zjištěných při údržbě letadla.



Stejným způsobem jako v předchozím případě lze i pojem „Failure“ rozkliknout pomocí kurzoru myši a dojde k zobrazení seznamu výrazů, které jsou k tomuto pojmu přiřazeny (Obrázek 16). V obou případech je možné s těmito seznamy pracovat a upravovat je dle potřeby uživatele.

Příklady závad ve Slovníku komponent a závad:

- Corrosion („koroze“)
- Worn („opotřebované“)
- Dent („vryp“)
- Leakage („únik“)
- Cracked („prasklina“)
- Unserviceable („neprovozoschopný“)
- Torn („roztržený“)



Obrázek 16 - Pojem Failure v nástroji TermIt

## 7.4 Použitelnost a rozsah zpracování

Slovník komponent a závad v sobě obsahuje pojmy, které se objevily v záznamech z údržby letadel typu Boeing 737 a Airbus A320. Z tohoto důvodu je vhodnější aplikovat automatickou textovou analýzu pomocí toho slovníku na dané typy letadel s ohledem na jejich těžkou údržbu na technické základně, protože se pojmy v něm budou týkat přímo těchto strojů a daného typu údržby. Není však vyloučené jeho použití na jiný typ letounů nebo také na jiný typ údržby. S největší pravděpodobností bude ale nutné tento záměr nejdříve vyzkoušet a otestovat na vzorových datech.

Vytvořený slovník je součástí přílohy této závěrečné práce. Jedná se o jeho elektronickou podobu, která byla exportována přímo z nástroje TermIt. Soubor obsahuje tabulku se všemi vytvořenými pojmy (Obrázek 17), kde jeden řádek tabulky představuje jeden pojem. Tabulka má několik sloupců, které obsahují jednotlivé vyplněné specifikace u daného termínu a reflektuje tak výše zmíněné podmínky pro zadávání pojmů. Jedná se především o Název („Label / Preferred Label“), Synonyma („Alternative label / Synonyms“), Vyhledávací texty („Hidden Labels / Search strings“) a Nadřazené pojmy („Parent term“). Slovník je možné otevřít v libovolném tabulkovém editoru, kde je možné s ním pracovat, vyhledávat pojmy nebo ho editovat.

1	Label	Alternative Labels	Hidden Labels	Parent term
2	APU compartment		apu compartment	Component
3	ATC control panel			Component
4	Access cover			Component
5	Access panel			Component
6	Acoustic panel		accoustical panel;acoustical panel	Component
7	Actuator			Component
8	Actuator assembly		actuator assy	Component
9	Adult life vest	Adult vest;Life jacket		Component
10	Aero blocker door seal			Component
11	Aerodynamic seal			Component
12	Aft rod			Component
13	Aft upper engine cowl			Component
14	Aileron trim			Component
15	Aileron trim actuator			Component
16	Air condition pack			Component
17	Air conditioning computer light		air cond compt lt	Component
18	Air cooling hose			Component
19	Air data module	ADM		Component
20	Air duct hose			Component
21	Air intake flap actuator			Component
22	Air outlet			Component
23	Alternate and gear retract brake system			Component
24	Aluminium plate			Component
25	Aluminum foil marker			Component
26	Anchorage fitting			Component
27	Angle		angels	Component
28	Angle bracket		bracket-angles	Component
29	Antenna			Component
30	Anti siphon valve		anti siphon-valve	Component
31	Anti-icing tube		tube anti ising	Component
32	Antichafing plate	Antichafing strip		Component
33	Antiskid harness			Component
34	Antiskid tape		antsikid tapes	Component
35	Arcing			Failure
36	Asna seal		asna seals	Component
37	Attach hole			Component
38	Attendant seat			Component

Obrázek 17 - Slovník komponent a závad

Slovník byl sestaven na základě manuálního procházení vybraného vzorku dostupných dat, který tvořil přibližně 3500 záznamů. Jednalo se o údržbové záznamy nálezů neplánovaných závad přibližně za období 3 měsíců. U tohoto datového setu byla řádně provedena anotace a tím došlo ke vzniku slovníku. Slovník obsahuje dohromady více než osm set pojmů, kde termíny pro komponenty tvoří majoritní podíl. Termíny pro závady představují jen určité procento z celkového počtu všech pojmů ve slovníku. Je to dané tím, že komponentů

na letadle je několikanásobně více než možných výrazů pro závady, a navíc některé výrazy pro závady byly sloučeny do jednoho pojmu z důvodu jejich stejného významu. Celou věc si lze představit tak, že komponent je jedinečný prvek na letadle, u kterého se může projevit různá závada.

Rozsah zpracování a kompletnost slovníku odpovídá termínům, které se objevily ve zpracovaných vzorových datech. Je pravděpodobné, že slovník nebude zcela úplný, ale vzhledem k provedení anotace přibližně 3500 záznamů je předpokladem, že slovník již obsahuje značnou část pojmů vyskytujících se v záznamech.

Slovník je volně dostupný a může posloužit jako podklad pro další použití. Je předpokladem, že bude dále rozšiřován a doplňován o další pojmy a následně sloužit jako podklad pro další provádění automatických textových analýz. Vytvoření tohoto slovníku představovalo jeden z dílčích úkolů této závěrečné práce a je součástí její přílohy.

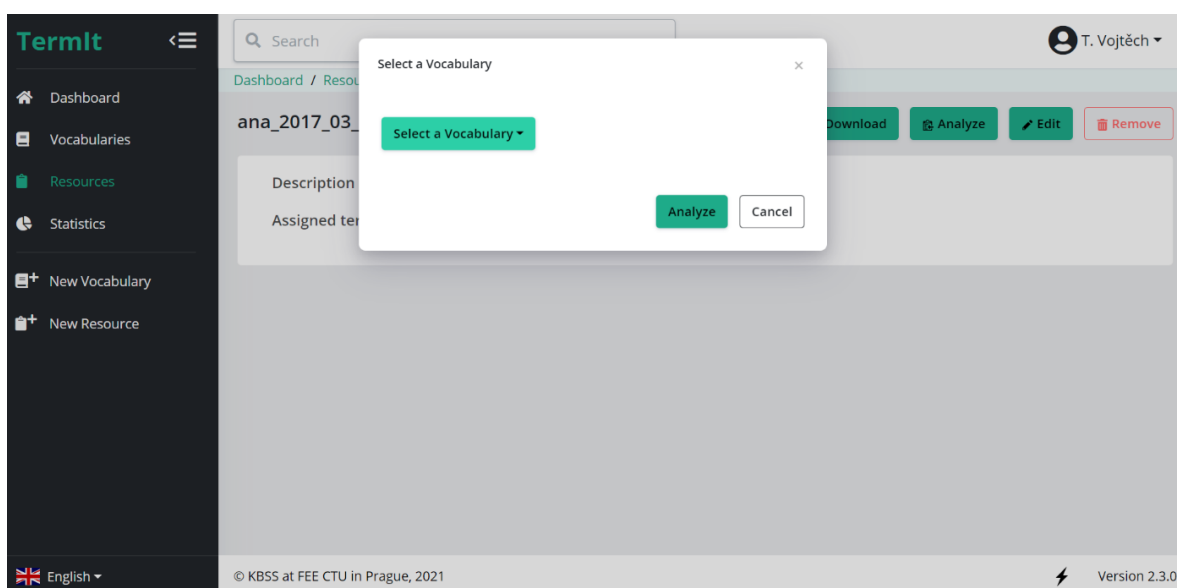
## 8. Provedení a vyhodnocení automatické textové analýzy

V této části práce je popsán postup, jakým byla provedena automatická textová analýza, její spuštění, nalezení a označení výskytu pojmů, včetně jejich vizuální reprezentace. Dále je v kapitole uvedena podoba výstupu z nástroje Termlt, se kterým je možné dále pracovat, vyhodnocovat výsledky textové analýzy a sestavovat statistiky výskytu pojmů.

### 8.1 Spuštění automatické textové analýzy

Spuštění automatické textové analýzy lze provést dvěma možnými způsoby. Jedním z nich je funkce *analyzovat*, která je k dispozici přímo v otevřeném dokumentu, se kterým uživatel momentálně pracuje. Po kliknutí na tlačítko „Analyze“ (Analyzovat) se zobrazí okno, ve kterém uživatel provede volbu vybraného slovníku, kterým chce daný dokument analyzovat. Tato volba se však nezobrazí, pokud je daný dokument již přiřazen k příslušnému slovníku.

Druhou možností, jakou lze spustit automatickou textovou analýzu, je přistoupit k dokumentům přes Datové zdroje nástroje Termlt. Vybere se příslušný soubor nebo se do nástroje nahraje a spuštění analýzy se provede pomocí tlačítka „Analyze“ (Analyzovat), které je možné nalézt na panelu nástrojů. V tomto případě je vždy uživateli zobrazena možnost volby požadovaného slovníku „Select a Vocabulary“ (Vyberte slovník), kterým má být dokument analyzován a musí jí uživatel potvrdit (Obrázek 18). Tento způsob spuštění nevyžaduje přímé otevření dokumentu a je možné ji spustit bez potřeby zobrazení dat v souboru, které vyžaduje zdlouhavé načítání záznamů. Je vhodný pro uživatele, kteří již nepotřebují anotovat dokument, mají slovník pojmů hotový a chtějí jen spustit automatickou analýzu.



