

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Šárka Hlušičková

Sada nástrojů pro manipulaci s bibliografickými údaji

Katedra softwarového inženýrství

Vedoucí bakalářské práce: RNDr. Michal Žemlička, Ph.D.

Studijní program: Informatika, obor Obecná informatika

2009

Ráda bych poděkovala svému vedoucímu, RNDr. Michalu Žemličkovi, Ph.D., za zajímavý námět pro bakalářskou práci a užitečné rady během vedení práce.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

v Praze
dne

Šárka Hlušičková

Obsah

1	Úvod	6
2	Přehled podporovaných bibliografických formátů	8
2.1	BibTeX	8
2.1.1	Formát .bib souboru	9
2.1.2	Zápis údajů v záznamech o dokumentech	10
2.1.3	Formát obsahu položek	10
2.1.4	Propojování záznamů o dokumentech	11
2.2	RIS format	12
2.3	Tagged „Endnote Import“ file	13
2.4	MARC 21	14
2.4.1	Struktura záznamu	15
2.4.2	Označení bibliografických dat	17
3	Části aplikace a celkový návrh řešení	18
3.1	Hlavní části aplikace	18
3.1.1	Konverzní nástroje	18
3.1.2	Nástroj pro tvorbu citací	19
3.1.3	Uložiště	19
3.2	Důvod vytvoření interního formátu	20
3.3	Volba způsobu vytváření citací	20
3.4	Omezení konverze	20
3.5	Volba způsobu konverze	20
3.6	Vliv způsobu konverze a tvorby citací na podobu interního formátu	21
4	Interní formát	22
4.1	Formát záznamu	22
4.2	Možnosti konfigurace	24
5	Řešení konverzí mezi formáty	25
5.1	Princip překlápění	25
5.2	Konfigurační soubory	25
5.3	Zrychlení konverze mezi formáty	28

5.4	Načítání BibTeX záznamů s referencí	29
5.5	Odlišnosti při importu formátu Marc 21	30
6	Vlastnosti konverzních nástrojů	32
6.1	Formátovací funkce	32
6.2	Obecné problémy při převodu mezi formáty	34
6.2.1	Import	34
6.2.2	Export	36
6.2.3	Konverze	36
6.3	Ukázky převodů záznamů mezi formáty	36
6.3.1	Import	36
6.4	Export	39
6.5	Konverze	40
7	Generování bibliografických citací	43
7.1	Vytvoření citace z interního záznamu	43
7.1.1	Typy pravidel v konfiguračním souboru	44
7.1.2	Příklady citací	46
7.2	Tvorba seznamů citací	47
7.2.1	Formát vstupního souboru	47
7.2.2	Formát konfiguračního souboru pro tvorbu seznamů	48
8	Závěr	50
	Literatura	52
A	Obsah přiloženého CD	54

Název práce: Sada nástrojů pro manipulaci s bibliografickými údaji

Autor: Šárka Hlušíčková

Katedra (ústav): Katedra softwarového inženýrství

Vedoucí bakalářské práce: RNDr. Michal Žemlička, Ph.D.

e-mail vedoucího: Michal.Zemlicka@mff.cuni.cz

Abstrakt: v předložené práci se zabýváme konverzí záznamů mezi několika vybranými bibliografickými formáty a tvorbou bibliografických citací. U konverzí jsme se zaměřili především na formáty, které jsou nativní pro nejrozšířenější citační aplikace nebo je lze snadno importovat. Konkrétně jsou to BibTeX format (BibTeX), RIS format (Reference Manager, Procite, EndNote) a Tagged „EndNote Import“ file (EndNote). Aplikace též podporuje konverze z formátu MARC 21, v němž jsou uloženy záznamy o dokumentech v databázích knihoven. Tvorba bibliografických citací se řídí pravidly norem ČSN ISO 960 a ČSN ISO 960-2 a je zaměřena na vytváření seznamů použité literatury, tzn. celých odkazů na dokumenty, nikoli na citace obsažené v textu.

Klíčová slova: bibliografický formát, konverze, bibliografické citace

Title: Tools for handling bibliographic data

Author: Šárka Hlušíčková

Department: Department of Software Engineering

Supervisor: RNDr. Michal Žemlička, Ph.D.

Supervisor's e-mail address: Michal.Zemlicka@mff.cuni.cz

Abstract: In this work we study conversions of bibliographic records between different bibliographic formats and making lists of bibliographic references. We have focused on the formats that are native for the most used citation managers or can be easily imported to them. The formats are the BibTeX format (BibTeX), the RIS format (Reference Manager, Procite, EndNote) and the Tagged "EndNote Import" file (EndNote). We also support conversions from the MARC 21 format, in which libraries store their records. The bibliographic references are created according to the rules of two standards ČSN ISO 960 and ČSN ISO 960-2, we do not consider making the in-text citations.

Keywords: bibliographic format, conversion, bibliographic reference

Kapitola 1

Úvod

Mnoho lidí se během své práce zabývá nejen prací samotnou, ale též publikací dosažených výsledků nebo opětovným zpracováním výsledků již známých, a právě zde naráží na potřebu vytvářet citace – uvádět prameny, ze kterých čerpali nebo se na ně ve své práci odkazují.

Údaje obsažené v citacích by měly být přesné a umožnit čtenáři jednoznačnou identifikaci použitého dokumentu. Požaduje se přehlednost a jednotný způsob uvádění citací v rámci dokumentu.

Citovat dokument lze několika způsoby. První možností je uvedení plného odkazu na použitý pramen již v textu. Druhou a častější možností je uvedení citací v textu a plných odkazů na jeho konci. Citace v textu jsou pouze krátké poznámky, jež odkazují na jiné místo do textu, kde je uvedený plný odkaz. Odkazy pak obsahují všechny důležité údaje pro identifikaci pramene. Soupisy odkazů bývá uvedeno nejčastěji pod názvem „Seznam použité literatury“ na úplném konci textu.

Z důvodu jednotnosti citování se ustálilo několik citačních stylů, některé z nich byly kodifikovány jako norma. Citačních stylů předepisují, která data se o dokumentech shromažďují a jakým způsobem se zapisují v citacích. Shromažďované údaje se mezi styly příliš neliší, údaje předepsané normou je obsáhlejší o volitelné prvky. Různí se ale zpracování jednotlivých údajů, zejména pořadí jejich zápisu. v České republice v současnosti platí pro citování norma, která má dvě části, a to ČSN ISO 960 [1] a ČSN ISO 960-2 [2]. Jde o převzaté mezinárodní normy ISO 960 [3] a ISO 960-2 [4].

Aby se práce s vytvářením citací zjednodušila, bylo vyvinuto mnoho programů, které jsou často označovány jako citační manažery. Ty uživatelům umožňují ukládat záznamy o jednotlivých dokumentech pro opakované použití, takže při citaci téhož pramene v několika dokumentech nemusí uživatel pokaždé vyhledávat potřebné údaje. Programy jsou schopné po vložení citací do textu automaticky vytvořit seznam odkazů a navíc na přání uživatele formátovat citace dle různých citačních stylů.

Nevýhodou pestrosti při výběru citačního manažeru je to, že si každý pro-

gram ukládá data v jiném formátu. a pokud si chtějí uživatelé různých citačních manažerů vyměňovat mezi sebou záznamy, pak je nutné záznamy konvertovat z jednoho bibliografického formátu do druhého.

V České republice se navíc uživatelé potýkají s problémem řazení národních znaků a neexistence citačního stylu dle platných norem ČSN ISO 960 [1] a ČSN ISO 960-2 [2]. Citační manažery bohužel nepodporují ani její mezinárodní verzi ISO 960 [3] a ISO 960-2 [4] a zaměřují se spíše na americké styly, neboť nejrozšířenější manažery (EndNote, ProCite, Reference Manager, BibTeX) jsou převážně amerického původu. Na druhou stranu jsou záznamy většinou distribuovány právě ve formátech těchto manažerů.

Jako výsledek této situace vznikla myšlenka napsat jednoduchý program pro správu bibliografických údajů, jenž by umožňoval záznamy ukládat a vytvářet seznamy odkazů dle normy ČSN ISO 690 a ČSN ISO 690-2. Do uložště by se záznamy přidávaly ručně nebo byly importovány z několika nejběžnějších formátů. Mezi tyto formáty byly vybrány BibTeX format (BibTeX), RIS format (Reference Manager, Procite, EndNote), Tagged „EndNote Import“ file (EndNote) a MARC 21 format. Záznamy z uložště by bylo možné do těchto formátů zase exportovat, s výjimkou formátu MARC 21, jenž by byl pouze importní.

Komerční programy často obsahují zásuvný modul do některého z komerčních editorů textu, právě kvůli vkládání citací do textu. v této práci jsem se rozhodla zaměřit se především na tvorbu odkazů, nikoli citací v textu, takže jako podporovaný editor jsem si vybrala program LaTeX, který mi toto umožní. Ráda bych ještě upozornila, že v dalších kapitolách místo dlouhého termínu „bibliografický odkaz“ užívám slovo „citace“.

Kapitola 2

Přehled podporovaných bibliografických formátů

Aplikace podporuje konverzi několika formátů. Ačkoli všechny slouží pro reprezentaci a výměnu bibliografických dat, daly by se rozdělit do dvou skupin dle jejich účelu. První skupinu tvoří formáty vytvořené pro konkrétní citační manažery, sem se řadí BibTeX formát (textový soubor .bib), RIS formát a Tagged „EndNote Import“ file. Druhou skupinou jsou formáty, v nichž ukládají svá data knihovníci, zástupcem této skupiny je formát MARC 21.

Formáty první skupiny slouží k uchování dat, jež jsou důležitá pro citování, případně dalších uživatelských poznámek a odkazů, které s citovanými dokumenty souvisejí. Záznamy si vytvářejí sami uživatelé nebo je přejímají od vydavatelů či z různých databází. Záznamy mají nekomplikovanou strukturu a snadno se s nimi pracuje. Na druhou stranu se na jejich obsah nelze úplně spolehnout.

Druhou skupinu zastupuje formát MARC 21, jenž vychází ze standardizovaných katalogizačních pravidel. Je poměrně robustní formát, takže s ním lze popsat opravdu každý dokument. Na druhou stranu jednotlivé položky mohou být provázané a katalogizační pravidla mají přednost před pravidly formátu. Tato skutečnost poněkud znesnadňuje strojovou interpretaci jednotlivých položek. Záznamy jsou vytvářeny knihovníky, některé položky procházejí validací a knihovna, jež záznam vytvořila nebo jej převzala, za něj ručí.

2.1 BibTeX

BibTeX je nástroj pro zpracování bibliografických citací, používaný společně LaTeXem, programem pro sazbu textu. BibTeX je ale také formát souborů s příponou .bib, které slouží stejnojmennému programu jako zdroj údajů o jednotlivých dokumentech.

2.1.1 Formát .bib souboru

Soubor .bib je textový. Protože se BibTeX používá společně s LaTeXem, soubor .bib by měl mít stejné kódování jako LaTeXový zdrojový soubor (přípona .tex). Standardní znakovou sadou je ASCII. Znaky, které nejsou obsaženy ve zvolené sadě, lze zapsat pomocí speciálních LaTeXových maker nebo sekvencí.

Každý záznam (entry) je uvozen znakem „@“, za ním pak následuje několika písmenný identifikátor typu záznamu (entry type), velikost písmen se nerozlišuje. Obsah záznamů se liší dle jejich typu, ale vždy je uzavřen ve složených nebo okrouhlých závorkách. Další text mezi záznamy BibTeX ignoruje. Pokud je v záznamu syntaktická chyba, BibTeX vypíše varování a přejde ke zpracování dalšího záznamu.

Záznamy lze rozdělit na čtyři základní druhy:

@STRING Pomocí tohoto typu záznamu lze definovat zkratky, jež lze opakovaně používat v obsahu položek. Definici lze uvést kdekoli v souboru, avšak vždy před použitím zkratky. Pro anglické názvy měsíců BibTeX nabízí standardní trojpísmenné zkratky.

Název zkratky by měl být tvořen písmeny, ale BibTeX připouští i názvy, které začínají písmenem ale obsahují číslice a další znaky jako třeba pomlčku či podtržítka. Velikost písmen se nerozlišuje. Oddělovačem mezi názvem zkratky a plným textem je rovnítko.

Plný text může být tvořen jedním nebo více částmi, které jsou spojeny znakem zřetězení „#“. Částí může být název již dříve definované zkratky nebo řetězec znaků uzavřený ve dvojitéch uvozovkách. Pokud se plný text skládá pouze z jednoho řetězce, může být řetězec ohraničen i složenými závorkami. Plný text nesmí obsahovat znaky složených závorek nebo dvojitéch uvozovek.

Příklad definic zkratky, druhá zkratka je definována pomocí první:

```
@STRING { ISBD = {International Standard Bibliographic  
Description}}  
@STRING { ISDBG = ISBD # }, part G}}
```

@PREAMBLE Obsah toho záznamu je určen LaTeXu. Mohou zde být definována nová makra LaTeXu, která se pak používají v položkách záznamů. Program BibTeX text pouze kopíruje na začátek výstupního souboru.

@COMMENT Tyto záznamy fungují jako jednořádkové komentáře.

@ARTICLE, @BOOK ... Všechny ostatní položky uvozené znakem „@“ se považují za záznamy o dokumentech, které jsou podrobněji popsány v následující sekci.

Příklad záznamu:

```
@BOOK{ kapitoly,  
author = {Ji{\v{r}}{\i} Matou{\v{s}}ek  
and Jaroslav Ne{\v{s}}et{\v{r}}il",  
title = "Kapitoly z~distr{\e}tn{\i} matematiky",  
edition = "2",  
publisher = "Nakladatelstv{\i} Karolinum",  
address = "Praha",  
year = 2003,  
note = {ISBN 80-246-0084-6 (2.opr.vyd.)}  
}
```

2.1.2 Zápis údajů v záznamech o dokumentech

Podle BibTeXu existuje pro každý standardní typ záznamu seznam povinných a nepovinných polí. Položky jsou v rámci záznamu neopakovatelné. Kompletní seznam typů záznamů a jejich položek, je uveden v manuálu k programu LaTeX [6], včetně krátkého doporučení o obsahu jednotlivých položek.

U číselných údajů se obvykle uvádí pouze čísla bez vysvětlivek. Například místo „no. 2“ se v položce number uvede pouze číslice a při výsledném formátování se zkratka „no.“ doplní automaticky, což ovšem předpokládá literaturu pouze v jednom jazyce, nejlépe anglickém.

Obsah polí, která obsahují údaje o autorech nebo editorech a obsahují více jmen, by měla mít obsah ohraničen dvojitými uvozovkami. Jednotlivá jména se pak oddělují slovem „and“, a pokud nejsou uvedena všechna, je jejich výčet ukončen slovy „and others“. Samotná jména je možné uvádět dvěma způsoby, v přirozeném nebo inverzním pořadí. Je-li jméno složitější, je lepší uvést ho v pořadí inverzním, aby bylo bylo jasné, které části jsou křestní jména a které jsou součástí příjmení. Inverzní způsob uvádění jména má obecně tvar: příjmení, přízvisko, křestní jména.

Příklady jmen:

```
Božena Němcová  
Erben, Karel Jaromír  
Havlíček Borovský, Karel  
Jones, Sr, David  
von Neumann, III, Hans
```

2.1.3 Formát obsahu položek

Název typu dokumentu (entry type) je uveden hned za znakem „@“. BibTeX má 15 standardních typů, pro každý z nich pak existuje seznam povinných a vo-

litelných polí vzhledem ke standardním formátovacím stylům BibTeXu (uloženy v souborech s příponou .bst). Na druhou stranu je možné styly upravit nebo si napsat vlastní, a tedy definovat vlastní typy dokumentů a vlastní pole. Pokud je použit styl, který dané pole nezná, jednoduše ho ignoruje. v případě neznámého typu dokumentu ignoruje celý záznam. Uživatelé této vlastnosti poměrně často využívají a vytvářejí vlastní položky a typy záznamů.

Hned za otevírací závorkou záznamu je uveden klíč, který musí být v celém souboru unikátní a slouží k identifikaci záznamu. Klíč (key) je následován čárkou. Jednotlivé položky jsou rovněž odděleny čárkami, čárka za poslední položkou je nepovinná. Mezi identifikátory, obsahy polí a oddělovači lze pro přehlednost psát mezery, tabulátory a znaky konce řádku, vše je bráno jako jediná mezera. Každá položka má identifikátor (keyword). Klíč a identifikátor položky by měly být složeny z písmen, ale ve skutečnosti mohou obsahovat číslice, podtržítka, dvojtečku, lomítko, pomlčku či plus. Velikost písmen se nerozlišuje.

Obsahem pole může být řetězec ve složených závorkách nebo dvojitých uvozovkách, ale také číslo, jež se do závorek či uvozovek uzavírat nemusí. Obsah pole může být i složený z více řetězců v dvojitých závorkách a předem definovaných zkratk. Jednotlivé části jsou pak spojeny znakem zřetězení „#“. Pokud řetězec ohraničíme dvojitými uvozovkami, nelze tento znak uvést jako součást řetězce. Pro závorky pak platí, že musí být správně uzávorkované, včetně těch zneplatněných lomítkem.

Použité znaky v položkách záznamů, stejně v plném textu zkratky, jsou omezeny pravidly TeXu. Znaky # \$ % & _ { } mají speciální význam, ale předchází-li jim znak \, svůj speciální význam ztrácí. Vlnovku, stříšku a zpětné lomítko jako obvyčejné znaky v textu použít nelze, ale je možné je zapsat příkazem \verb. Samotná vlnovka značí nedělitelnou mezeru. Znaky <> | lze psát pouze v tzv. matematickém módu, což znamená, že jsou obklopeny například dvojicí znaků \$.

V položkách se také mohou vyskytovat složené závorky, BibTeX jejich obsah nepracovává. Obvykle se takto označují velká písmena v názvech, jež mají jako velká zůstat i po výsledném zformátování, nebo se jimi ohraničují příkazy.

Položky také mohou obsahovat příkazy pro zpracování v LaTeXu, ty také bývají ohraničeny složenými závorkami. Mohou to být příkazy k formátování, ale též sekvence pro zápis znaků, které nejsou součástí použité znakové sady nebo příkazy pro vytvoření speciálních symbolů.

Podrobný popis formátování obsahu položek, včetně příkladů, lze najít v manuálu k programu LaTeX [5].

2.1.4 Propojování záznamů o dokumentech

Záznamy o dokumentech lze propojit. Máme-li například v databázi více článků z jednoho časopisu, stačí uvést společné údaje o časopisu pouze v jednom záznamu. Ostatní záznamy se pak budou na tento záznam odkazovat. Odkaz se vytvoří pomocí pole crossref, kde se jako hodnota uvede klíč záznamu, na nějž se odkazu-

jeme. Záznam pak převezme z odkazovaného záznamu všechna povinná a volitelná pole, jež zůstala neuvedena. Odkazovaný záznam je v souboru umístěn až za všemi záznamy, které se na něj odkazují.

Příklad propojení záznamů:

```
@INPROCEEDINGS{Lenzerini,  
author = "Maurizio Lenzerini",  
title = "Data Integration Is Harder than You Thought",  
crossref = "DBLP:conf/coopis/2001",  
pages = "22--26",  
}
```

```
@PROCEEDINGS{DBLP:conf/coopis/2001,  
editor = {Carlo Batini and Fausto Giunchiglia and  
          Paolo Giorgini and Massimo Mecella},  
title = {Cooperative Information Systems, 9th International  
         Conference, CoopIS 2001, Trento, Italy, September 5-7,  
         2001, Proceedings},  
booktitle = {CoopIS},  
publisher = {Springer},  
series    = {Lecture Notes in Computer Science},  
volume    = {2172},  
year      = {2001},  
isbn      = {3-540-42524-1},  
bibsource = {DBLP, http://dblp.uni-trier.de}  
}
```

2.2 RIS format

Tento formát byl vytvořen společností Research Information Systems, Inc. pro Reference Manager.

Záznamy se skládají z různého počtu položek (field). Každá položka začíná na novém řádku (ANSI 13 10) a je uvozena značkou (tag) určující její obsah. Všechny značky jsou 6-znakové: velké písmeno, velké písmeno nebo číslice, mezera (ANSI 32), mezera (ANSI 32), pomlčka (ANSI 45) a mezera (ANSI 32). První položka záznamu musí být se značkou „TY - “ a poslední se značkou „ER - “, ostatní mohou být v libovolném pořadí, některé položky se mohou v záznamu vyskytnout opakovaně.

Pokud délka dat v některém z polí přesahuje 70 znaků, je možné, ale nikoliv však nutné, vložit znak nového řádku (ANSI 13 10) za 70. znak a pokračovat na další řádce. Nicméně běžnou praxí je, že dlouhé položky jsou rozděleny na více

řádků pouze přibližně po 70 znacích a navíc další řádky bývají někdy odsazeny tabulátorem nebo mezerami na délku značky.

RIS používá znakovou sadu ANSI (Windows 1252 – West European Latin), nicméně EndNote při exportu do tohoto formátu zachovává znakovou sadu UTF-8 a v některých polích jsou obsaženy znaky konce řádků, např. pole s poznámkami nebo klíčovými slovy.

Příklad záznamu:

```
TY - CHAP
AU - Whitney, J.R.
AU - Billoski, T.V.
AU - Jones, V.R.
PY - 1997
BT - New Directions in Paleontology
ED - Billoski, T.V.
CT - Evidence for Triceratops in Antarctica
CY - New York
PB - Academic Press
SP - 24
EP - 27
N1 - Evidence for Triceratops in Antarctica
KW - Antarctica
KW - Triceratops
ER -
```

Úplný popis formátu, včetně všech značek a jejich použití, se nachází na webových stránkách programu Reference Manager [7] a v uživatelském manuálu [8] k programu lze najít i seznamy položek pro jednotlivé typy záznamů. Bohužel nelze tyto informace z důvodu ochrany vlastnických práv zařadit do této práce.

2.3 Tagged „Endnote Import“ file

Formát je odvozen od Refer/BibIX formátu a je určen pro import záznamů do programu Endnote. v souborech se používá znaková sada UTF-8.

Každá položka je uvozena značkou (tag), která je tvořena znakem „%“ a velkým písmenem. Jednotlivé záznamy jsou odděleny jedním nebo více prázdnými řádky.

Typ záznamu určuje značka v poli „%0“, typy záznamů jsou předdefinované. Úplný seznam typů předdefinovaných záznamů a značek lze najít v manuálu k programu EndNote [9] v sekci Creating a Tagged „EndNote Import“ File (s. 217 – 220). Přehled standardních typů záznamů programu EndNote a položek, jež obsahují, se nachází v tomtéž manuálu v příloze Appendix C, Reference Types and

Fields (s. 645–672). Tyto informace bohužel nelze z důvodu ochrany vlastnických práv zařadit do této práce.

Jediný typ pole, který má předepsaný formát, jsou pole s odpovědností, tj. se jmény autorů, editorů apod. Každé jméno má vlastní položku, tzn. pole, v nichž se jména vyskytují, jsou opakovatelná. Jména se zapisují v přirozeném pořadí nebo v invertované podobě. Křestní jméno je možné uvést celé nebo zapsat pouze iniciály.