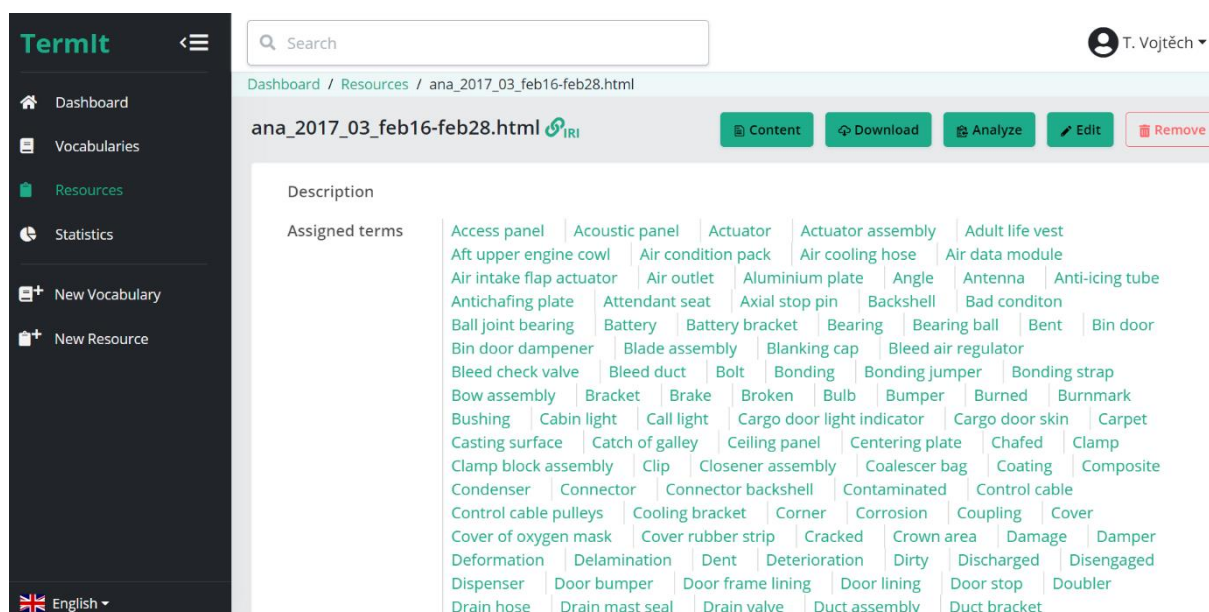
Obrázek 18 - Spuštění automatické textové analýzy

## 8.2 Formy výstupu dokončené textové analýzy

Rozhraní nástroje Termlt umožňuje dva způsoby reprezentace výsledků textové analýzy. Po jejím dokončení je možné si zobrazit seznam pojmů, které textová analýza v daném dokumentu našla nebo si přímo zobrazit daný dokument, ve kterém jsou nalezené pojmy vizuálně označeny a zvýrazněny. Hlavní náplní této práce je však analýza výstupního souboru extrahovaného přímo z databáze nástroje Termlt, který obsahuje další důležité výstupy a informace o provedené analýze. Vyhodnocení analýzy přímo v nástroji nebylo v době tvorby této práce možné.

### 8.2.1 Zobrazení seznamu výskytu pojmů

Při rozkliknutí příslušného analyzovaného dokumentu se zobrazí seznam všech pojmů, které odhalila textová analýza. Tyto pojmy je možné nalézt pod označením „Assigned terms“ (Přiřazené pojmy) a představují výskyt pojmů obsažených ve slovníku, které byly nalezeny při analýze vybraného dokumentu. (Obrázek 19). Tento typ výstupu z nástroje představuje pouze znalost toho, že se určitý pojem ze slovníku nachází ve vybraném dokumentu alespoň jednou.



Obrázek 19 - Výskyt pojmů v dokumentu

### 8.2.2 Vizuální reprezentace dat

Druhou možností, jakou je umožněno uživateli pracovat s výstupem z nástroje, je vizuální reprezentace dat. Po dokončení analýzy je možné si otevřít příslušný dokument a zobrazit si jeho obsah, resp. v případě této práce to jsou jednotlivé záznamy z údržby. Nástroj Termlt

na základě nalezení příslušného pojmu provede jeho označení v textu a zvýrazní ho (Obrázek 20). Uživatel je tedy hned schopný vidět, zda byly v daném dokumentu nalezeny ty pojmy, které si definoval ve slovníku a jsou pro něho klíčové. Pokud výstup nesplnil očekávání uživatele, je potřeba se přesunout zpět do slovníku, upravit jeho strukturu, jednotlivé pojmy anebo ho doplnit o výrazy, které se ve slovníku ještě neobjevily a následně znovu provést automatickou textovou analýzu.

TermIt	Hledat	T. Vojtěch
Hlavní strana	19. 4351019 4350819	item.2-ext-the lead bonding on toilet service door was found cracked.
Slovníky	20. 4351019 4350819	item.2-ext the lead bonding on water service door was found cracked.
Datové zdroje	21. 4351538 4343699	ldg:external leak was found on the yellow hydr. hp filter bowl.
Statistiky	22. 4344174 4343698	oring seal was found visible deteriorated after filter bowl removal.
Nový slovník	23. 4351538 4343699	ldg: the streamers of nlg and mlg ground lock pins and sleeves were found in a poor condition.
Nový zdroj	24. 4346687 4343699	perform leak check ref amm subtask 29-13-14-710-050-a and subtask 29-13-45-790-050-a.
	25. 4346687 4343699	int: during preweekly check megaphone was found unserviceable.
	26. 4346687 4343699	int: during preweekly check 2pcs loop belts were found missing.
	27. 4346687 4343699	int: during preweekly check 3pcs crew life vests were found torn.
	28. 4351005 4350715	the cockpit fire extinguisher was found unsealed. (seal was found missing)
	29. 4350932 4343699	emergency floor proximity light between row 15-20 inop. inverter was found faulty.
	30. 4345562 ZL-575-01-1	ext - fuel water drain valve was found leaking in blue hyd compartment.
	31. 4351591 4351335	ext. the panel out fixed trailing edge (r/h aileron) was found deteriorated to screw hole.
	32. 4348888 110000-CSA-00001-000	the shade last h sidewall was found missing
	33. 4346713 110000-CSA-00002-000	the placards on aft cargo door and their door selector were found unreadable and poor condition.
	34. 4346713 110000-CSA-00002-000	the placards in cockpit on f/o oxygen panel, thirth occupant seat and lg pins stowage were found unreadable.
	35. 4343727 323100-05-1	ldg: the identification plates of nlg main and aft doors were found missing.
	36. 4348888 110000-CSA-00001-000	the placard on inlet to air condition pack lh was found in poor condition.
	37. 4346713 110000-CSA-00002-000	the placards on g 1 were found missing and poor conditiot (box 101,103,131,171)
	38. 4346713 110000-CSA-00002-000	the placards (no cigarette disposal) on lav a and d were found in very poor condition.
	39. 4343697 323100-05-1	ldg: the markers of nlg door operating mechanism were found missing or damaged.
	40. 4343688 4343653	int: several exterior placards from passenger and cargo comp. doors were found unreadable.
	41. 4346712 4343820	the placards were found deteriorated: slat no.3 on l/h wing hoist.point slat extension angle on l/h wing inb. flap on l/h wine hoist.point

Obrázek 20 - Označení nalezených pojmů

### 8.2.3 Export z databáze nástroje TermIt

Na základě požadovaných filtrů byl proveden export výsledků automatické textové analýzy údržbových záznamů. Z databáze nástroje TermIt byl vygenerován výstupní soubor, který má podobu tabulky formátu excel a obsahuje několik sloupců představující vybrané parametry, podle kterých je možné výsledek analýzy vyhodnotit. (Obrázek 21)

Každý řádek odpovídá jednomu údržbovému záznamu, který je rozdělen na části dle zvolených parametrů. Je možné zde nalézt číslo pracovního příkazu, jaký komponent a závadu označila automatická textová analýza, nebo pokud jich označila v jednom záznamu více, z jakého dokumentu pochází záznam a další.

Parametry obsažené ve výstupním souboru

- DocumentId – identifikuje příslušný dokument, ve kterém se záznam nachází, slouží pro zpětné vyhledání záznamu
- Document Line Number – označuje příslušný řádek v dokumentu, ve kterém se záznam nachází, slouží pro zpětné vyhledání záznamu
- WorkOrder Id – specifické číslo pracovního příkazu přiřazené k záznamu



- TaskCard Id – označení technologické karty, při které vzniknul daný záznam
- Component URI – hypertextový odkaz na daný komponent ve slovníku nástroje TermIt
- Component Label – komponent označený automatickou textovou analýzou
- Component Score – hodnocení, jak moc se shoduje označený text v záznamu s vybraným komponentem („pravděpodobnost, jak moc si je analýza označením jistá“)
- Multiple Components – výpis komponentů, pokud jich textová analýza označila více
- Failure URI – hypertextový odkaz na danou závadu ve slovníku nástroje TermIt
- Failure Label – závada označená automatickou textovou analýzou
- Failure Score – hodnocení, jak moc se shoduje označený text v záznamu s vybranou závadou („pravděpodobnost, jak moc si je analýza označením jistá“)
- Multiple Failures – výpis závad, pokud jich textová analýza označila více
- Is Confirmed – informace, zda byl označený pojem manuálně potvrzen uživatelem
- Original Text – text záznamu

DocumentId	WorkOrderId	TaskCardId	ComponentLabel	ComponentScore	MultipleComponents	FailureLabel	FailureScore	MultipleFailures	IsConfirmed	OriginalText
csat/full-data/1	43	17	4299815 <a href="#">http://or Crew surface tablet</a>	1.0	Hydraulic return tube; <a href="#">http://or Scratched</a>	1.0	Scratched; Missing	1.0	false	1 ea crew surfe
csat/full-data/2	43	79	4267894 <a href="#">http://or Light</a>	1.0	Light; Cabin light; Pass <a href="#">http://or Inoperative</a>	1.0		1.0	false	some lights are
csat/full-data/3	43	22	4299815 <a href="#">http://or Lamp</a>	1.0	Cockpit voice recorder <a href="#">http://or Burned</a>	1.0	Burned; Missing	1.0	false	there were four
csat/full-data/4	43	22	4299815 <a href="#">http://or Seat belt</a>	1.0	Seat belt; Bulb <a href="#">http://or Burned</a>	1.0	Burned; Missing	1.0	false	several bulbs h
csat/full-data/5	43	29	4280718 <a href="#">http://or Light</a>	1.0				1.0	false	light 62lg inop.
csat/full-data/6	43	39	4299815 <a href="#">http://or Placard</a>	1.0	Wing fairing; I profile; F <a href="#">http://or Deterioration</a>	1.0	Deterioration; Miss	1.0	false	ext: l/h and r/h
csat/full-data/7	43	02	050000-CSA- <a href="#">http://or First aid kit</a>	1.0	First aid kit; Security se <a href="#">http://or Missing</a>	1.0		1.0	false	during wy on fir
csat/full-data/8	43	33	20-120-01-01 <a href="#">http://or Clamp</a>	1.0	Clamp; Harness assen <a href="#">http://or Damage</a>	1.0	Missing; Damage	1.0	false	there was foun
csat/full-data/9	43	75	050000-CSA- <a href="#">http://or Adult life vest</a>	1.0				1.0	false	during the weel
csat/full-data/10	43	31	050000-CSA- <a href="#">http://or Primary latch</a>	1.0	Door frame; Primary la <a href="#">http://or Unserviceable</a>	1.0	Unserviceable; Mi	1.0	false	during weekly c
csat/full-data/11	43	10	050000-CSA- <a href="#">http://or Door frame</a>	0.5	<a href="#">http://or Incorrect install</a>	0.5		0.2	false	pls install servic
csat/full-data/12	43	10	050000-CSA- <a href="#">http://or Lower hinge</a>	1.0	Door frame; Hydraulic <a href="#">http://or Cracked</a>	1.0	Cracked; Missing	1.0	false	the door from t
csat/full-data/13	43	37	4301224 <a href="#">http://or Spoiler</a>	1.0				1.0	false	after technical r
csat/full-data/14	43	39	4291375 <a href="#">http://or Test stop button</a>	0.33				0.3	false	pls perform this
csat/full-data/15	43	36	4315306 <a href="#">http://or Outboard main flap</a>	0.67				0.6	false	pls perform t/e
csat/full-data/16	43	37	4301224 <a href="#">http://or Outboard fixed trailing</a>	0.5				0.5	false	after technical r
csat/full-data/17	43	14	4291385 <a href="#">http://or Filter</a>	1.0	Light indicator; Pull out <a href="#">http://or Missing</a>	0.5		0.5	false	the lp filter fron
csat/full-data/18	43	14	4291385 <a href="#">http://or Filter</a>	1.0				1.0	false	perform replac
csat/full-data/19	43	39	4291375 <a href="#">http://or Fuel pipe</a>	1.0	<a href="#">http://or Damage</a>	1.0		1.0	false	eng 1, fwd pylc
csat/full-data/20	43	48	4315306 <a href="#">http://or Landing gear control int</a>	0.4				0.4	false	perform lubrica
csat/full-data/21	43	18	4301418 <a href="#">http://or Washer</a>	1.0	<a href="#">http://or Corrosion</a>	1.0	Missing; Corrosion	1.0	false	#1# ldg; some t
csat/full-data/22	43	14	4291385 <a href="#">http://or Light indicator</a>	0.5				0.5	false	do an operator
csat/full-data/23	43	15	4297618 <a href="#">http://or Brake overtemperature</a>	0.33				0.3	false	pls perform lub
csat/full-data/24	43	31	4299639 <a href="#">http://or Exit light</a>	1.0	Galley compartment; E <a href="#">http://or Broken</a>	1.0		1.0	false	during removal
csat/full-data/25	43	15	4297618 <a href="#">http://or Rudder servocontrol</a>	0.5				0.5	false	12-22-21 ruddc
csat/full-data/26	43	15	4297618 <a href="#">http://or Aileron trim</a>	0.5				0.5	false	12-22-11 ailerc
csat/full-data/27	43	15	4297618 <a href="#">http://or Elevator</a>	1.0				1.0	false	12-22-31 eleva
csat/full-data/28	43	37	4291385 <a href="#">http://or Runway turnoff light</a>	1.0	<a href="#">http://or Missing</a>	0.5		0.5	false	left runway turn
csat/full-data/29	43	15	4297618 <a href="#">http://or Control box</a>	0.5				0.5	false	12-22-41 stab

Obrázek 21 - Výstupní soubor z databáze

## 8.3 Ověření správnosti textové analýzy testovací množiny dat

Pro ověření správnosti a úspěšnosti označování pojmů pomocí automatické textové analýzy bylo provedeno porovnání stejné reprezentativní sady dat, resp. záznamů z údržby, kde jedna sada dat obsahovala manuálně označené pojmy tvořící hledané komponenty a závady a na druhou byla aplikována textová analýza pomocí vytvořeného Slovníku komponent a závad.

K tomuto porovnání a vyhodnocení je vhodná již dříve anotovaná sada dat, která byla použita pro tvorbu tohoto definičního slovníku. Ta obsahuje manuálně potvrzené (anotované) pojmy, které jsou z hlediska této analýzy potřeba najít a tvoří tak vhodný podklad pro ověření správnosti a úrovně přesnosti vyhledávání a označování pojmů pomocí nástroje TermIt. Druhá sada dat, obsahuje totožné záznamy a byla použita pro zkoušku automatické textové analýzy bez jakéhokoliv manuálního označování pojmů.

Obě sady dat byly následně porovnány z hlediska nalezení a přiřazení správného pojmu komponentu a závady v daném záznamu o údržbě. K tomuto porovnání byly použity výstupní soubory z databáze nástroje TermIt popsané v předchozí kapitole.

### 8.3.1 Porovnání manuální anotace s výsledkem textové analýzy

Ve výstupním souboru z databáze nástroje TermIt lze nalézt sloupce s označením „ComponentLabel“ a „FailureLabel“. Tyto sloupce obsahují nástrojem vybrané a označené pojmy komponent a závad, které byly přiřazeny k danému údržbovému záznamu při provedení automatické textové analýzy. Tyto pojmy byly následně porovnány s těmi, které byly manuálně anotovány v reprezentativní sadě dat.

Na základě tohoto porovnání bylo zhodnoceno, zda byla v daném záznamu nalezena správná komponenta i závada, nebo pouze jedna z nich. Nastaly i případy, kdy nebyla nalezena správně ani komponenta ani závada, anebo šlo o záznam neobsahující ani jednu z těchto dvou kategorií a jednalo se tak o nerelevantní záznam.

Možnosti, které mohli nastat při vyhodnocování dat:

- Komponenta označena správně a závada označená správně
- Komponenta označena správně a závada označená nesprávně
- Komponenta označena nesprávně a závada označená správně
- Komponenta označena nesprávně a závada označená nesprávně
- Nerelevantní nebo nejasný záznam (bez komponenty i závady)



Porovnávaná sada dat tvořila přes 3500 údržbových záznamů. Každý z nich byl manuálně zkontrolován a porovnán se vzorovými daty a bylo rozhodnuto, zda byl daný záznam analyzován zcela správně, jen částečně nebo zcela neúspěšně.

Nadcházející odstavce obsahují tabulky s příklady analyzovaných záznamů textovou analýzou. Je zde vždy uvedena vybraná komponenta a závada, a k tomu je uveden daný údržbový záznam s hledaným spojením.

V následující tabulce jsou uvedeny příklady správně nalezených a vybraných komponent a závad, kterou provedla automatická textová analýza nástroje TermIt (Tabulka 2). Je předpokladem, že tímto způsobem bude textová analýza probíhat a označovat hledanou kombinaci komponenty a závady v daném údržbovém záznamu.

Tabulka 2 - Správně nalezené komponenty a závady

Komponenta	Závada	Analyzovaný údržbový záznam + hledané spojení
Lamp	Burned	there were found <b>burned lamps</b> in the cockpit
„žárovka“	„prasklá“	„v kokpitu byla nalezena prasklá žárovka“
Cargo door skin	Dent	there was found <b>dent</b> on fwd <b>cargo door skin</b> . location between sta455.050 and sta460.625
„dveře nákladového prostoru“	„promáčknutí“	„byly nalezeny promáčknuté dveře nákladového prostoru“
Window shade	Cracked	int: the pass cabin <b>window shades</b> were found <b>cracked</b> .
„okenní stínidlo“	„prasklé“	„v kabině byla nalezena prasklá okenní stínidla“

V následující tabulce jsou uvedeny příklady nesprávně nalezených a vybraných komponent nebo závad, kterou provedla automatická textová analýza nástroje TermIt (Tabulka 3). Tyto případy je zapotřebí blíže přezkoumat a zajistit označování správných pojmů. Může se například jednat o prioritizaci některých pojmů, úpravu stávajícího definičního slovníku, odstranění chyb algoritmu analýzy programu a další.

Tabulka 3 - Nesprávně nalezené komponenty a závady

Komponenta	Závada	Analyzovaný údržbový záznam + hledané spojení
Light	Cracked	during disassembly aft strobe light were found <b>cracks</b> on <b>conector</b> strobe light (component corroded).
„světlo“	„prasklý“	„při rozebírání zadního zábleskového světla byly nalezeny praskliny na konektoru“
Bolt	Missing	eng #1 during fuel filter removal found <b>missing</b> two <b>retaining rings</b> from d-headed bolts
„šroub“	„chybí“	„během vyjmutí palivového čerpadla byly nalezeny 2 chybějící pojistné kroužky“
Bonding jumper	Missing	the water tank <b>bonding jumper</b> was found <b>tor</b> .
„kostřící kabel“	„chybí“	„kostřící kabel vodní nádrže byl nalezen roztržený“

V následující tabulce jsou uvedeny příklady nerelevantních záznamů (Tabulka 4) neobsahující žádnou kombinaci hledané komponenty a závady. Tyto záznamy jsou pouze informativní a není v nich možné nalézt nic, co by se týkalo problematiky této práce. Automatická analýza by měla tyto případy zcela přeskočit.

Tabulka 4 - Nerelevantní záznamy

Komponenta	Závada	Analyzovaný údržbový záznam
-	-	for access to inspection flight control cables have been removed in pax cabin raceways
-	-	„z důvodu přístupu ke kontrole řídicích lan byla vyjmuta...“
-	-	galley 7 has been removed from a/c
-	-	„galley 7 byla vyjmuta z letadla“
-	-	pls.perform t /e inbd,outb flap lubrication {amm 12-22-51}due to technical cleaning
-	-	„provedte promazání vnitřní a vnější klapky... “

### 8.3.2 Zhodnocení úspěšnosti automatické textové analýzy

Z provedeného srovnání manuálně anotovaných záznamů se záznamy, na které byla použita automatická textová analýza bylo možné zjistit počet správně i nesprávně označených pojmů v testovací množině. Ze zjištěných údajů byla následně určena hladina úspěšnosti textové analýzy a úroveň její přesnosti.

Testovací množina obsahovala 3604 údržbových záznamů. Více než 2500 záznamů bylo z hlediska této textové analýzy relevantní a obsahovala hledané spojení komponenty a závady. Zbytek testovací množiny tvořily záznamy, které nebyly pro tuto analýzu relevantní a neobsahovaly hledané pojmy. Analýza by měla tyto záznamy zhodnotit jako nerelevantní.

Správně označené hledané spojení komponenty a závady v jednotlivých údržbových záznamech představovalo přibližně 60 % úspěšnost označení ze všech relevantních záznamů. Jednalo se o 1505 záznamů z celkového počtu 2531, ve kterých nástroj TermIt správně označil příslušnou komponentu a její závadu uvedenou v daném záznamu.