Příklad záznamu:

```
%0 Book
%1 6448
%A Alfred V. Aho
%A Ravi Sethi
%A Jeffrey D. Ullman
%T Compilers: principles, techniques, and tools
%D 1986
%@ 0-201-10088-6
%P 796
%I Addison-Wesley Longman Publishing Co., Inc.
```

Přestože to není ve formátu uvedeno, každá položka je uvedena na zvláštním řádku, jak napovídá příklad. Nicméně některé položky přesahují jeden řádek, typicky se jedná o abstrakt, poznámky nebo klíčová slova. Další řádky víceřádkové položky nejsou odsazeny.

Endnote umožňuje upravovat názvy standardních typů záznamů a též množinu položek, které obsahují. Pokud má uživatel některé typy záznamů upraveny, měl by při importu používat nová pojmenování.

2.4 MARC 21

Název MARC je zkratkou slov „MACHine-Readable Cataloging“, což v překladu znamená katalogizování ve strojem čitelné podobě. MARC je ve skutečnosti celá skupina standardů a zahrnuje více formátů, jež jsou standardy pro reprezentaci a sdílení bibliografických a dalších s nimi spojených dat. MARC 21 formát pro bibliografická data (orig. MARC 21 format for Bibliographic Data) byl vytvořen na základě MARC formátů používaných v USA a Kanadě jako formát pro 21. století. v MARC 21 je možné ukládat informace o tištěných nebo rukopisných textech, počítačových souborech, mapách, hudbě, pokračujících zdrojích, vizuálních materiálech a smíšených dokumentech. v České republice s ním pracují snad všechny elektronické katalogy knihoven, takže MARC 21 záznamy jsou snadno přístupné.

Celý MARC 21 záznam má v podstatě tři úrovně pravidel, kterými se řídí: struktura záznamu, značení obsahu a samotný obsah záznamu. Struktura záznamu

je implementací mezinárodního standardu ISO 2709 (Format for Information Exchange) a jeho amerického protějšku ANSI/NISO Z39.2 (Bibliographic Information Interchange). Bibliografická data jsou v rámci záznamu rozčleněna do polí, příp. ještě podpolí, označena kódy (tag) z důvodu strojového zpracování. MARC 21 formát pro bibliografická data definuje právě tyto kódy, jež identifikují prvky údajů v bibliografických záznamech. Zpracování bibliografických dat je řízeno bibliografickými standardy, např. AACR2 (Anglo-American Cataloguing Rules), ISBD (International Standard Bibliographic Description) a dalšími, používá se také tezaurus a uplatňují místní pravidla knihovnických organizací. Každý záznam v sobě nese údaj o tom, kdo jej vytvořil, modifikoval a je vlastně odpovědný za jeho obsah.

2.4.1 Struktura záznamu

Záznamy jsou textové, používané jsou dvě znakové sady MARC8 a UTF-8. Záznam se má 3 základní části: návěští (leader), adresář (directory) a proměnná pole.

Návěští Obsahuje data pro zpracování záznamu, včetně délky celého záznamu a délek některých identifikátorů. Délka návěští je dána pevně, prvních 24 znaků záznamu (znak = 1 byte). Data jsou interpretována dle jejich pozice.

Adresář Je to seznam trojic o 12 znacích: kód (tag) proměnná pole – 3 znaky, délka obsahu pole – 4 znaky a relativní pozice pole vzhledem k začátku záznamu – 5 znaků. Číselné hodnoty jsou zleva doplněny nulami na potřebný počet znaků. V adresáři jsou nejprve uvedeny tzv. kontrolní pole (mají značku 00X, kde X je číslice), a to v rostoucích posloupnosti. Pak následují datová pole, jež jsou řazena vzestupně podle prvního znaku značky. Řazení datových polí v záznamu nemusí odpovídat řazení jejich značek v adresáři, duplicitní pole se rozlišují podle pozice. Adresář je ukončen znakem ASCII 0x1E (field terminator).

Proměnná pole Bibliografická data v MARC 21 záznamu jsou rozdělena do polí proměnných délek a každé z nich je označeno číselným kódem (tag) o délce tři. Kód není uveden přímo před polem, ale pouze v adresáři. Celý záznam je ukončen znakem ASCII 0x1D (record terminator). Rozlišujeme dva typy polí proměnné délky:

Kontrolní pole Tyto pole mají kód 00X, kde X je číslice od 1 do 9, a nejsou již dále podrozděleny. Obsahují buď jediný údaj, nebo více různých údajů, které jsou identifikovány pozicí v rámci pevné délky pole, jež přísluší konkrétnímu poli s daným kódem.

Datová pole Datová pole tvoří zbytek údajů v záznamu, jejich kód nezačíná dvěma nulami. Datová pole, narozdíl od polí kontrolních, obsahují dva jednoznakové indikátory (indicator), které jsou uvedeny na začátku pole.

Identifikátory zpřesňují povahu dat uložených v poli nebo je doplňují další informací. Identifikátory se interpretují odděleně, jsou to malá písmena abecedy nebo číslice. Identifikátor může být i nedefinován, pak se na jeho místo píše mezera ASCII SPACE, mezera ale může mít i konkrétní význam.

Datová pole jsou navíc podrozdělena a každé podpole (subfield) je uvozeno kódem podpole (subfield code). Kód podpole se skládá ze dvou znaků, oddělovače ASCII 0x1F a identifikátoru podpole, což je malé písmeno nebo číslice. Identifikátory podpolí jsou definovány zvlášť pro každé pole, ale pokud mají pole obdobný význam, jsou i podpole definována stejně. Pořadí podpolí není určeno alfabetským řazením identifikátorů, nýbrž je řízeno datovým obsahem a katalogizačními pravidly.

Záznamy se obvykle nezobrazují ve formátu dle Z39.2, ale v mnohem snadněji čitelné formě, kdy se vypouští adresář a kód pole je umístěn přímo před pole. Některé nezobrazitelné znaky se nahrazují zobrazitelnými, např. kód podpole ASCII 0x1F se zobrazuje jako „\$“ a mezera značící nedefinovaný identifikátor se jako #. Níže uvedený příklad (přejatý ze stránek formátu Marc [10]) vyznačuje prázdné místo pro systémem generované údaje hvězdičkami, jindy se používají pomlčky, záleží na konkrétní implementaci.

```
Leader/00-23   *****cam##22*****7a#4500
001  <control number>
003  <control number identifier>
005  19921113071152.5
008/00-39  870102s1962####at#####000#0#eng#d
040  ##$a[organization code]$c[organization code]
100  1#$aKelleher, James M.
245  10$aRoman fever :$ba story of the struggle by the Catholic
      people of New South Wales to remedy the injustice caused by the
      Education Acts of 1866 and 1880--an observer's opinion of the
      causes which so far have rendered that struggle in vain /$cby
      James M. Kelleher.
260  ##$aSurrey Hills, N.S.W. :$bCatholic Press Newspaper Co.,
      $c[1962?]
300  ##$a139 p. ;$c16 cm.
```

Další podrobnosti o výplňových znacích v polích apod. lze najít na stránkách formátu Marc 21 [11].

2.4.2 Označení bibliografických dat

Proměnná pole jsou uspořádána podle prvního číslice kódu (tagu) pole do bloků, které kromě několika výjimek identifikují bibliografickou funkci údaje v rámci záznamu. Zbylé dvě číslice pak zpřesňují o jaký údaj se jedná.

0XX Kontrolní informace, identifikační čísla, kontrolní pole, klasifikační znaky

1XX Hlavní záhlaví

2XX Názvy a údaje o vydání, nakladatelské údaje atd.

3XX Údaje fyzického popisu atd.

4XX Údaje o edici

5XX Poznámky

6XX Věcné selekční údaje

7XX Vedlejší záhlaví s výjimkou předmětů a edic; propojovací pole

8XX Vedlejší záhlaví pro edici, knihovní jednotky atd.

9XX Pole pro národní využití

V rámci bloků 1XX, 4XX, 6XX, 7XX a 8XX jsou v označení polí až na výjimky zachovány určité paralely.

X00 Osobní jméno

X10 Jméno korporace

X11 Jméno akce

X30 Unifikovaný název

X40 Bibliografické názvy

X50 Věcná témata

X51 Geografická jména

Záznam připouští násobnost všech polí a podpolí. Nicméně z podstaty některých ukládaných údajů vyplývá, že tato pole jsou označena jako neopakovatelná. Formát MARC 21 tuto informaci u všech polí a podpolí uvádí.

V záznamu lze pomocí podpolí \$8 jednotlivá pole propojovat. Lze propojit i celé záznamy pomocí názvu díla nebo identifikačních čísel.

Podrobné informace o polích, podpolích, jejich obsahu, opakovatelnosti atd. lze získat na stránkách MARC 21 standardu [10].

Kapitola 3

Části aplikace a celkový návrh řešení

Aplikace má dvě základní funkce: konverzi mezi vybranými formáty a tvorbu citací podle normy. První funkci zajistí konverzní nástroje, které záznamy dávkově zpracují. Implementace druhé funkce bude vyžadovat kromě nástroje pro tvorbu citací ještě uložičtě, aby záznamy mohly být využívány opakovaně. Jak jsem už zmínila v úvodu, zvolila jsem si LaTeX jako podporovaný editor, protože citace v textu si spravuje sám a požaduje pouze seznam plných odkazů. Aplikace tedy bude citace generovat též dávkově. Pro aplikaci tak postačí jednoduché ovládání pomocí příkazové řádky, navíc LaTeX je ovládán stejným způsobem.

Aplikace by se tedy dala rozdělit tři základní části: konverzní nástroje, nástroj pro tvorbu citací a uložičtě. v následujících odstavcích jsou popsány možné přístupy k řešení jednotlivých částí. v dalších sekcích této kapitoly je pak zdůvodněn výběr řešení, jež je v aplikaci implementováno.

3.1 Hlavní části aplikace

3.1.1 Konverzní nástroje

Konverzi mezi bibliografickými formáty lze provést v podstatě dvěma způsoby, přímo z výchozího formátu do cílového nebo všechny konverze provádět přes společný formát. Výhodou přímé konverze je, že ztrácí pouze ty informace, které cílový formát nepodporuje, zatímco u konverze přes společný formát se navíc ztrácí i informace, jež neuchovává společný formát. Nevýhodou přímé konverze je pak implementace převodu pro každou dvojici formátů.

Převod mezi dvěma formáty lze řešit pomocí pevně zabudovaných pravidel nebo také konfiguračními soubory, jež by konverzní pravidla obsahovala. Předností pevně nastavených pravidel, je převod záznamu jako celku. To umožňuje komplexní rozhodování na základě obsahu položek nebo jejich existence. Nevýhodou je staticnost tohoto rozhodování. Existují totiž formáty jako například BibTeX,

u nichž uživatelé mohou přidávat nové položky do záznamů nebo dokonce definovat vlastní typy. Zde je na místě použít převod pomocí konfiguračních souborů, kde může uživatel libovolně upravovat formát vstupních a výstupních záznamů. Nicméně čím složitější pravidla umožníme definovat v konfiguračním souboru, tím bude soubor pro uživatele méně přehledný.

Formáty, s nimiž by aplikace měla pracovat, tvoří jednu homogenní skupinu. Pro formáty citačních manažerů (BibTeX, EndNote, RIS) je vhodné implementovat převod pomocí konfiguračních souborů. U BibTeXu je to téměř nutností. Konverzní pravidla budou spíše jednoduchá, neboť seznam položek sice není pevný, ale údaje jsou elementární a jejich povaha je daná označením položky.

U formátu MARC 21 by byl lepším řešením převod funkcí. Kromě několika položek, jež jsou určeny pro národní použití, je obsah všech položek jednoznačně určen normou. Typ záznamu určuje nastavení v několika polích. Povaha některých položek je podmíněna obsahem jiné položky. a navíc údaje v položkách jsou prokládány interpunkcí podle katalogizačních pravidel.