Zbytkovou množinu tvořila kombinace správně označené pouze komponenty nebo pouze závady. Ta dohromady tvořila přibližně 35 % označených záznamů z celkového počtu, z toho přibližně 15 % označených pouze komponent a přibližně 20 % označených pouze závad ve zbylé množině dat. Lze tedy konstatovat, že automatická textová analýza pracuje lépe s výrazy pro závady než pro komponenty. To je pravděpodobně způsobeno menším počtem pojmů ve slovníku definující závady a jejich jednoduchost, narozdíl od složitějších pojmů popisující letadlové komponenty nebo místo označení nálezu závady.

Zbývajících 5 % porovnávaných pojmů představovalo ty případy, kdy automatická textová analýza v daném záznamu neoznačila správně ani komponentu a ani závadu, nebo v záznamu neoznačila žádný pojem, přestože se v něm hledané spojení nacházelo.

V následující tabulce jsou uvedeny výsledky zjištěné na základě porovnání s testovací množinou (Tabulka 5)

Tabulka 5 - Úspěšnost automatické textové analýzy – testovací množina

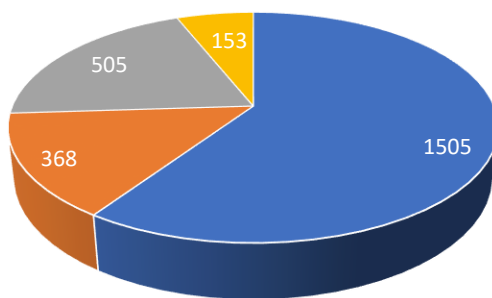
Testovací vzorek	Relevantní záznamy	Nerelevantní záznamy	Celkem
Počet záznamů	2531	1073	3604
Ze všech relevantních záznamů:			
Správně označená komponenta i závada	1505	Úspěšnost označení	59,46 %
Správně označeno komponent	1873	Úspěšnost označení	74 %
Správně označeno závad	2010	Úspěšnost označení	79,41 %
Správně označená pouze komponenta	368		
Správně označená pouze závada	505		
Neoznačená komponenta i závada	153	Neúspěšnost označení	6,05 %

Z celkového počtu 2531 relevantních údržbových záznamů bylo automatickou textovou analýzou nástroje TermIt provedeno u 1505 záznamů správné označení jak komponenty, tak i závady. U zbývajících počtu záznamů pak došlo k označení pouze správné komponenty a to ve 368 případech. V dalším případě se jednalo o označení pouze správné závady, kterou tvořilo 505 případů z celkového počtu. Ve 153 případech nebyla textová analýza schopná označit správně ani komponentu ani závadu.

Vyhodnocení automatické textové analýzy testovací množiny z hlediska rozdělení dle úspěšnosti je zobrazeno v grafu na následující straně (Graf 1). Graf jako celek představuje celkovou množinu všech relevantních dat (2531 záznamů) a jeho jednotlivé části zobrazují počet záznamů, u kterých bylo provedeno dané označení nástrojem TermIt. Druhý graf (Graf 2) zobrazuje celkovou procentuální úspěšnost označení komponent a závad.

Největší úspěšnost má nástroj TermIt při označování termínů pro závady. Je to způsobeno menším počtem pojmů ve slovníku a jejich jednoduchostí. O něco málo menší úspěšnost má nástroj ve správném označování pojmů pro komponenty. Pojmů pro komponenty je ve slovníku více a obsahují někdy i komplikovanější výrazy, které mohou nástroji v některých případech činit obtíže. Poslední množinu, a také hlavní výstup této analýzy, tvoří průnik těchto dvou množin, které představují úspěšnost správného označení současně jak komponenty, tak i závady. Nástroj tuto funkci v testovací množině provedl na hladině úspěšnosti 60 %.

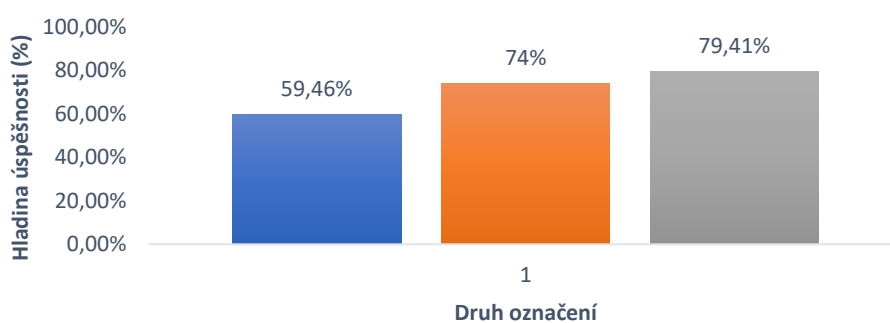
### Rozdělení záznamů dle výsledku textové analýzy



- Správně označená komponenta i závada
- Správně označená pouze komponenta
- Správně označená pouze závada
- Neoznačená komponenta i závada

Graf 1 - Rozdělení záznamů dle výsledku textové analýzy – testovací množina dat

### Hladina úspěšnosti označování pojmů



- Úspěšnost označení komponenty i závady
- Úspěšnost označení komponent
- Úspěšnost označení závad

Graf 2 - Hladina úspěšnosti označování pojmů – testovací množina dat

## 8.4 Ověření textové analýzy na nezávislé sadě dat

Předchozí textová analýza a její vyhodnocení bylo provedeno na testovací množině dat, pomocí které byl vytvořen Slovník komponent a závad. Získané výsledky však samy o sobě nemohou být reprezentovány jako konečná hladina úspěšnosti označení pojmů, jelikož data nebyla na slovníku nezávislá. Z tohoto důvodu byla provedena další automatická textová analýza na nezávislé sadě dat pomocí stejného slovníku, aby bylo možné ověřit a vyhodnotit reálnou úspěšnost nástroje Termlt při označování definovaných pojmů.

### 8.4.1 Zhodnocení úspěšnosti automatické textové analýzy

Stejným způsobem, jako u předchozí textové analýzy, bylo provedeno manuální vyhodnocení výstupního souboru, který obsahoval nástrojem označené pojmy. Ty byly následně porovnány s textem obsaženým v údržbovém záznamu a bylo posouzeno, zda došlo ke správnému označení pojmů. Tímto způsobem bylo dosaženo počtu správně i nesprávně označených pojmů v nezávislé sadě dat a následně určena úspěšnost textové analýzy.

Nezávislá sada dat obsahovala 3438 údržbových záznamů. Více než 2200 záznamů bylo z hlediska textové analýzy relevantní a obsahovaly hledané spojení komponenty a závady. Zbývající záznamy byly z hlediska textové analýzy nerelevantní a nebyly použity.

Po provedení textové analýzy této datové sady byla úspěšnost správného označení obou pojmů v daných záznamech přibližně 45 %. U zbývajících záznamů byly správně označeny pouze komponenty nebo pouze závady. Tato množina představovala přibližně 46 % označených záznamů z celkového počtu, z toho přibližně 14 % označených pouze komponent a přibližně 32 % označených pouze závad ve zbylé množině dat.

Ve zbývajících případech, které představovaly přibližně 8 % záznamů, automatická textová analýza neoznačila správně ani komponentu a ani závadu, nebo v záznamu neoznačila vůbec nic, přestože se v něm hledané spojení nacházelo.

Vyhodnocení textové analýzy nezávislé sady dat ukázalo, obdobně jako u předchozí analýzy, že textová analýza nástroje Termlt pracuje úspěšněji s výrazy pro závady než pro komponenty. Opět je možné přisoudit tento vliv menšímu počtu pojmů ve slovníku definující závady a jejich jednoduchosti narozdíl od složitějších pojmů popisující letadlové komponenty nebo místo označení nálezu závady.

V následující tabulce jsou uvedeny výsledky zjištěné na základě vyhodnocení textové analýzy nezávislé sady dat (Tabulka 6)

Tabulka 6 - Úspěšnost automatické textové analýzy – nezávislá sada dat

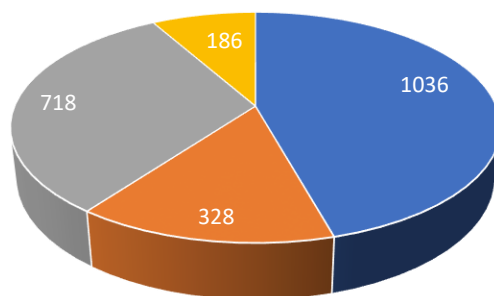
Nezávislá sada dat	Relevantní záznamy	Nerelevantní záznamy	Celkem
Počet záznamů	2268	1170	3438
Ze všech relevantních záznamů:			
Správně označená komponenta i závada	1036	Úspěšnost označení	45,68 %
Správně označeno komponent	1364	Úspěšnost označení	60,14 %
Správně označeno závad	1754	Úspěšnost označení	77,34 %
Správně označená pouze komponenta	328		
Správně označená pouze závada	718		
Neoznačená komponenta i závada	186	Neúspěšnost označení	8,2 %

Z celkového počtu 2268 relevantních údržbových záznamů bylo automatickou textovou analýzou nástroje Termlt provedeno u 1036 záznamů správné označení jak komponenty, tak i závady. U zbývajících počtu záznamů pak došlo k označení pouze správné komponenty, a to ve 328 případech. V dalším případě se jednalo o označení pouze správné závady, kterou tvořilo 718 případů z celkového počtu. Ve 186 případech nebyla textová analýza schopná označit správně ani komponentu ani závadu.

Vyhodnocení automatické textové analýzy nezávislé sady dat z hlediska rozdělení dle úspěšnosti je zobrazeno v grafu na následující straně (Graf 3). Graf jako celek představuje celkovou množinu všech relevantních dat (2268 záznamů) a jeho jednotlivé části zobrazují počet záznamů, u kterých bylo provedeno dané označení nástrojem Termlt. Druhý graf (Graf 2) zobrazuje celkovou procentuální úspěšnost označení komponent a závad.

Detailní porovnání výsledků obou textových analýz je předmětem až následující kapitoly, pokud ale provedeme rychlé srovnání je možné si všimnout, že došlo k výraznému snížení celkové úspěšnosti správného označení komponenty i závady, oproti textové analýze testovací množiny. Podrobnější srovnání je obsaženo v kapitole 8.4.2 Porovnání výsledků obou textových analýz.