3.1.2 Nástroj pro tvorbu citací

Při tvorbě citací podle normy ČSN ISO 960 a ČSN ISO 960-2 je důležité nejprve správně určit typ dokumentu. Norma totiž striktně rozlišuje mezi dokumenty tištěnými a elektronickými. U tištěných se omezuje pouze na dokumenty vydané a patenty. Dále se citace liší podle toho, jedná-li se o jednorázové dílo nebo o seriálovou publikaci. a též se rozlišuje dílo jako celek, části díla (např. kapitoly) a nezávislé příspěvky do díla. Formát, který by sloužil pro vytváření citací, by měl také samozřejmě uchovávat všechna potřebná data, minimálně však údaje, jež jsou v normě pro příslušný typ dokumentu označeny jako povinné. Navíc by tyto údaje měly být ve formě, ve které jsou uvedeny v dokumentu.

Z formátů, jež by aplikace měla podporovat, výše uvedené požadavky splňuje pouze formát MARC 21. Formáty citačních manažerů často obsahují údaje v té nejstručnější formě, z nichž není možné zpětně rekonstruovat přesný text uvedený v dokumentu. Navíc typy záznamů jsou podle obsahu dokumentu a ne vždy lze určit, zda je dílo vydané či má elektronickou podobu.

Samotné vytváření citací je možné řešit podobně jako konverzi, funkcí nebo konverzními soubory. Řešení s konfiguračními soubory jsou samozřejmě výhodnější, protože uživatel může nastavovat výskyt volitelných polí v citaci. Pokud by šlo v konfiguračním souboru nastavit i řazení a formát jednotlivých prvků, pak lze samozřejmě vytvářet i jiné citační styly pouze vytvořením jiného konfiguračního souboru.

3.1.3 Uložiště

Uložiště uchovává své záznamy pro opětovné použití. Přestože je možné uchovávat v uložisti záznamy v různých bibliografických formátech, je mnohem praktičtější

zvolit jeden konkrétní. Výhodou je rychlejší práce s aplikací při tvorbě citací, protože již není třeba záznamy konvertovat. a uživateli je mnohem příjemnější, může-li záznamy editovat a přidávat jedním způsobem.

3.2 Důvod vytvoření interního formátu

Jelikož součástí aplikace je uložení, bylo nutné vybrat jeden formát, ve kterém budou uloženy všechny záznamy. Formát zvolený pro ukládání by měl přirozeně reflektovat požadavky, které jsou uvedené v sekci o tvorbě citací. z formátů, jež aplikace podporuje, je to pouze formát MARC 21. Na druhou stranu je ale zbytečně robustní a poměrně komplikovaný, takže uživateli by se záznamy těžko vytvářely a spravovaly. Druhou možností je vzít jeden z formátů citačních manažerů a vytvořit si jiné, vhodnější typy záznamů a doplnit chybějící položky. To se ale vzhledem k počtu změn téměř rovná vytvoření nového formátu. Poslední variantou je vytvoření vlastního formátu na míru požadavkům. Právě tato varianta se mi jevila jako nejvýhodnější, s ohledem na fakt, že z formátů BibTeX, EndNote a RIS nevyhovoval ve stávající podobě žádný. v dalších sekcích a kapitolách je zde navrhovaný nový formát označován jako interní formát.

3.3 Volba způsobu vytváření citací

Citace jsou vytvářeny podle pravidel uvedených v konfiguračním souboru. Ačkoli formát souboru je složitější, možnost vytvořit citace i podle jiného citačního stylu, tuto nevýhodu plně vyvažuje. Uživatel si navíc může zvolit, zda chce, aby volitelná pole byla součástí citace.

3.4 Omezení konverze

Jak jsem se již zmínila, MARC 21 je velmi robustní a na tvorbu komplikovaných záznam. a právě vzhledem k povaze dat, jež obsahuje, je tento formát používán pouze pro import. Formát MARC 21 užívají především knihovny a knihovník by nepřejal záznam, za jehož správnost neručí jiná knihovna. Navíc s kusými údaji z formátů z citačních manažerů, lze jen těžko vytvořit záznam, který by mohl někdo používat, aniž by ho s dokumentem v ruce doplnil.

3.5 Volba způsobu konverze

Při návrhu konverzních nástrojů se s o ohledem na vytvoření interního formátu nabízela otázka, zda z něj neudělat společný formát pro konverzi. Jedním z důvodů je nutnost použití převodu pomocí konfiguračních souborů pro BibTeX a konverze přes společný formát zajistí rozumně udržitelné množství těchto konfiguračních

souborů. Interní formát je pak samozřejmě nutné rozšířit tak, aby se při převodu ztrácelo pouze minimum informací.

Převod z vnějšího formátu do interního a naopak se řídí konfiguračními soubory. Tyto soubory obsahují informace, odkud a kam se data přesouvají. Pokud je nutné data ještě nějak upravit, je kromě základní informace o přesunu ještě uvedeno jméno formátovací funkce. Funkce jsou součástí aplikace, nelze je přidávat ani modifikovat. Pole jsou zpracovávána odděleně, takže není možné vytvářet podmínky na existenci jiného pole nebo se řídit jeho obsahem.

Převod záznamu MARC 21 do interního formátu je realizován taktéž konfiguračním souborem, ale záznam je nejprve předzpracován. Funkce vytvoří řetězec, jež slouží jako umělý typ záznamu, z několika údajů v různých položkách. Dále jsou přidány umělé položky pro údaje, které nejsou elementární, a také pro ty, jež nejsou určeny pouze svým označením (tagem) ale i obsahem jiné položky. Tím odpadá nutnost podmínek v konfiguračním souboru. Konfigurační soubor je tak přehlednější a k uživateli přívětivější.

3.6 Vliv způsobu konverze a tvorby citací na podobu interního formátu

Interní formát byl nakonec vytvořen jako variabilní s pevným počtem typů záznamů a omezeným maximálním počtem položek v záznamu. Řešení vytváření citací a konverze mezi formáty pomocí konfiguračních souborů toto plně umožňuje. Aplikace pracuje se všemi položkami stejně a značí je pouze číselnými identifikátory. Obsah jednotlivých položek je dán nastavením právě v konfiguračních souborech. Stejně tak informace o číslech položek, nad nimiž jsou vybudovány indexy nebo mají speciální určení a tedy i formát, jsou zaneseny v konfiguračním souboru.

Výhodou je možnost volby obsahu položek. Počet položek je mírně naddimenzován, aby si uživatel mohl ukládat i své poznámky. Protože aplikace se neřídí textovými popisky, ale číselnými identifikátory, mohou být popisky v libovolném jazyce.

V defaultním nastavení položky se stejným číslem obsahují stejné nebo podobné údaje, pokud je to možné. Nicméně pro snadnou orientaci uživatele jsou pole pojmenována co nejpřesněji dle typu záznamu, například název zdrojového dokumentu, je u kapitoly v knize nazván název knihy a u článku název časopisu.

Nevýhodou této volby je nutnost udržení konzistence konfiguračních souborů pro správnou funkci aplikace.

Kapitola 4

Interní formát

Interní formát je v podstatě rozšířením klasických formátů citačních aplikací (BibTeX, EndNote „Import File“ Format, RIS) o údaje, které jsou důležité pro citování a v těchto formátech chybí. Rozdělení typů dokumentů je obdobné jako u formátů citačních manažerů, protože formát je zároveň používán jako mezistupeň při konverzi.

4.1 Formát záznamu

Záznamy používají znakovou sadu UTF-8. Každý záznam obsahuje typ dokumentu, dvě značky a jednotlivé položky. První značka určuje, zda byl dokument vydán, a druhá, zda má elektronickou podobu. Přípustné hodnoty značek jsou „Y“ a „N“ z anglických slov yes a no.

Každý typ záznamu obsahuje určitý počet položek dle svého typu. Maximální počet položek v záznamu je pro všechny typy záznamů stejný. Název položky je zvolen tak, aby napovídal uživateli, co má položka obsahovat. Položky jsou v rámci záznamu neopakovatelné. Pokud by tedy bylo třeba uložit více údajů stejného typu, jsou uloženy v jedné položce oddělené středníkem.

Údaje se do jednotlivých položek zapisují stejně jako u jiných formátů, speciálně formátové položky jsou uvedené níže.

Klíč Jednoznačně identifikuje dokument v rámci uložení. Posloupnost ASCII znaků – písmen, čísel a interpunkčních znaků.

Popisek Používá-li se aplikace společně s LaTeXem, tento popisek se může zobrazit jako součást citace v textu.

Pole s odpovědností Jména se zapisují v invertované podobě, ve tvaru „příjmení, přízvisko, křestní jména“. Jednotlivá jména se oddělují v rámci pole od sebe středníky bez mezer. Nejsou-li uvedena všechna jména, je za posledním jménem uvedeno znaménko +, též odděleno středníkem.

Datum Datum se uvádí ve formátu „rok/MM/DD/text“. Lomítka jsou uvedena vždy a určují délky jednotlivých částí. Měsíc a den se uvádí číselně podobě, jednomístné číslo je zleva doplněné nulou. Rok je obvykle tvaru „YYYY“, ale může být i tvaru „c2002“, jedná-li se o copyright, nebo u dokumentů vydávaných v rozpětí několika let ve tvaru „1989-2007“.

Pole s údaji o lokaci v dokumentu Obyčejně se v polích, které obsahují údaje jako svazek, číslo časopisu, ročník apod. uvádí pouze číslice. v rámci tohoto formátu z důvodu citování je nutné tyto údaje zapisovat včetně slovních zkratk, tedy 2. sv., 5. roč., 4 vol. atd. Výjimkou jsou údaje o stranách, ty se zapisují pouze číslicemi.

Chybějící údaje Pokud údaj neexistuje, je položka vyplněna znakem „#“. Aplikace pak nebude zbytečně upozorňovat na jejich doplnění.

Příklad záznamu v interním formátu. Všechny položky jsou jednořádkové, text je zalomen pouze z důvodu omezené šířky stránky.

```
05-ARTICLE [YY]
01-author_analytic: Catmull, E.;Clark J.
03-title_analytic: Recursively generated B-spline surfaces on
  arbitrary topological meshes
02-periodical: Computer-Aided Design
08-medium_designator:[online]
09-edition: #
12-place_published: #
13-publisher: Elsevier Ltd.
14-date: 1978
15-last_update: #
16-access_date: 2009/07/20/
17-volume: vol. 10
19-issue: no. 6 (november 1978)
20-pages: 350-355
21-series: #
22-series_volume: #
23-note:
25-ISSN: 0010-4485
26-DOI: #
28-URL: http://www.idi.ntnu.no/~fredrior/files/Catmull-Clark%20197
8%20Recursively%20generated%20surfaces.pdf
30-type_of_periodical: journal
31-language: eng
32-key: catmull_clark
33-keywords: Catmull-Clark;subdivision
```

4.2 Možnosti konfigurace

Typy dokumentů jsou v aplikaci označeny pouze číselnými identifikátory. Celkový počet typů je omezen. Názvy typů, tedy rozdělení dokumentů, jsou určeny nastavením v konfiguračním souboru „editFile.cfg“. Číslo typu dokumentu odpovídá řádku (číslováno od 0) v konfiguračním souboru a jeho název je první slovo uvedené na řádku. Pak následuje seznam položek ve tvaru „číslo položky, meze- ra, název položky, mezera, případná funkce pro načítání“. Číselné identifikátory označují položky, jež jsou přiřazeny k danému typu, včetně jejich názvů. Položky jsou předcházeny znakem #.