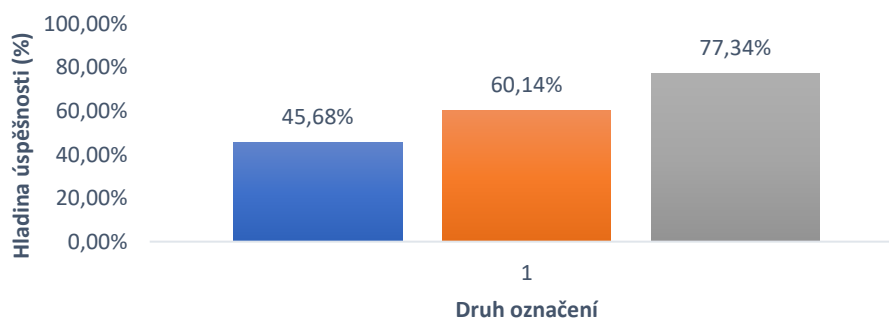
### Rozdělení záznamů dle výsledku textové analýzy



- Správně označená komponenta i závada
- Správně označená pouze komponenta
- Správně označená pouze závada
- Neoznačená komponenta i závada

Graf 3 - Rozdělení záznamů dle výsledku textové analýzy – nezávislá sada dat

### Hladina úspěšnosti označování pojmů



- Úspěšnost označení komponenty i závady
- Úspěšnost označení komponent
- Úspěšnost označení závad

Graf 4 - Hladina úspěšnosti označování pojmů – nezávislá sada dat

## 8.4.2 Porovnání výsledků obou textových analýz

Při porovnání výsledků textové analýzy testovací množiny dat s výsledky textové analýzy nezávislé sady dat je možné zpozorovat přibližně 15 % propad v celkové úspěšnosti správného označení komponenty i závady, v případě nezávislé sady. Zároveň došlo k nárůstu počtu správně označených pouze závad. Nárůst počtu označených pouze závad byl způsoben nedostatečným počtem správně označených komponent. Ty by jinak dohromady se závadami tvořily celkovou úspěšnost označení.

V případě obou automatických textových analýz bylo zjištěno, že textová analýza nástroje Termlt je schopna pracovat úspěšněji při označování pojmů pro závady. Z výsledků je možné si všimnout, že úspěšnost správného označení závad byla v obou datových sadách téměř stejná. Jednalo se přibližně o 78 % úspěšnost označení.

Naopak úspěšnost správného označení komponent v nezávislé sadě dat poklesla přibližně o 15 %. Je pravděpodobné, že tento pokles byl způsoben tím, že se v záznamech vyskytly nové pojmy pro komponenty, které se neobjevily v testovací množině dat a nemohly tak být zařazeny do Slovníku komponent a závad. Druhou variantou, která mohla nastat, je větší komplikovanost analyzovaných záznamů, se kterou textová analýza nástroje Termlt nebyla schopná řádně pracovat a označit hledané pojmy pro komponenty.

Posledním, ale také nejvíce přispívajícím faktorem, který má vliv na úspěšnost textové analýzy, je nástrojem nesprávně zvolený pojem z více nalezených výrazů v jednom údržbovém záznamu. O této limitaci a dalších nedostatcích je pojednáno v následující kapitole.

V následující tabulce jsou pro srovnání uvedeny výsledky obou textových analýz zjištěné na základě jejich vyhodnocení (Tabulka 7).

Tabulka 7 - Porovnání výsledků obou textových analýz

	Testovací množina dat	Nezávislá sada dat
Počet záznamů	2531	2268
Ze všech relevantních záznamů:		
Správně označená komponenta i závada	1505 (59,46 %)	1036 (45,68 %)
Správně označeno komponent	1873 (74 %)	1364 (60,14 %)
Správně označeno závad	2010 (79,41 %)	1754 (77,34 %)
Správně označená pouze komponenta	368 (14,54 %)	328 (14,46 %)
Správně označená pouze závada	505 (19,95 %)	718 (31,66 %)
Neoznačená komponenta i závada	153 (6,05 %)	186 (7,41 %)



## 8.5 Limitace současného stavu automatické textové analýzy

Během práce s nástrojem TermIt a jeho funkcí provádět automatickou textovou analýzu bylo možné upozorovat několik nedostatků, které se týkaly označení hledaných pojmů. Především se jednalo o vyhledání a označení nesprávných pojmů za pomoci definovaného slovníku. V jistých případech, které představovaly složitější textové údržbové záznamy, vyhledávací algoritmus dával nevhodně přednost jiným pojmům a došlo tak k označení nevyhovujících výrazů. Dále se pak jednalo o označování vícero pojmů pro komponenty a závady v jednom údržbovém záznamu, což představuje problém z hlediska identifikace vlastního hledaného spojení komponenty a závady. V neposlední řadě se jedná o optimalizaci nástroje a práce s ním, z hlediska analýzy velkého objemu dat, doladění importu dokumentů a dalších funkcí.

Pro větší přesnost textové analýzy a snadnější práci s nástrojem je třeba vyřešit následující nedostatky:

- Označení nesprávné komponenty nebo závady
- Označení vícero komponent a závad najednou
- Uživatelská přívětivost nástroje

### 8.5.1 Označení nesprávné komponenty nebo závady

Při manuálním porovnávání analyzovaných dat s testovací množinou byly zaznamenány případy, ve kterých docházelo k nesprávnému označování komponent nebo závad. K těmto případům, kdy byly textovou analýzou označeny nesprávné pojmy v některých údržbových záznamech, docházelo na základě chybné automatické analýzy. Její algoritmus s největší pravděpodobností nesprávně upřednostňoval jiné pojmy, přestože byly řádně zaneseny dle dříve definovaných podmínek ve Slovníku komponent a závad.

Jednalo se o upřednostnění jiného pojmu, resp. výrazu uvedeného v jeho vyhledávacím textu („Search string“) před jiným vhodnějším pojmem a jeho hlavním názvem („Preferred label“), viz kapitola 6.4.1 Podmínky pro vytváření pojmů.

Tento problém se týká především pojmu pro závadu „Missing“ („chybějící“), resp. výrazu uvedeném v jeho vyhledávacím textu, kterým byl výraz „not found“ („nebyl nalezen“). Výraz „not found“, resp. pouze jeho část „found“ se v údržbových záznamech objevovala velmi často a byla jí algoritmem v některých případech nesprávně udělována přednost před jinými pojmy ve slovníku s výše nadřazenými výrazy (Tabulka 8).

Tabulka 8 - Nesprávné označení pojmu "Missing"

Komponenta	Závada	Analyzovaný údržbový záznam + hledané spojení
Runway turnoff light	Missing	left runway turnoff light was found faulty.
„odbočné světlo“	„chybí“	„levé odbočné světlo bylo nalezeno nefunkční“
Drain hose	Missing	the drain hose from galley 4 was found torn.
„vypouštěcí hadice“	„chybí“	„vypouštěcí hadice z galley 4 byla nalezena roztržená“

Jedním z možných řešení je výrazy ponechat ve slovníku nezměněné s tím, že tento problém bude odstraněn v dalších vývojových verzích nástroje TermIt. O této skutečnosti byli vývojáři informováni a byla s nimi vedena komunikace o těchto nedostatcích. K tomu sloužil příslušný reportovací systém. Další možností je odladění nástroje pro konkrétní problémové výrazy.

### 8.5.2 Označení více komponent a závad v jednom záznamu

Vzhledem k tomu, že se v některých údržbových záznamech objevuje více komponent nebo závad najednou, tak tato skutečnost může způsobovat nepřesnost a sníženou spolehlivost v automatické textové analýze.

Předpokladem je, že by se v jednom údržbovém záznamu měla objevit pouze jedna komponenta a jedna závada. Ale to není vždy pravda. V případech, kdy se podrobněji popisuje umístění závady, tak se může stát, že se v záznamu objeví více komponent. Pomocí nich technik popisuje přesné umístění závady a jedna z nich je pak i konečnou komponentou, která byla nalezena jako poškozená. Nástroj TermIt však při provádění analýzy označí všechny ty výrazy, které se shodují s použitým definičním slovníkem, a proto se může stát, že v jednom záznamu bude označeno více pojmů. To představuje problém z hlediska výběru správné komponenty nebo závady ze všech označených a následné zpracování všech záznamů.

Tabulka 9 - Označení více komponent a závad v jednom záznamu

Komponenta	Další označ. komponenty	Závada	Další označ. závady	Analyzovaný údržbový záznam + hledané spojení
Fuselage	Fuselage; Drain valve; Lower fuselage	Cracked	Cracked; Missing	ext. fuselage - the drain valves (2x) on lower fuselage skin were found cracked .
„trup/plášť“	„trup/plášť“ „výpustný ventil“ „spodní plášť“	„prasklé“	„prasklé“ „chybí“	„vnější trup – výpustné ventily na zadním trupu byly nalezeny prasklé“
Fuselage	Fuselage; Drain valve	Cracked	Cracked; Missing	the fuselage water drain valves were found cracked .
„trup/plášť“	„trup/plášť“ „výpustný ventil“	„prasklé“	„prasklé“ „chybí“	„výpustné ventily na trupu byly nalezeny prasklé“

Algoritmus, kterým byl vygenerován výstupní soubor z databáze nástroje Termlt byl pravděpodobně nastaven tak, že pokud bylo označeno automatickou textovou analýzou více komponent nebo závad v jednom záznamu, tak byl vybrán ten výraz, který byl označen v pořadí jako první (Tabulka 9). V některých případech se jednalo o správný pojem pro komponentu nebo závadu, protože se v záznamu nacházely hned na začátku, ale v jiných případech se již hledané pojmy nacházely na dalších pozicích.

Omezení tohoto charakteru není snadné odstranit. Vyžadovalo by to sestavení seznamu vzorových záznamů, ve kterých by bylo definováno, v jakém pořadí by měl algoritmus vybírat správné pojmy tak, aby mohl být generován korektní výstupní soubor pro vyhodnocení provedené textové analýzy. Toto řešení není z hlediska velkého počtu údržbových záznamů a jejich různorodosti zcela praktické a vyžadovalo by to náročný proces manuálního zpracování dat, kterým by byl vytvořen požadovaný vzor.

Jednodušším řešením by mohlo být zavedení nových pravidel pro tvorbu údržbových záznamů. Mechanici by byli školeni tak, že při tvorbě údržbových záznamů by příslušnou komponentu, u které byla lokalizována závada, uvedli v záznamu jako první. Tím by byla zajištěna jednoznačnost určení komponenty a automatické textové analýze by to výrazně usnadnilo její vyhledání a označení.

### **8.5.3 Uživatelská přívětivost nástroje**

V době psaní této práce nástroj Termlt procházel řadou změn z hlediska jeho pracovních verzí a funkcí, které umožňoval. S každou novou verzí byly opraveny známé chyby, které se projevíly při používání nástroje, nebo přidány nové funkce, které dosud nebyly dostupné.

Za jednu z nevýhod nástroje lze považovat vyšší hardwarovou náročnost a potřebu kvalitního internetového připojení. Tato kombinace hraje roli při procházení a vytváření slovníků, načítání a úpravu jednotlivých pojmů nebo při zobrazování importovaných dokumentů a práci s nimi. Pokud není počítač dostatečně výkonný a internetové připojení dostatečně rychlé, může to při větším množství pojmů a rozsáhlejších dokumentech způsobovat zdlouhavé načítání, které výrazně zpomaluje práci s nástrojem a následnou automatickou textovou analýzu.

Jednou z omezujících funkcí nástroje Termlt je použití výhradně souborů datového typu HTML. Při potřebě automatické textové analýzy jiných datových typů, muselo tomuto kroku předcházet složité transformování vybraných dokumentů. Vývojáři nástroje jsou si vědomi tohoto nedostatku. Funkce, která bude umožňovat import souborů i v jiných datových typech, než je formát HTML, bude do nástroje implementována v některých budoucích verzích nástroje. Následně bude import souborů do nástroje jednodušší a uživatelsky přívětivější.

Další věcí, kterou je třeba zmínit je, že nástroj TermIt není komerčním softwarem. Jedná se o produkt univerzitního výzkumu fakulty elektrotechnické ČVUT v Praze a především o prototypovou aplikaci, která je stále ve vývoji a pracuje se na jejím odladování. Proto při práci s nástrojem někdy docházelo k tomu, že se objevily nepředpokládané chyby nebo zaseknutí nástroje, které znemožňovaly vytváření nových pojmů, jejich editaci a uložení, což výrazně komplikovalo anotaci dokumentů a postup práce, protože bylo třeba se k některým věcem několikrát vrátit, než byly zadány korektně. Je však předpokladem, že tyto problémy budou odstraněny v některých následujících verzích nástroje. Pomocí reportovacího systému je možné vývojářům nahlásit objevené chyby nebo jiné problémy, na které uživatel narazil při používání nástroje. Do budoucna bude jistě probíhat jeho optimalizace a přidávání nebo oprava nových funkcí, které usnadní práci s nástrojem.

## **8.6 Možnosti vyhodnocení výstupu automatické textové analýzy**

Pro vyhodnocení automatické textové analýzy, resp. výstupního souboru z databáze nástroje TermIt je možné využít celou řadu multifunkčních nástrojů, které umí pracovat s tabulkovými daty. Jedná se hlavně o funkci výběru a třídění dat ze zdrojového souboru, jejich rozdělení a filtraci za pomoci specifického klíče, dotazování na specifické parametry a další.

Těmito funkcemi disponuje většina tabulkových editorů, např. Microsoft Excel a jeho alternativy, nebo speciální programy umožňující tvorbu relačních databází, např. Microsoft Access a jiné databázové nástroje.

Pro názornost a představu možnosti využití jednotlivých nástrojů k vyhodnocení automatické textové analýzy je stručně přiblížen příklad práce s daty pro následující metody:

- Kontingenční tabulka (graf) – tabulkový editor (MS Excel, LibreOffice Calc,..)
- Databázové nástroje – relační databáze (MS Access, MySQL,..)

### **8.6.1 Tabulkový editor – kontingenční graf**

Jednou z možností, jakým lze vyhodnotit výstup z automatické textové analýzy je použití funkce kontingenčního grafu. Ten umožňuje zpracování velkého množství dat s různým počtem parametrů, podle kterého lze data uspořádat a reprezentovat. Data je možné uspořádat do řádků i sloupců, dle preference uživatele tak, aby z nich bylo možné jednoduše zjistit hledané informace nebo sestavit požadovanou statistiku.

V případě této práce, ve které byla prováděna analýza údržbových záznamů je zajímavá kombinace Komponenty („Component“) a Závady („Failure“) vždy pro jednu specifickou

technologickou kartu, tedy určitou práci vykonanou na letadle dle příslušné technologické karty, při které byla nalezena závada a vzniknul z ní příslušný údržbový záznam.

Pro uspořádání záznamů obsažených ve výstupním souboru nástroje TermIt („výstup z textové analýzy“) lze použít specifické číslo technologické karty („TaskCard Id“), které slouží jako unikátní identifikátor. Ten se objeví v tabulce jako nadřazená položka, která v sobě seskupuje všechny údržbové záznamy, ve kterých se daná technologická karta vyskytla. Tímto krokem je odstraněna duplicita a dochází ke zvýšení přehlednosti v datech.

Následně je možné do tabulky přiřadit příslušná data obsahující pojmy pro komponenty a závady označené automatickou textovou analýzou v jednotlivých údržbových záznamech. Dalším krokem pak může být přiřazení počtu výskytu daných pojmů, který může následně sloužit pro vyhodnocení vybraných dat z hlediska statistického zpracování.

Kombinace parametrů, které byly vloženy do kontingenčního grafu, vytvoří hierarchickou strukturu, kterou je možné vidět na obrázku níže (Obrázek 22). Pokud dojde k rozkliknutí příslušné technologické karty, tak se zobrazí seznam jednotlivých komponent, které jsou k vybrané kartě přidruženy. Ten představuje komponenty nalezené automatickou textovou analýzou v kombinaci s danou technologickou kartou. Jednotlivé komponenty lze dále rozkliknout, stejným způsobem jako technologické karty, a dojde k zobrazení seznamu závad, které byly textovou analýzou přidruženy k danému komponentu.

Za pomoci této hierarchické struktury, která pro každou technologickou kartu přesně udává nalezenou komponentu a závadu, která se u něj vyskytla, je možné provádět rozsáhlé statistické analýzy. Ty se mohou týkat například analýzy výskytu specifických kombinací komponent a závad, pouze dané závady a jejího výskytu na jednotlivých komponentech nebo lze provést sestavení statistiky nejčastěji se vyskytujícími se závadami, které byly zaznamenány při provádění prohlídky na letadle dle dané technologické karty.

	A	B	C	D	E	F
3	<b>Technologické karty (Komponenty a závady)</b>	<b>Nalezený počet</b>				
4	000001-A000220	5				
5	000001-A000230	15				
6	000002-A000230	2				
7	050000-CSA-00031-000	68				
8	05-020-00-90					
9	05-023-00-08	6				
10	052000-001	290				
11	05-HAP-010	1				
12	05-HAP-011	2				
13	05-HAP-051	1				
14	05-HAP-080	12				
15	Access panel	1				
16	Cracked	1				
17	Blocker door wear pad	1				
18	Bolt	1				
19	Chafed	1				
20	Door hinge	1				
21	Corrosion	1				
22	Fairing panel					
23	Folding step					
24	Fuselage	1				
25	Dent	1				
26	Gusset plate	1				
27	Gouge	1				
28	Jumper	1				
29	Lens assembly	1				

Obrázek 22 - Kontingenční tabulka technologických karet (Komponenty a závady)

## 8.6.2 Databázové nástroje – relační databáze a dotazy

Databáze obecně slouží k uchování a organizaci velkého množství dat. Existuje mnoho druhů databází, ale z hlediska této práce nejvíce vyhovuje nejpoužívanější typ, kterým je relační databáze. Tyto databáze uchovávají data v tabulkách, se kterými je možné jednoduše pracovat a manipulovat s jednotlivými záznamy. Každá tabulka by měla mít svůj tzv. primární klíč, který slouží k jednoznačné identifikaci záznamu v tabulce. K databázovým dotazům slouží jazyk SQL („Structured Query Language“ – strukturovaný dotazovací jazyk). Jedná se o jazyk uzpůsobený pro práci s daty, pomocí kterého uživatel vytvoří požadovaný dotaz a databáze vyhledá příslušná data.

Hlavním rozdílem mezi tabulkovými editory a databázovými nástroji je ve způsobu uložení dat a manipulaci s nimi. Tabulkové editory jako je MS Excel slouží většinou méně náročným uživatelům, kteří nevyžadují komplikovanější práci s daty, resp. postačí jim pouze jednoduché dotazy. Databáze naopak umožňují uchovávat větší množství dat, která jsou přístupná více uživatelům současně, a jsou schopny zajistit rychlý přístup k datům, práci s nimi a zadávání komplexních dotazů.

Na obrázku níže je zobrazen dotaz v nástroji MS Access, který z výstupního souboru nástroje Termit vypíše jednotlivě všechna čísla technologických karet, spolu s označenými pojmy pro Komponenty a závady a počtem, kolikrát byla tato shoda v údržbových záznamech nalezena (Obrázek 23).

TaskCardId	Pole	ComponentLabel	Pole	FailureLabel	Pole	Počet výskytů
53-882-00-01		Fuselage		Dent		74
202800-01-1		Spoiler		Torn		73
ZL-471-01-1		Fairing		Cracked		67
276400-02-1		Spoiler		Loosened		63
110000-006		Placard		Missing		57
53-380-00-01		Intercostal		Cracked		56
INSPEC-11TAV-00019		Placard		Missing		55
110000-004		Placard		Missing		54
252100-006		Plastic part		Damage		53
INSPEC-51TAV-00047		Fuselage		Dent		53
252700-001		Literature pocket		Unserviceable		49
ZL-334-01-1		Bonding jumper		Unserviceable		46
INSPEC-25TAV-01182-02		Placard		Missing		44
27-235-00-02		Control cable		Low tension		43
ZL-500-02-1		Seal		Torn		42
INSPEC-25TAV-01182-02		Kickstrip		Cracked		40

Obrázek 23 – Dotaz v databázovém nástroji MS Access

## 9. Diskuse

Cílem této práce bylo seznámit se s možnostmi využití automatické textové analýzy k rozboru a vyhodnocení údržbových záznamů, v případě této práce záznamů z prostředí letecké údržby, a zhodnotit možný potenciál a přínosy její implementace do systému údržbových organizací. Nástroje umožňující provádění textové analýzy slouží k vyhodnocování rozsáhlých textových souborů, které mohou být v podobě strukturovaného, tak i nestrukturovaného textu, a poskytují uživatelům snadnější a rychlejší analýzu svých dat oproti manuálnímu zpracování.