Příklad řádku z konfiguračního souboru „editFile.cfg“:

```
GENERIC#00 author_monographic#01 author_analytic#02 title_monogr  
aphic#03 title_analytic#04 editor_monographic#05 subsidiary_resp  
onsibility#06 organisation_name#07 organisation_place#08 medium_  
designator#09 edition#10 publishing#11 periodicity#12 place_publ  
ished#13 publisher#14 date D#15 last_update#16 access_date D#17  
volume#18 chapter#19 issue#20 pages#21 series#22 series_volume#2  
3 note#24 ISBN I#25 ISSN#26 DOI#27 number#28 URL#29 availability#3  
0 type_of_work#31 language#32 key K#33 keywords#34 label L#35 re  
search_notes#36 abstract#37 cited_docs#38 unused#39 unused
```

Před manipulací s interním formátem se informace ze souboru načtou do vnitřní struktury aplikace. Ta pak zajišťuje přepis číselných identifikátorů na slovní názvy nebo kontrolu formátu u zvolených položek při načítání. Pokud chce uživatel vytvořit nový záznam, slouží tyto informace k vytvoření šablony pro zvolený typ.

Kapitola 5

Řešení konverzí mezi formáty

Všechny konverze mezi podporovanými formáty jsou realizovány přes společný, interní formát. Konverzi, jejíž výchozí i cílový formát jsou podporované formáty, budu i nadále označovat jako konverzi. Konverzi z podporovaného formátu do interního formátu budu označovat jako import a konverzi z interního formátu do podporovaného jako export.

5.1 Princip překlápění

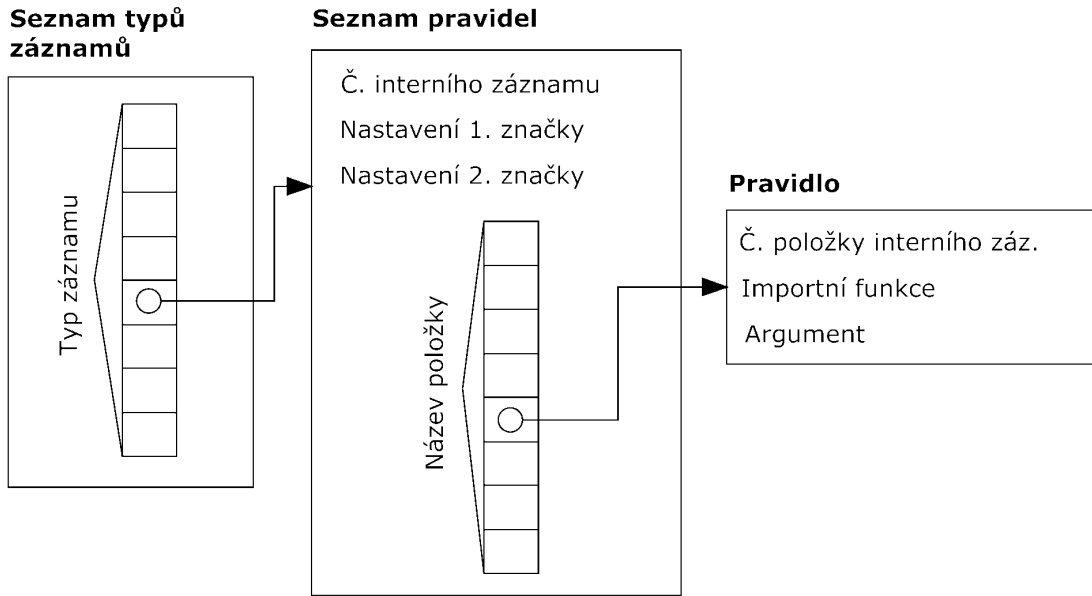
Konverze, import a export fungují podobně. Nejprve se načtou pravidla z konfiguračních souborů do paměti. Pak se ze zdrojového souboru načítá záznam po záznamu, okamžitě se překlápějí do cílového formátu a zapisují do cílového souboru. s výjimkou speciálního případu u BibTeXu, jenž je diskutován v sekci 5.4, je v paměti vždy jen jeden záznam.

V případě BibTeX záznamů, jež mohou obsahovat povolené LaTeXové příkazy a jinak upravené znaky, se po načtení tyto sekvence nahrazují odpovídajícími UTF-8 znaky. Při exportu do BibTeXu se naopak zase nahrazují znaky, jež nejsou ASCII.

Obrázek 5.1 ukazuje, jak vypadají vnitřní struktury pro import. v tabulce vlevo se nejprve podle typu záznamu nalezne příslušná subtabulka. Pokud neexistuje, nelze záznam tohoto typu překlápat. Při překlápění záznamu se pracuje pouze s příslušnou subtabulkou. Pro každou položku se nalezne pravidlo. v pravidle je uvedeno číslo cílové položky interního záznamu. Obsah se buď pouze přesune, nebo se uplatní importovací funkce s argumentem.

5.2 Konfigurační soubory

Pro každý podporovaný formát, s výjimkou MARC 21, existují dva konfigurační soubory, jeden pro import (převod do interního formátu) a druhý pro export



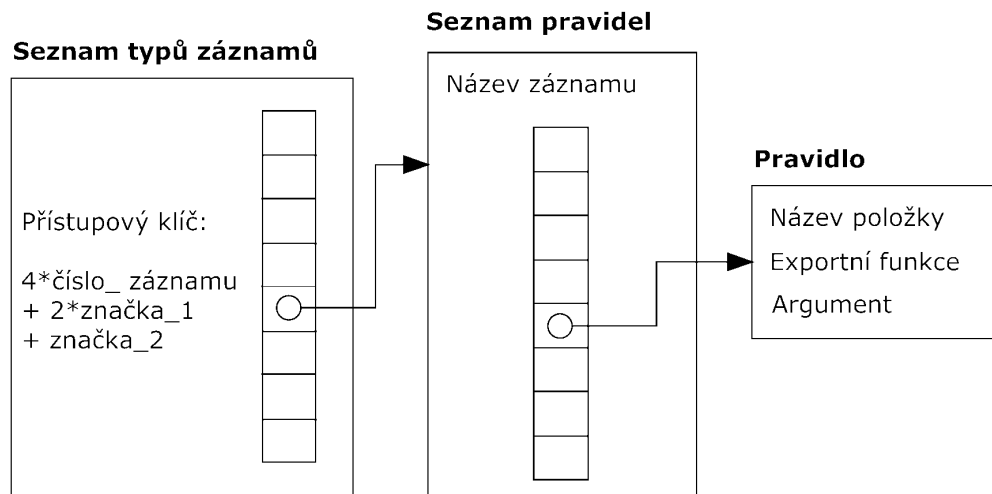
Obrázek 5.1: Vnitřní reprezentace importních pravidel

(převod z interního formátu). Soubory mají podobný formát, liší se především povolenými funkcemi.

Pro každý typ záznamu vstupního formátu je určeno, na jaký typ záznamu interního formátu se přepíše včetně nastavení, zda jde o vydaný nebo elektronický dokument. Pro jednotlivé položky jsou pak definována pravidla, do které položky interního záznamu se mají přepsat. Způsob přepisu určuje případná funkce v pravidle. Funkce může mít navíc i řetězový argument. Za komentáře jsou považovány řádky, jež mají na začátku znak „%“.

Příklad přepisovacích pravidel pro článek v časopisu z formátu BibTeX do interního formátu:

```
ARTICLE > 05-ARTICLE [YN]
author > 01-author_analytic (impAuthors "I| and |others")
title > 03-title_analytic
journal > 02-periodical
year > 14-date (toDate "YYYY")
month > 14-date (toDate)
volume > 17-volume
number > 19-issue
pages > 20-pages
note > 23-note
key > 32-key
```



Obrázek 5.2: Vnitřní reprezentace exportních pravidel

Jako první je uveden název výchozího záznamu, následuje většítko oddělené mezerami a za ním typ interního záznamu, zde rozhoduje číslice před pomlčkou, nikoli popisek. Za mezerou jsou v hranatých závorkách uvedeny značky, zda jde o vydaný nebo elektronický dokument. Značky jsou vyhodnocovány nezávisle, první označuje, zda byl dokument vydán, druhá, jestli má elektronickou podobu. Povolené hodnoty jsou „Y“ a „N“ z anglických slov yes a no. Na dalších řádcích následují pravidla pro položky obsažené v záznamu – název položky, mezerami oddělené většítko a označení cílové položky interního záznamu. Opět rozhoduje číslo, nikoli popisek. Pokud má být obsah položky nějak zpracován, následuje za označením cílové položky mezera a pak v kulatých závorkách jméno funkce. Případný argument je uveden v dvojitéch uvozovkách (angl.) a od názvu funkce je oddělen mezerou. Sekvence pravidel pro daný typ výchozího záznamu je zakončena prázdným řádkem.

Funkce umožňují základní formátovací úkony jako vložení položky na začátek, resp. na konec, nebo přidání popisku k obsahu. Dále jsou k dispozici importovací a exportovací funkce pro položky se speciálním formátem jako datum, isbn a položky se jmény osob. Přesný výčet povolených funkcí lze najít v příložené uživatelské dokumentaci.

Export z interního formátu probíhá obdobně, viz obrázek 5.2. Subtabulka se vyhledává podle čísla typu záznamu a nastavení značek, tj. subtabulek je čtyřikrát více než typů záznamů interního formátu a používá se přímý přístup pomocí výpočtu uvedeného v obrázku. Při překlápění záznamu se opět pracuje pouze se subtabulkou. Je-li položka interního záznamu neprázdná, vezme se pravidlo a použije se. Pravidla mohou obsahovat funkce. K pravidlům se přistupuje pomocí číselných identifikátorů, tedy přímo.

Příklad prepisovacích pravidel z interního formátu do formátu BibTeX:

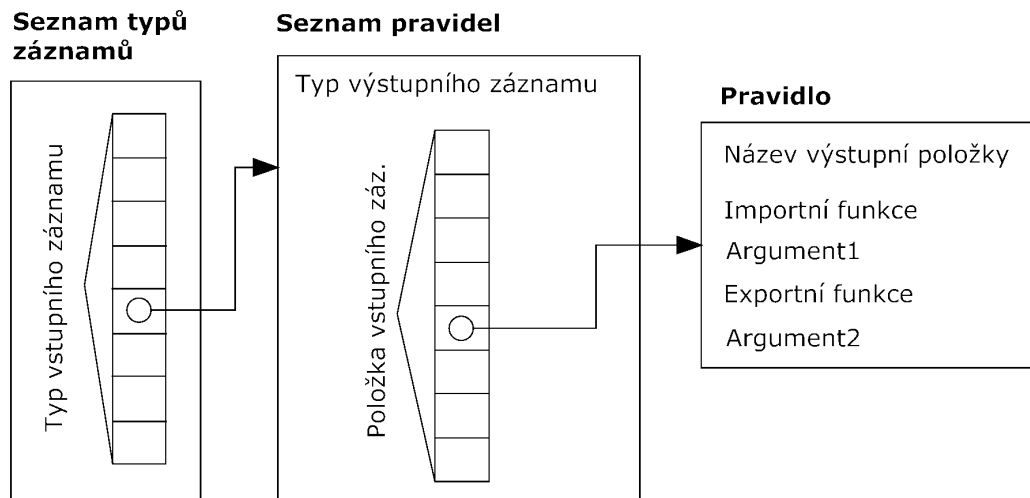
```
05-ARTICLE [--] > ARTICLE
01-author_analytic > autor (expAuthors " and |others")
03-title_analytic > title
02-periodical > journal
14-date > year (toDate "YYYY")
17-volume > volume (getNo)
19-issue > number (getNo)
20-pages > pages
23-note > note
25-ISSN > issn
26-DOI > doi
28-URL > url
32-key > key
36-abstract > abstract
```

V konfiguračním souboru je nejprve uveden typ interního záznamu, pak za mezerou jsou v hranatých závorkách uvedeny značky, jestli jde o vydaný nebo elektronický dokument. Značky jsou opět vyhodnocovány nezávisle, povolené hodnoty jsou „Y“, „N“ a navíc „-“, která označuje, že tento soubor pravidel bude aplikován nezávisle na hodnotě značky. Následuje většítka oddělené mezerami a za ním název cílového záznamu. Na dalších řádcích jsou opět pravidla pro položky obsažené v záznamu – označení cílové položky interního záznamu, opět rozhoduje číslo, mezerami oddělené většítka a název cílové položky. Pokud má být obsah položky nějak zpracován, je uvedeno ještě jméno funkce, formát je stejný jako u konfiguračního souboru pro import. Sekvence pravidel pro daný typ interního záznamu je zakončena prázdným řádkem.