Práce je zaměřena na automatickou textovou analýzu údržbových záznamů, které obsahují nálezy neplánovaných závad objevených při údržbě letadel, během provádění výrobcem předepsaných a provozovatelem objednaných prohlídek dle údržbového plánu. Snahou je automatické vyhodnocení dostupných údržbových záznamů, z hlediska identifikace jednotlivých letadlových komponent a jejich závad, které byly objeveny na letadle a následně najít možnosti využití informací získaných z historických dat, pro budoucí plánování údržby. Práce byla vypracována na základě úzké spolupráce s údržbovou organizací CSAT, která mi poskytla všechny potřebné materiály k provedení příslušných datových analýz.

Nejdříve bylo potřeba se seznámit s postupem údržby letadel v údržbové organizaci a způsoby, jakými vznikají záznamy z údržby. Následně bylo zapotřebí zjistit jaké informace jsou obsahem jednotlivých záznamů, jak vypadá formát záznamu, a jakým způsobem jsou jednotlivé údaje uspořádány. Tento krok byl důležitý z hlediska toho, abych byl schopen údržbovým záznamům rozumět, pracovat s nimi a dokázal v nich identifikovat příslušné informace vhodné k textové analýze.

Nejdůležitější položkou v záznamu je pole popisu závady. Popis je ručně vyplňován mechaniky při nálezů závady na letadle a je ve formě volně psaného textu. Obsahuje letadlovou komponentu, na které byla nalezena závada nebo popis umístění závady na konkrétním letadlovém celku. Automatická textová analýza, kterou se zabývám v této práci, je zaměřena na vyhodnocení těchto nestrukturovaných záznamů z hlediska nalezení specifické komponenty a její závady. Vzhledem k velkému množství dat z údržby se strojové zpracování jeví jako vhodný nástroj a prostředek k jednoduchému a rychlému rozboru vlastních dat organizace.

Dalším krokem, který jsem provedl, bylo zjištění požadavků potřebných k provádění textových analýz nestrukturovaných dat. Narazil jsem na několik příkladů využití v údržbě, nikoliv ale v té letecké. Všechny případy měly ale jedno společné, a tím byla potřeba mít k dispozici definiční slovník výrazů, podle kterého by textová analýza byla schopna vykonávat požadovanou činnost. Vzhledem k tomu, že jsem nebyl schopen dohledat podobné využití



tohoto konceptu v letecké údržbě, tedy ani žádný existující definiční slovník komponent a závad, přistoupil jsem k jeho vlastnímu vytvoření. Slovník jsem nazval „Slovník komponent a závad“ a je rozdělen do dvou nadřazených pojmů, kterými jsou „Komponenty“ a „Závady“. V nich jsou obsaženy pojmy, které jsem manuálně extrahoval z vybrané vzorové sady údržbových dat, a v této práci mi sloužily jako podklad pro provedení textové analýzy. Nevýhodou je, že takto vytvořené slovníky nemusí být zcela kompletní a některé pojmy mohou chybět. Vycházel jsem ale z předpokladu, že většina pojmů se již vyskytla v záznamech ve vzorové sadě a slovník tak bude dostatečně obsáhlý pro analýzu dalších dostupných dat.

Nástroj, který jsem vybral pro provedení automatické textové analýzy, je nástroj s názvem Termlt. Na základě porovnání s dalšími nástroji se mi jevil tento nástroj jako vhodný. Pozitivně hodnotím jeho jednoduchost a způsob, jakým lze snadno vytvářet a spravovat slovníky pojmů a provádět textové analýzy, na rozdíl od ostatních nástrojů, které byly uživatelsky náročnější, a ne zcela přehledné. Jedná se o nástroj, který vzniknul univerzitním výzkumem fakulty elektrotechnické ČVUT v Praze, a měl jsem příležitost komunikovat přímo s jeho vývojáři. To mělo výhodu v tom, že jsem byl schopný konzultovat s nimi svůj záměr s nástrojem a oni mi byli ochotni poradit a také nástroj pro moji práci v rámci možností přizpůsobit. Jedná se však o experimentální aplikaci, která v sobě skrývá některé nedostatky a limitace, na kterých se pracuje, jelikož se jedná o nekomerční nástroj, který je stále ve vývoji.

V této práci jsem se rozhodnul provést dvě automatické textové analýzy za pomoci nástroje Termlt a mnou vytvořeného Slovníku komponent a závad. Jednu textovou analýzu jsem provedl na testovací množině dat, ze které byl zároveň vytvořen tento slovník, abych ověřil úspěšnost označování pojmů nástrojem v anotovaných záznamech. Druhou analýzu jsem provedl na zcela nezávislých datech, aby výsledky analýzy byly objektivní a simulovaly tak použití v reálném prostředí při použití daného slovníku. Výsledky obou analýz jsem následně mezi sebou porovnal a vyhodnotil z hlediska míry úspěšnosti označení pojmů v záznamech podle toho, jestli nástroj správně označil komponentu i závadu nebo pouze jednu z nich.

V testovací množině byla automatická textová analýza jednoznačně úspěšnější, což ale také bylo předpokladem, jelikož z ní vzniknul použitý slovník. Celková úspěšnost označení byla kolem 60 %, na rozdíl od 45 % úspěšnosti označení v textové analýze nezávislých dat. Zbývající pojmy byly správně označeny pouze jako komponenta nebo pouze závada, ale také docházelo k tomu, že byly označeny nesprávné pojmy nebo nebylo označeno nic. Takto nízkou celkovou úspěšnost označení příkládám současným limitacím nástroje a jeho textové analýzy, které jsem popsal v kapitole 8.5 Limitace současného stavu automatické textové analýzy. Přesto výsledky analýzy z hlediska správně označené komponenty dosahovaly úspěšnosti blížící se k 75 % v testovací množině a 60 % v nezávislé sadě dat, a z hlediska správně označené závady v obou testovaných datových sadách k 80 %.

V případech, ve kterých automatická textová analýza neoznačila požadované pojmy v záznamech musíme zvážit několik aspektů. Tím jsou především limitace, které jsem zjistil během práce s nástrojem, a které zamezovaly větší úspěšnosti označování správných pojmů. Nízká úspěšnost označování byla nejčastěji způsobována vyhledávacím algoritmem. Ten v určitých případech dával přednost jiným, méně prioritním výrazům uvedených ve slovníku a docházelo tak k označení nerelevantních pojmů a zanesení chyby do textové analýzy. Tento problém se týkal převážně označování závad. Další limitací vyskytující se převážně u komponentů bylo, že pokud došlo v jednom údržbovém záznamu k označení více pojmů současně, tak nebylo možné ovlivnit, který z označených výrazů má prioritu a bude zanesen do databáze jako nástrojem vybraný pojem. Zpravidla tak docházelo k tomu, že se jednalo o pojem označený analýzou v pořadí jako první. Ne vždy se však hledaný komponent nacházel na prvním místě, a proto byl analýzou vybrán nesprávný pojem. Jednoduché řešení by spočívalo v nastavení pravidla uvádění porouchané komponenty jako první, již při vzniku údržbového záznamu, což by výrazně přispělo na úspěšnosti označování. Složitější řešení by představovalo sestavení souboru vzorových záznamů s určenou prioritou označení a jeho implementace do algoritmu. Toto řešení by však nemuselo být konečné, z hlediska všech možných variací, které se mohou objevit v údržbových záznamech.

Je však nutné poukázat na to, že i když dojde k částečnému nebo úplnému odstranění zjištěných limitací nástroje, anebo k úpravě vyhledávacího algoritmu, tak nástroj TermIt nebude možné zcela odladit tak, aby pracoval se 100 % úspěšností. Je to dáno především rozsáhlou variací pojmů v údržbových záznamech, které vyplňují technici údržby vlastními slovy. Každý proto může popsat svoji provedenou práci jiným způsobem, použít jiné, ale podobné výrazy nebo se dopustit gramatických chyb, překlepů, použití nesprávných výrazů apod. Všechny tyto aspekty mohou činit obtíže z hlediska strojového zpracování záznamů, ale je však možné některé z těchto zábran odstranit přidáním dalších variací pojmů do slovníku. Ne vždy je však možné zahrnout všechny nedokonalosti. Tím jsem se dostal k tomu, že převážně záležína obsahu použitého slovníku, a zda je udržován v aktuálním stavu vzhledem k provádění požadované automatické textové analýzy.

Závěr mé práce je zaměřen na možnosti vyhodnocení automatické textové analýzy. Navrhnul jsem dva možné způsoby vyhodnocení. Jedním je vyhodnocení pomocí kontingenční tabulky, kde jsou seskupeny záznamy se stejným referenčním číslem a seznam výskytu pojmů komponent a závad v dané skupině. Druhou možností je využití relační databáze a dotazů, tak aby bylo dosaženo požadovaných výstupů, stejných jako u kontingenční tabulky. V obou případech se však jedná o návrh a nezahrnuje kompletní řešení.

Výhodou myšlenky této práce je celkový vznik návrhu provádění automatické textové analýzy nestrukturovaných údržbových záznamů. Údržbová organizace CSAT žádnou podobnou

analýzu svých dat momentálně neprovádí, což mi umožnilo mít volnou ruku při tvorbě tohoto návrhu. Absencí analýzy vlastních dat může údržbová organizace přicházet o důležité informace, které by mohla využít z hlediska budoucího plánování údržby letadel. Jedná se především o zjištění vývoje a trendu nálezu neplánovaných závad z historických dat. Jde o to vyhledat v záznamech závady, které se při konkrétních údržbových pracích na letadle vyskytují často a opakují se. S těmito informacemi lze při plánování stejné údržbové práce počítat, že může dojít k nálezu této neplánované závady a je možné vzhledem k této skutečnosti upravit dle potřeby celkový plán údržby. Tím je myšleno přizpůsobit dostupné kapacitní i personální možnosti, naplánovat dostupnost potřebného materiálu a nářadí, spolu s časovým plánem a harmonogramem prací a docílit tak efektivnějšího plánování údržby v těchto organizacích.

Jako nevýhodu považuji rozsáhlou variaci výrazů v nestrukturovaných údržbových záznamech, které mohou způsobovat komplikace z hlediska navrhovaného automatického zpracování. Důležité je mít kvalitně zpracovaný definiční slovník pojmů, který bude zahrnovat co nejvíce variací vybraných pojmů, nikdy však nebude možné zahrnout všechny tyto variace s ohledem na gramatické chyby, překlepy nebo jiné chybně použité výrazy apod. Nestrukturovaný text není nijak limitován a každý mechanik si ho vyplňuje po svém. Porouchané komponenty jsou často uváděny v jiném pořadí a tím dochází k tomu, že algoritmus vybere nesprávné komponenty, v případě označení více komponent v jednom záznamu. Tento problém by mohl být vyřešen tím, že by porouchaný komponent byl uváděn v záznamu jako první. Po odstranění této limitace a dalších, které jsem zmínil výše předpokládám, že dojde k celkovému nárůstu úspěšnosti označování správných pojmů automatickou textovou analýzou.