5.3 Zrychlení konverze mezi formáty

Následující obrázky naznačují, jak vypadají vnitřní struktury pro uchování pravidel. Konfigurační soubory pro různé formáty mají stejnou syntaxi, takže informace jsou uchovávány v jednotných strukturách a načítány stejnou funkcí. Celé překlápění je též napsáno jednotně, rozlišuje se pouze import, export a konverze. Záznamy mají vytvořená jednotná rozhraní a odlišnosti jsou překlenuty pomocí dědičnosti.

Právě tento jednotný přístup umožňuje zrychlit konverzi. Každá konverze se skládá z importu a následného exportu, při nichž se používá interní formát, který je jako pomocná struktura v závěru vymazán. Zrychlení se dá dosáhnout, jestliže se bude zapisovat přímo do výsledného záznamu a interní záznam se vůbec tvořit nebude. Toho se lze dosáhnout sloučením struktur pro import a export do jedné konverzní tabulky. Podmínkou je ale právě jednotnost tabulek.



Obrázek 5.3: Vnitřní reprezentace konverzních pravidel

Způsob sloučení naznačuje následující obrázek 5.2. Nejprve se vytvoří struktury pro import i export. z nich se pak vytvoří jedna struktura pro konverzi. Prochází se tabulka pro import, její klíče (název typu vstupního záznamu) se převádí do tabulky pro konverzi a subtabulky jsou nahrazeny subtabulkami pro konverzi. Pro každou importní subtabulku se nalezne exportní subtabulka podle čísla interního záznamu a nastavení značek. Do konverzní subtabulky se z exportní subtabulky převezme název výstupního záznamu a z importní subtabulky klíče k seznamu pravidel (názvy položek vstupního záznamu). Importní pravidla jsou nahrazena konverzními. K importnímu pravidlu se nalezne exportní pravidlo podle čísla položky interního záznamu. Do konverzního pravidla se z importního pravidla převezme importní funkce a argument a z exportního pravidla exportní funkce, argument a název položky výstupního formátu. Při určitých kombinacích jsou funkce a jejich argumenty nahrazeny nebo vyměněno jejich pořadí tak, aby bylo dosaženo požadovaného výsledku.

Konverze se může zrychlit i zmenšením prohledávaných struktur. Po sloučení mohou totiž ze subtabulek pravidla pro některé položky vstupního záznamu vypadnout, pokud se nenažde odpovídající exportní pravidlo. Teoreticky tak může vypadnout i celá subtabulka pro nějaký typ vstupního záznamu.

5.4 Načítání BibTeX záznamů s referencí

BibTeX záznamy obsahující pole „crossref“ nejsou úplné. Odkazují se záznam, jehož klíč je v tomto poli uveden. Tento záznam je pak uveden za všemi záznamy, které se na něj odkazují. Ve chvíli, kdy aplikace načte neúplný záznam, nemůže jej hned překlomit. Při načítání BibTeXu se tedy používá pomocná struktura, v níž

jsou tyto neúplné záznamy uvedeny. Při načtení nového záznamu se zkontroluje, zda obsahuje pole „crossref“, a pokud ano, je záznam přidán do seznamu. Jestliže takové pole neobsahuje, pak je záznam úplný a je možné jej překlopit. Nejprve se ale ověří, jestli se na něj neodkazují nějaké neúplné záznamy. Když ano, dojde k jejich doplnění. Doplněné záznamy se spolu s aktuálně načteným záznamem překloupí.

Toto je jediný případ, kdy je v paměti více než jeden záznam. Nicméně záznamy bývají velice krátké, takže maximální počet záznamů v paměti není v aplikaci omezen.

5.5 Odlišnosti při importu formátu Marc 21

Záznamy formátu MARC 21 se liší od ostatních podporovaných formátů. z důvodu vybudování společného rozhraní pro všechny podporované formáty, je třeba záznam po načtení upravit. Záznamy MARC 21 nemají určený typ záznamu v jedné položce, takže je třeba tuto položku uměle vytvořit. Dále jsou přidány do záznamu informačně duplicitní položky pro podpole, jež obsahují údaje, které nejsou elementární, a pro údaje, jež jsou interpretovány podle jiného podpole. Nově přidané položky jsou už elementární a jejich označení vystihuje přesněji povahu obsahu.

V souboru pro import MARC 21 jsou jako název typu záznamu uvedeny masky. Narozdíl od ostatních souborů, v tomto je důležité, jak jdou masky za sebou, neboť ve shodném pořadí jsou uvedeny i ve vnitřní struktuře. Při hledání pravidel pro import, se pak tyto masky postupně prochází, dokud se nenajde odpovídající danému umělému typu. Délka masky a umělého typu záznamu je shodná. Znak „-“ v masce znamená jakýkoli znak.

V konfiguračním souboru je vždy uvedeno pole následováno kódem pro podpole, import je totiž omezen na proměnná datová pole, kontrolní pole a návěští zahrnuté nejsou.

Příklad importovacích pravidel pro formát MARC 21, pravidla s tagy z malých písmen se týkají právě umělých položek:

```
am-m----- > 10-THESIS [YN]
disp > 00-author
thsp > 04-teacher
autc > 06-school_name
245a > 02-title (pushFront " : ")
245b > 02-title (pushBack " : ")
020a > 24-ISBN
250a > 09-edition
260a > 12-place_published
260b > 13-publisher
260c > 14-date (toDate)
```

300a > 20-extent
440a > 21-series
440v > 22-series_volume
490a > 21-series
490v > 22-series_volume
500a > 23-note (pushFront " ")
502a > 23-note (pushBack " ")
041a > 31-language
655a > 30-type_of_work

Kapitola 6

Vlastnosti konverzních nástrojů

V této kapitole jsou shrnuty výsledky pokusných převodů mezi formáty. První část se zaměřuje na formátovací funkce a jejich chování. Druhá část se zabývá problémy vyplývajícími z vlastností jednotlivých formátů. A poslední část je věnována příkladům. Podrobnosti ohledně načítání (znakové sady, povolené znaky) a výstupu jsou uvedeny v uživatelské dokumentaci.

6.1 Formátovací funkce

Aplikace umožňuje přímý převod položek bez úprav nebo převod pomocí formátovacích funkcí. Formátovací funkce dovolují následující úpravy:

Slévání položek Převáděná položka se přidá do pole. Obsahuje-li pole nějaké údaje, přidá se položka na začátek nebo konec pole dle zvolené funkce. Nově přidanou položku je možné oddělit od původních údajů zvoleným oddělovačem.

Spojování násobných položek Násobné položky téhož druhu jsou přidávány do pole a oddělovány středníkem bez mezer. Tato funkce je dostupná pouze při importu.

Rozdělování položek s násobnými údaji Násobné údaje obsažené v jedné položce jsou rozděleny do více polí se stejným označením (tagem), jako hranice mezi údaji slouží středník. Tato funkce je dostupná pouze při exportu.

Popisky Popisek je možné přidat na začátek nebo na konec převáděné položky.

Formátování seznamu jmen Pro import jmen v polích s odpovědností jsou určeny dvě funkce. První funkce, která jména postupně přidává a odděluje je od sebe středníkem, se používá u formátů s násobnými položkami. Druhá funkce přebírá všechna jména v jedné položce a pouze vymění původní

oddělovače za středníky a případnou značku pro neúplný seznam za značku interního záznamu (+). V obou funkcích je možné nastavit, aby se jména při importu převedla do inverzní podoby.

Příklad:

Michel Chaudron and Kees van Hee Chaudron, Michel;van Hee, Kees;

W. M. P. van der Aalst and others 04 editor: van der Aalst, W. M. P.;

A1 - Laumann,E.O. Laumann,E.O.;Gagnon,J.H.;Michaels,
A1 - Gagnon,J.H. S.;Michael,R.T.;Coleman,J.S.
A1 - Michaels,S.
A1 - Michael,R.T.
A1 - Coleman,J.S.

Úprava jmen při exportu záleží na cílovém formátu, buď jsou jména rozdělena podle středníků do více polí, nebo jsou jen vyměněny oddělovače a značka pro neúplný seznam. Jména zůstávají v invertované podobě.

Datum Vzhledem k tomu, že interní formát má předepsaný tvar data, jsou položky s daty převáděny pomocí funkcí. Připomeňme, že formát data interního záznamu je „rok/MM/DD/text“, kde v části rok může být klasický tvar „YYYY“, copyright „c2002“ nebo rozsah „2004-2005“.

Importní funkce je schopna importovat každé datum, pokud se jí nepodaří vstupní údaje rozdělit, zařadí je do části text. Pomocí argumentu lze funkci upozornit, že se jedná pouze o údaj roku nebo že jde o datum americké, tj. měsíc předchází den.

Příklad dat před a po importu:

c2006	c2006///
2003/4/5	2003/04/05/
20030521	2003/05/21/
April 22, 2003	2003///April 22,
05/21/89	1989/05/21/

Pokud záznam obsahuje datum ve dvou částech dojde k postupnému doplnění.

2007	2007///
May	2007///May
August 13	///August 13
1998	1998///August 13

Datum je možné exportovat ve standardní podobě „YYYY-MM-DD“, ve formátu pro RIS „YYYY/MM/DD/text“ nebo americkém „MM/DD/YY“. Lze zvolit i export s rozdělením data „sepYear|tag“, kde rok se uloží do pole, k němuž se váže pravidlo, zbylá část data do pole „tag“.

Příklady exportu dat s různými maskami. v pravém sloupci je nejprve uvedeno datum v interním formátu a pod ním to samé datum po exportu s různými maskami. Masky jsou uvedeny vlevo od exportovaných datumů.

interní formát	„2004/08/03/3. srpna“
YYYY-MM-DD	„2004-08-03“
YYYY/MM/DD/text	„2004/08/03/3.srpna“
MM/DD/YY	„08/03-04“
sepYear tag	„2004“ a v poli tag „3.srpna“
bez masky	„3. srpna 2004“
interní formát	„c2004///“
YYYY-MM-DD	„2004“
YYYY/MM/DD/text	„2004“
MM/DD/YY	„2004“
sepYear tag	„2004“ a pole tag se nevytvoří
bez masky	„c2004“

ISBN Funkce určená pro import ISBN kontroluje jeho délku a počet obsažených pomlček. Přísnější nastavení (určeno argumentem funkce) pak kontroluje i správnost kontrolní číslice, pomlčky odstraní a číslo opět rozdělí. V případě, že je na vstupu správné, ale nerozdělené ISBN, funkce pomlčky doplní. Pokud je číslo chybné, připojí k němu značku „(sic)“.

Získání čísla z položky Tato funkce je určena především pro export polí, jako je číslo edice, ročníku atd., do formátů, jež tyto údaje uvádí pouze jako číselné bez vysvětlujících zkratk. Pokud položka číslo neobsahuje, převezme se celý její obsah.

Podrobný přehled včetně názvů funkcí, možných argumentů a popisu chování se nachází v uživatelské dokumentaci.

6.2 Obecné problémy při převodu mezi formáty

6.2.1 Import

Základním problémem je, že formáty různě rozdělují dokumenty. Při přechodu z jemnějšího rozdělení do méně jemného se některé informace obsažené v názvu

typu mohou ztratit. V opačném případě se část záznamů zbytečně zařadí pod obecnější typ.