## 10. Závěr

Tato práce je zaměřena na návrh způsobu vyhodnocení údržbových záznamů z prostředí letecké údržby za pomoci automatické textové analýzy. Strojové zpracování je výhodnější z hlediska rychlosti a počtu zpracovaných záznamů na rozdíl od manuálního zpracování a umožňuje kvalitní a přesný rozbor vlastních dat organizace. Reálná data byla poskytnuta údržbovou organizací CSAT, která podobný rozbor svých údržbových záznamů v současné době neprovádí, což přispělo ke vzniku této práce. Použitelnost návrhu se však striktně nevztahuje jen k této organizaci a může být využit i jinými údržbovými organizacemi.

Práce obsahuje vlastní návrh provádění automatické textové analýzy údržbových záznamů v letecké údržbě. Je v ní uveden princip provádění textové analýzy, postup práce s nástrojem a anotace pojmů, ale také doporučený způsob vyhodnocení dat. Další důležitou a přínosnou částí této práce je Slovník komponent a závad. Ten obsahuje vybrané pojmy leteckých komponentů a jejich možná selhání. Byl vytvořen pro potřeby provedení vlastních textových analýz na reálných datech s následným vyhodnocením a ověřením úspěšnosti označování. Slovník je volně přístupný a může sloužit pro další práci se zaměřením na tuto problematiku.

Při zavedení této metody analýzy záznamů do reálného provozu by mohlo docházet k získávání důležitých informací o provedené údržbě z historických dat organizace, které mohou být využity pro budoucí plánování údržby. Analýza by sloužila k odhalování specifických komponent a jejich závad, které se opakovaně vyskytují v údržbových záznamech v souvislosti s konkrétními provedenými pracemi na letadle. Se znalostí těchto dosud neznámých dat by bylo možné již při procesu plánování údržby předpokládat možnost výskytu neplánované závady během konkrétní prohlídky letadla a předem se na opravu připravit. Tím je myšleno přizpůsobit plán revize z hlediska harmonogramu prací a potřebného času na opravu, zvážit kapacitní možnosti organizace a dostupnost personálu, nebo také dopředu zajistit přítomnost nezbytného materiálu a nářadí potřebného k provedení opravy. Využití textové analýzy a získaných dat pro plánování údržby by bylo přínosem především z ekonomického hlediska. Mohlo by docházet k efektivnějšímu plánování údržby, zlepšení využití dostupných kapacit organizace a jejích zdrojů a v důsledku také ke snížení nákladů vynaložených na opravu. Přínos by mohl být také z hlediska bezpečnosti, vzhledem k možné predikci pravděpodobného výskytu závady na letadle již v procesu plánování údržby.

Z pohledu limitací této práce byly provedeny pouze dvě automatické textové analýzy dvou navzájem nezávislých sad údržbových záznamů. Další textové analýzy dostupných údržbových dat nebyly provedeny z hlediska rozsahu této práce a jejího zaměření na předložení návrhu automatického strojového zpracování dat. Výsledky obou provedených

analýz byly manuálně zpracovány a vyhodnoceny pouze z hlediska úspěšnosti označování pojmů, aby bylo možné určit s jakou přesností pracuje automatická textová analýza. Je potřeba uvést, že vzhledem k zaměření práce výsledky analýz také nebyly vyhodnoceny z pohledu statistiky výskytu pojmů, a proto v práci není uveden seznam nejčastěji se vyskytujících komponent a závad v údržbových záznamech.

Tato práce může sloužit jako základ pro budoucí vývoj a další využití možností textové analýzy v prostředí letecké údržby. Stejně tak může posloužit vytvořený Slovník komponent a závad, který obsahuje pojmy z vybrané sady údržbových záznamů a je pro provádění textových analýz nezbytný. Je možné ho využít jako podklad pro další práce se zaměřením na textovou analýzu údržbových záznamů letounů typu Boeing B737 a Airbus A320, nebo také po příslušných úpravách i pro jiné typy letadel. Slovník je možné upravovat a dále rozšiřovat pro potřeby vlastních analýz. Implementace automatické textové analýzy do praxe by mohla vést k získání důležitých, dosud neznámých informací z údržbových dat organizace. Zjištěné trendy na základě historických dat by mohly přispět k optimalizaci plánování údržby, odhalování a predikci neplánovaných závad a tím tak ke zvýšení celkové produktivity práce celého procesu údržby na technické základně.

## **Seznam příloh**

Příloha 1: Slovník komponent a závad

## Zdroje

- [1] The A, C and D of aircraft maintenance. *QANTAS News Room* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.qantasnewsroom.com.au/roo-ales/the-a-c-and-d-of-aircraft-maintenance/>
- [2] Nařízení Komise (EU) č. 1321/2014 ze dne 26. listopadu 2014 o zachování letové způsobilosti letadel a leteckých výrobků, letadlových částí a zařízení a schvalování organizací a personálu zapojených do těchto úkolů. *EUR - Lex, Access to European Union law* [online]. [cit. 2021-03-25]. Dostupné z: <https://eur-lex.europa.eu/legal-content/CS/TXT/PDF/?uri=CELEX:32014R1321&from=CS>
- [3] Základní informace plynoucí ze změny Nařízení Komise (EU) 1321/2014. *ÚŘAD PRO CIVILNÍ LETECTVÍ ČR* [online]. [cit. 2021-03-25]. Dostupné z: [https://www.caa.cz/wp-content/uploads/2020/01/Z%C3%A1kladn%C3%AD-informace-plynouc%C3%AD-ze-zm%C4%9Bny-Na%C5%99%C3%ADzen%C3%AD-1321\\_final.pdf](https://www.caa.cz/wp-content/uploads/2020/01/Z%C3%A1kladn%C3%AD-informace-plynouc%C3%AD-ze-zm%C4%9Bny-Na%C5%99%C3%ADzen%C3%AD-1321_final.pdf)
- [4] O společnosti Czech Airlines Technics (CSAT). *Czech Airline Technics (CSAT)* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.csatechnics.com/cs/about-us>
- [5] Výroční zpráva společnosti Czech Airline Technics, a.s. za rok 2019. *Czech Airline Technics (CSAT)* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.csatechnics.com/media/files/5/6/Vyrocni-zprava-Czech-Airlines-Technics-a.s.-2019.pdf>
- [6] Služby společnosti Czech Airlines Technics (CSAT). *Czech Airline Technics (CSAT)* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.csatechnics.com/cs/sluzby-2>
- [7] Aircraft Maintenance. *SKYbrary* [online]. [cit. 2021-03-25]. Dostupné z: [https://www.skybrary.aero/index.php/Aircraft\\_Maintenance](https://www.skybrary.aero/index.php/Aircraft_Maintenance)
- [8] Přijatelné způsoby průkazu a Poradenský materiál k Part-66 a Part-147 a postup CAA-ZLP-121. *ÚŘAD PRO CIVILNÍ LETECTVÍ ČR* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.caa.cz/wp-content/uploads/2019/07/121-2-AML-Part-66.pdf>
- [9] Complete List of Aircraft ATA 100 Chapters. *Aviation Sourcing Solutions* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.aviationsourcingsolutions.com/ata-chapters.aspx>
- [10] Maintenance Programme. *SKYbrary* [online]. [cit. 2021-03-25]. Dostupné z: [https://www.skybrary.aero/index.php/Maintenance\\_Programme](https://www.skybrary.aero/index.php/Maintenance_Programme)
- [11] AMOS Alternatives. *SOURCEFORGE* [online]. [cit. 2021-03-25]. Dostupné z: <https://sourceforge.net/software/product/AMOS/alternatives>
- [12] AMOS, An MRO software solution. *SWISS AviationSoftware* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.swiss-as.com/amos-mro>

- [13] Aeronautical Publications. *FLIGHT MECHANIC* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.flight-mechanic.com/publications/>
- [14] Text Analysis Examples. *WONDERFLOW* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.wonderflow.ai/blog/text-analysis-examples>
- [15] A Guide: Text Analysis, Text Analytics & Text Mining. *Towards Data Science* [online]. [cit. 2021-03-25]. Dostupné z: <https://towardsdatascience.com/a-guide-text-analysis-text-analytics-text-mining-f62df7b78747>
- [16] Natural Language Processing of Maintenance Records Data. *Digitála vetenskapliga arkivet DiVA* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.diva-portal.org/smash/get/diva2:975548/FULLTEXT01.pdf>
- [17] Edwards, Brett & Zatorsky, Michael & Nayak, Richi. (2008). Clustering and Classification of Maintenance Logs using Text Data Mining.. 87. 193-199. *ResearchGate* [online]. [cit. 2021-03-25]. Dostupné z: [https://www.researchgate.net/publication/221338019\\_Clustering\\_and\\_Classification\\_of\\_Maintenance\\_Logs\\_using\\_Text\\_Data\\_Mining](https://www.researchgate.net/publication/221338019_Clustering_and_Classification_of_Maintenance_Logs_using_Text_Data_Mining)
- [18] Logbook in the Cloud. *Aircraft Maintenance Technicians Logbook* [online]. [cit. 2021-03-25]. Dostupné z: <https://www.amtlogbook.com/index>
- [19] Ledvinka, Martin & Kremen, Petr & Saeeda, Lama & Blasko, Miroslav. (2020). TermIt: A Practical Semantic Vocabulary Manager. *ResearchGate* [online]. [cit. 2021-03-25]. Dostupné z: [https://www.researchgate.net/publication/341541783\\_TermIt\\_A\\_Practical\\_Semantic\\_Vocabulary\\_Manager](https://www.researchgate.net/publication/341541783_TermIt_A_Practical_Semantic_Vocabulary_Manager)
- [20] Mini-introduction to brat. *Brat rapid annotation tool (BRAT)* [online]. [cit. 2021-03-25]. Dostupné z: <https://brat.nlplab.org/introduction.html>
- [21] Overview and Introduction to GATE. *General achitecture for text engineering (GATE)* [online]. [cit. 2021-03-25]. Dostupné z: <https://gate.ac.uk/overview.html>
- [22] GATE Developer. *General achitecture for text engineering (GATE)* [online]. [cit. 2021-03-25]. Dostupné z: <https://gate.ac.uk/family/developer.html>