Při importu může druhý případ nastat u formátu Marc 21. V záznamech totiž nebývají vyplněna některá důležitá kontrolní pole a dokument má tak obecnější zařazení. Ve výsledku pak nejsou použita pravidla specifická pro daný dokument a v importovaném záznamu chybí údaje. Jedinou možností je doplnění specifických pravidel do obecnějších typů.

Občas může dojít ke zcela špatnému určení typu, hlavně částí dokumentů. Ve formátu Marc 21 se nedá z obsahu kontrolních polí rozlišit příspěvek v monografii a článek v periodiku. Podobný problém je u formátů RIS a EndNote "Import File" format, kde splývají kapitola a příspěvek v monografii.

Rozdělení dokumentů také málokdy reflektuje, zda byl dokument vydán a zda má elektronickou podobu. Toto nastavení by měl po importu uživatel zkontrolovat, stejně jako typ dokumentu, neboť tyto údaje jsou důležité pro citace.

Potíže se mohou vyskytnout i u importu jednotlivých položek.

Nejčastějším problémem jsou položky pro číslo edice, ročníku apod. Obvykle se totiž zapisují pouze číslice bez příslušných zkratk. Norma [1, 2] ale požaduje i uvedení zkratky. Tato chybějící slova nelze doplnit automaticky, protože se dokumenty liší jazykem a někdy připadá v úvahu i více zkratk (svazek, ročník...). Uživateli pak nezbyvá než zkratky doplnit podle dokumentu.

Při automatickém invertování jmen je možné též chybovat. Například RIS formát a EndNote „Import File“ Format mají pro jména osob a názvy korporací společnou položku, takže se snadno stane, že aplikace invertuje název korporace.

Potíže mohou působit i pole s místem vydání. Norma předepisuje, že místo vydání je možné upřesnit přidáním zkratky státu uvedené v kulatých závorkách. v záznamech se ale tyto zpřesňující údaje vyskytují spíše v nezkrácené podobě, oddělené čárkou nebo středníkem. Úpravu pole nelze provádět automaticky, protože místo a zpřesňující údaj bývají různém pořadí. Takže tuto úpravu musí uživatel zajistit sám.

Na uživateli je také rozřídění obsahu položek z poznámkami. Pokud se v záznamu pro údaj nevyskytuje příslušná položka, obvykle je údaj vložen do poznámky. Uživatelé BibTeXu, kteří se drží standardních polí, by do poznámek řadili i ISBN, ISSN a další důležité údaje.

Nakonec bych chtěla zmínit ještě otázku násobných nakladatelských údajů. Interní formát nepřipouští opakování polí. Je možné v něm zapsat záznam o dokumentu s více místy vydání, aniž by to negativně ovlivnilo tvorbu citací. Pokud ale dokument vyšel na více místech v různých nakladatelstvích, pak tyto informace interní formát není schopen uložit. Na druhou stranu to však neovlivňuje kvalitu citací, protože podle normy stačí uvést nejvýraznější nebo první místo vydání a k němu příslušného nakladatele.

6.2.2 Export

Při exportu se opět potýkáme s problémem různého rozdělení dokumentů v jednotlivých formátech, a to při přechodu z méně jemného do jemnějšího rozdělení. Záznam se pak zařadí k obecnějšímu nebo vůbec chybnému typu a položky se mohou uložit do nesprávných polí.

Jediným řešením tohoto problému je zavést rozhodování o výstupním typu záznamu i podle toho, která pole jsou v záznamu vyplněna a umožnit zápis těchto pravidel v konfiguračním souboru. Typ interního záznamu a nastavení značek je plně postačující pouze tehdy, pokud je rozdělení interního záznamu stejně jemné nebo jemnější než rozdělení všech výstupních formátů.

Možným problémem při exportu jsou položky, které výstupní formát požaduje jako čistě číselné. Pokud údaj, který do takové položky exportujeme neobsahuje číslo, pak se do položky vloží text. Například, uvedeme-li vydání jako „2. vyd“, vloží se dvojka, v případě slovního vyjádření se položka okopíruje celá. Výstupní položka tak sice nesplňuje formát, ale je to lepší, než když údaj vypustíme úplně.

6.2.3 Konverze

V konverzi se nejvíce projeví slabé stránky importu a exportu, což je cena za konverzi přes společný formát. Předejít kumulativním chybám se dá pouze přímou konverzí. Pokud však uživatel bude používat import pro vkládání nových záznamů, které si před dalším použitím zkontroluje, a export pro vykazování dokumentů v příslušném formátu, kumulativní chyby se neprojeví.

Vzhledem k tomu, že inicializace konverze probíhá spojením pravidel pro import a export, může se stát, že funkce nepůjdou spojit tak, aby byl zachován původní význam pravidel. Uživatel je upozorněn, ve kterých typech záznamů k tomu může dojít. Kompletní výčet kombinací funkcí, které způsobují tyto potíže lze najít v uživatelské dokumentaci.

K výše uvedené situaci může dojít například tak, že import do položky, probíhá spojením obsahu několika položek vstupního záznamu a export do cílové položky probíhá rovněž spojením několika položek (interního záznamu). Pak nelze zaručit správné pořadí v cílové položce ani použití správných oddělovačů. Použití nesprávného oddělovače je způsobeno buď tím, že některá z spojovaných položek chybí. Špatné pořadí ve výstupní položce zase nemožností určit pořadí převodu vstupních položek.

6.3 Ukázky převodů záznamů mezi formáty

6.3.1 Import

Příklady importu BibTeX záznamů. Položky interního záznamu jsou pro přehlednost zarovnané k odpovídajícím položkám původního záznamu, všechny položky

interního záznamu jsou jednořádkové. Pokud uživatel používá pouze standardní sadu položek a typů záznamů, bude pravděpodobně nutné některé údaje doplnit, aby záznam bylo možné použít pro tvorbu citací.

```

@inproceedings{
  ChaudronAspol,
  author = "Michel Chaudron and
  Kees van Hee and Lou Somers",
  title = "Use Cases as Workflows",
  editor = "W. M. P. van der Aalst
  and others",
  booktitle = "BPM 2003",
  series = "LNCS",
  volume = 2678,
  pages = "88--103",
  year = 2003,
  publisher = "Springer Verlag",
  address = "Berlin Heidelberg"
}
07-INPROCEEDINGS [YN]
32-key: ChaudronAspol
01-author: Chaudron, Michel;
van Hee, Kees;Somers, Lou
03-title: Use Cases as Workflows
04-editor: van der Aalst, W. M. P.:+
02-proceedings: BPM 2003
21-series: LNCS
22-series_volume: 2678
20-pages: 88-103
14-date: 2003///
13-publisher: Springer Verlag
12-place_published: Berlin Heidelberg

```

Příklad importu záznamu ve formátu EndNote "Import File"Format. Položky interního záznamu jsou pro přehlednost zarovnány k odpovídajícím položkám původního záznamu, všechny položky jsou jednořádkové.

```

%0 Electronic Article
%A Polly, P. David
%D 2004
%T On the Simulation of the
Evolution of Morphological
Shape: Multivariate Shape
Under Selection and Drift
%B Palaeontologia Electronica
%V 7
%N 2
%8 April 19, 2005
%@ 1094-8074
%R 7.2.7A
%U http://palaeo-electronica
.org/2004_2/evo/issue2_04.htm
%G English
05-ARTICLE [NY]
01-author: Polly, P. David
14-date: 2004///
03-title: On the Simulation of the
Evolution of Morphological Shape:
Multivariate Shape Under Selection
and Drift
02-periodical: Palaeontologia Electronica
17-volume: 7
19-issue: 2
16-access_date: 2005///April 19,
25-ISSN: 1094-8074
26-DOI: 7.2.7A
28-URL: http://palaeo-electronica.org/
2004_2/evo/issue2_04.htm
31-language: English

```

Příklad importu záznamu z formátu Marc 21, pro větší přehlednost jsou záznamy psány za sebou. Položky výstupního záznamu jsou jednořádkové, text zalomen pouze kvůli omezené šířce stránky.

```
LDR -----nam-a22-----a-4500
FMT BK
001 cpk20041311118
003 CZ-PrNK
005 20060510142920.0
007 ta
008 040408s2004----xr-a---e-----101-0-eng--
020 |a 80-86732-19-3 (brož.)
040 |a ABA001 |b cze
072 7 |a 004 |x Počítačová věda. Výpočetní technika. Informační
      |a technologie |2 Konspekt |9 23
080 |a 004 |2 MRF
080 |a (062.534) |2 MRF
1112 |a SOFSEM 2004 |c (Měřín, Žd'ár nad Sázavou, Česko)
      |7 kn20040420018
24510 |a SOFSEM 2004 :
      |b theory and practice of computer science : 30th
      |b conference on Current Trends in Theory and Practice of
      |b Computer Science : Měřín, Czech Republic, January 2004 :
      |b proceedings. |n Volume II /
      |c Peter Van Emde Boas ... [et al.] (eds.)
24630 |a Theory and practice of computer science
24630 |a 30th conference on Current Trends in Theory and Practice
      |a of Computer Science
24630 |a Conference on Current Trends in Theory and Practice of
      |a Computer Science
250 |a Vyd. 1.
260 |a Prague : |b Matfyzpress, |c 2004
300 |a ix, 263 s. ; |c 24 cm
504 |a Obsahuje bibliografie, bibliografické odkazy a rejstřík
65007 |a počítačová věda |2 czenas
65009 |a computer science |2 eczenas
655 7 |a sborníky konferencí |2 czenas
655 9 |a proceedings of conferences |2 eczenas
7001 |a Emde Boas, Peter van |7 xx0018499 |4 edt
901 |f 1. vyd. |o 20040401
9101 |a ABA001
SYS 001311118
```

06-PROCEEDINGS [YN]
 02-title: SOFSEM 2004 : theory and practice of computer science : 30th conference on Current Trends in Theory and Practice of Computer Science : Měříň, Czech Republic, January 2004 : proceedings. Volume II
 04-editor: Emde Boas, Peter van;+
 09-edition: Vyd. 1.
 12-place_published: Prague
 13-publisher: Matfyzpress
 14-date: 2004///
 20-extent: ix, 263 s.
 24-ISBN: 80-86732-19-3 (brož.)
 30-: sborníky konferencí
 38-conference_name: SOFSEM 2004
 39-venue: (Měříň, Žďár nad Sázavou, Česko)

6.4 Export

Příklad exportu z interního záznamu do formátu RIS:

06-PROCEEDINGS [YN]	TY - CONF
02-title: Database and expert systems applications : 12th international conference, DEXA 2001, Munich, Germany, September 3-5, 2001 : proceedings	JF - Database and expert systems applications : 12th international conference, DEXA 2001, Munich, Germany, September 3-5, 2001 : proceedings
04-editor: Mayr, Heinrich C.;	ED - Mayr, Heinrich C.
12-place_published: Berlin	CY - Berlin
13-publisher: Springer	PB - Springer
14-date: 2001///	P1 - 2001///
20-extent: xix, 991	
21-series: Lecture notes in computer science	T3 - Lecture notes in computer science
22-series_volume: 2113	VL - 2113
33-keywords: Teorie dat; Umělá inteligence	KW - Teorie dat KW - Umělá inteligence
38-conference_name: DEXA 2001	T2 - DEXA 2001
39-venue: Mnichov, Německo	M2 - Mnichov, Německo ER -

Příklad exportu z interního záznamu do formátu BibTeX:

01-BOOK [YN]	@BOOK{
32-key: OpenGL_guide	OpenGL_guide,
00-author: Shreiner, Dave;+	author = {Shreiner, Dave and others},
02-title: OpenGL Programming Guide : The Official Guide to Learning OpenGL, Version 1.4	title = {OpenGL Programming Guide : The Official Guide to Learning OpenGL, Version 1.4},
06-organisation_name: OpenGL Architecture Review Board	
09-edition: 4th ed.	edition = {4},
12-place_published: Boston	address = {Boston},
13-publisher: Addison-Wesley	publisher = {Addison-Wesley},
14-date: c2004///	year = {2004},
20-extent: xliii, 759	
24-ISBN: 0321173481	isbn = {0-321-17348-1},
31-language: eng	}
33-keywords: OpenGL;computer graphics	

Příklad exportu z interního záznamu do formátu Endnote „Import File“ Format:

14-STANDARD [NN]	%0 Standard
27-identifier: ČSN ISO 690	%V ČSN ISO 690
02-title: Bibliografické citace - Obsah, forma a struktura	%T Bibliografické citace - Obsah, forma a struktura
12-place_published: Praha	%C Praha
13-publisher: Český normalizační institut	%I Český normalizační institut
14-date: 1996///	%D 1996
20-extent: 31	%P 31

6.5 Konverze

Příklad konverze z formátu Marc 21 do formátu EndNote "Import File" Format:

```
LDR -----nai-a22-----a-4500
FMT SE
001 web20081794679
003 CZ-PrNK
005 20080305092127.0
007 cr-mn-
```


008 080305c19999999xr--x-w-s-----0---b2cze--
 040 |a ABA001 |b cze
 0410 |a cze |a eng
 072 7 |a 004.7 |x Počítačové sítě |2 Konspekt |9 23
 072 9 |a 004.6 |x Interfacing and communications |2 Conspectus
 |9 23
 080 |a 004.738.52 |2 MRF
 080 |a 004.439:004.55HTML |2 MRF
 080 |a 004.439:004.55HTML |2 MRF
 080 |a (0.034.2)004.738.12 |2 MRF
 1001 |a Kosek, Jiří |7 xx0025270 |4 aut |4 pbl
 24510 |a Vše o WWW |h [elektronický zdroj] :
 |b vše, co jste kdy chtěli vědět, ale báli jste se zeptat
 24633 |a kosekcz
 260 |a [Praha : |b Jiří Kosek]
 310 |a Aktualizováno nepravidelně
 3621 |a Vychází od 1999?
 500 |a Název z titulní obrazovky (verze z 5.3.2008)
 520 |a Velice obsáhlý zdroj informací tvorbě WWW-stránek,
 včetně skriptů, programování webových aplikací apod
 538 |a Způsob přístupu: World Wide Web
 65007 |a World Wide Web |2 czenas
 65007 |a DHTML |2 czenas
 65007 |a kaskádové styly |2 czenas
 65009 |a World Wide Web |2 eczenas
 65009 |a DHTML |2 eczenas
 65009 |a cascading style sheets |2 eczenas
 655 7 |a www dokumenty |2 czenas
 655 9 |a www documents |2 eczenas
 85640 |u http://www.kosek.cz/ |q text/html |4 N
 929 |a Text
 930 |a cop.
 SYS 001794679

%0 Web Page
 %9 [elektronický zdroj]
 %A Kosek, Jiří
 %C [Praha]
 %G cze;eng
 %T Vše o WWW : vše, co jste kdy chtěli vědět, ale báli jste se zeptat
 %U http://www.kosek.cz
 %Z Název z titulní obrazovky (verze z 5.3.2008)

Příklad konverze z formátu BibTeX do formátu RIS:

@TECHREPORT{	TY - RPRT
report1999,	
author = "Jaroslav Král and	AU - Král, Jaroslav
Michal Zemlička",	AU - Žemlička, Michal
title = "Autonomous Components	TI - Autonomous Components and
and Wide Area Information Systems",	Wide Area Information Systems
institution = "Charles Univerity,	
Faculty of Mathematics and Physics,	
Department of Software Engineering",	
type = "Technical Report",	M1 - Technical Report
number = "10",	VL - 10
month = dec,	
year = 1999,	P1 - 1999//December
address = "Prague, Czech Republic",	
note = "http://kocour.ms.mff.cuni.cz	N1 - http://kocour.ms.mff.cuni.cz
/~zemlicka/ps/	/~zemlicka/ps/AC_and_WAIS.ps
AC_and_WAIS.ps"	
}	ER -

Kapitola 7

Generování bibliografických citací

V této kapitole popsáno, jak se vytvářejí citace. Celý problém je rozdělen na dvě části, vytvoření jednotlivých citací a sestavení seznamů.

Citace lze vytvářet pouze z interních záznamů. Je možné vytvořit citace ze všech záznamů v souboru nebo pomocí klíčů určit, které záznamy budou citovány. Vstupní soubor tedy obsahuje klíče a také názvy souborů, v nichž se záznamy s těmito klíči nachází. Soubor má stejný formát jako LaTeXem generovaný soubor .aux, aby uživatel mohl v LaTeXu používat příkaz `\cite` k vytvoření citace v textu a pak si jednoduše vygenerovat příslušný soupis použité literatury. Výstupním formátem může být soubor .bbl pro další zpracování LaTeXem nebo jednoduchý seznam citací v html nebo xml. Formát jednotlivých citací i formát celých seznamů se řídí nastavením v konfiguračních souborech.

7.1 Vytvoření citace z interního záznamu

Citace se vytvoří aplikací pravidel v konfiguračním souboru na interní záznam. Pro každý typ interního záznamu a kombinaci značek (publikovaný dokument, elektronický dokument) je možné definovat jiný soubor pravidel. Následující příklad ukazuje soubor pravidel pro vydanou knihu:

```
01-BOOK [Y-]
1(00-author Y -editAuthors "; |, et al. ").
2(02-title Y @E).
4(09-edition Y).
3{[N (04-editor N -editEditors "; |, et al.")(06-organisation_name
  N)].(05-subsidiary_responsibility N)}.
9(24-ISBN Y -labelFirst "ISBN ").
5{(12-place N):(13-publisher N),(14-date Y -toDate "YYYY")}.
6(20-extent N -labelAfter " s. ").
7{(21-series N):(22-series_volume N)}.
8{(23-note N) (28-URL N -labelFirst "Dostupné též na: ")}
```

7.1.1 Typy pravidel v konfiguračním souboru

Pravidla umožňují definovat seznamy položek včetně oddělovačů, formátovat položky i celé seznamy pomocí předdefinovaných funkcí a uzavírat je do závorek.

Každý soubor pravidel obsahuje jeden číslovaný seznam. Pořadí zpracování položky číslovaného seznamu je určeno pořadím zápisu v konfiguračním souboru. Čísla u položek pak udávají pořadí položky ve výstupní citaci. Každý údaj z interního záznamu smí být použit pouze jednou, což reflektuje fakt, že každý údaj o dokumentu se v citaci uvádí taktéž pouze na jednom místě. Převede se tedy podle prvního pravidla, které je pro něj nalezeno. To umožňuje určit priority při zaplňování polí v citaci.

Položkou číslovaného seznamu není přímo údaj v interním záznamu, ale vnořená položka. Vnořených položek je více typů, v konfiguračním souboru se rozlišují pomocí různých typů závorek. Některé opět obsahují vnořené položky. Hloubka zanoření je v podstatě neomezená, zpracování probíhá průchodem do hloubky.

Typy vnořených položek:

Jednoduchá položka Položka obsahuje číslo pole interního záznamu, což vzhledem k zanoření znamená konec rekurze. Při zpracování vrací položka obsah příslušného pole. Pokud je pole neprázdné, může být jeho obsah uzavřen do závorek a po té ještě upraven formátovací funkcí.

Každá jednoduchá položka má v konfiguračním souboru uvedenou značku, zda je v rámci citace povinná. Pokud je položka povinná a pole v interním záznamu prázdné, vypíše se upozornění, aby uživatel záznam doplnil. Pokud údaj neexistuje, je možné výpis upozornění potlačit uvedením znaku „#“ v příslušném poli interního záznamu.

Jednoduché položky se zapisují do kulatých závorek. Číslo 16 je v následujícím příkladu číslo pole v interním záznamu. Y znamená, že položka je povinná (v opačném případě uvedeno N). Za pomlčkou následuje název formátovací funkce a v uvozovkách je uveden argument funkce.

```
(16-access_date Y -labelFirst "cit. ")
```

Seznam položek Seznam položek je tvořen výčtem vnořených položek. Položky jsou zpracovány a spojeny oddělovači v pořadí, které odpovídá zápisu v konfiguračním souboru. Výsledný řetězec je možné ještě uzávorkovat a nakonec upravit formátovací funkcí.

Seznam položek se značí složenými závorkami. Následující příklad obsahuje seznam dvou jednoduchých položek. Oddělovačem položek je středník. Za oddělovače se automaticky doplňují mezery.

```
{(21-series N);(22-series_volume N)}
```

První platná položka ze seznamu Tento typ seznamu využívá vlastnosti, že každý údaj smí být použit pouze jedenkrát. Položky v seznamu jsou postupně zpracovávány a první neprázdná položka je výstupem celého seznamu. Výstupní řetězec může být ještě opatřen závorkami.

Tento typ položky se uvádí v hranatých závorkách. V následujícím příkladu písmeno Y značí, že výstup ze seznamu je povinný (v opačném případě uvedeno písmeno N). V uvozovkách je pak uvedena chybová hláška, jež se zobrazí, pokud se po zpracování celého seznamu žádný výstup neobjeví. Jako poslední je uveden výčet položek bez oddělovačů.

```
[Y "Chybi udaj o dostupnosti." (29-dostupnost N)(28-URL N)]
```

Jak již bylo zmíněno, údaje získané z interního záznamu lze uzavírat do závorek. V konfiguračním souboru se použití závorek označuje znakem „@“, za ním následuje další znak určující typ uzávorkování (zkratky vycházejí z anglických názvů). Kulaté závorky se zapíší pomocí znaku „P“, hranaté pomocí „S“ a složené pomocí „C“. Pokud je uveden znak „E“, znamená to, že se kolem textu vloží speciálních znaky pro začátek a konec zvýraznění textu. Tyto znaky jsou při sestavování seznamu citací nahrazeny podle nastavení v konfiguračním souboru pro tvorbu seznamů.

Příklad uzávorkování uvedeného v jednoduché položce:

```
(08-medium_designator Y @S)
```

Příklad uzávorkování uvedeného v seznamu položek:

```
{@S (16-access_date Y -labelFirst "cit. ")}
```

Kromě uzávorkování lze položky také zpracovávat formátovacími funkcemi. Výčet formátovacích funkcí pokrývá běžné úpravy potřebné ke zformátování citace.

Přehled formátovacích funkcí:

Přidání popisku k textu K jednotlivým polím je možné na začátek nebo na konec doplňovat text.

Úprava formátu data Úprava data je podobná jako při exportu, výstupem může být datum ve tvaru „YYYY“ nebo „YYYY-MM-DD“.

Formátování jmen Formátovat lze jak seznam jmen, tak jednotlivá jména. Formátování se řídí maskou.

V masce je stanoveno, kolik jmen je možné uvést a čím se oddělí. Oddělovač mezi posledními dvěma jmény lze nastavit odlišný. Je tam také uvedeno,

kolik jmen smí obsahovat neúplný seznam, a značka, jež vyjadřuje neúplnost seznamu.

Maska obsahuje i tvar jednotlivých jmen. U křestního jména je možné volit mezi plnými křestním jménem nebo iniciálami. Příjmení lze psát velkými písmeny.

Pokud se formátují jména editorů, přidává se nakonec seznamu jmen zkratka „(ed.)“. U zkratky je možné zvolit velikost písmene e a také, jestli se při uvádění více jmen má použít množné číslo.

Ukázka formátování jména editora. Nejprve je uvedena položka interního záznamu, pak maska a nakonec výsledné jméno:

```
Mayr, Heinrich C.;+
3; |; |1, et al.|L, |f ,|sn
MAYR, Heinrich C. (ed.), et al.
```

Další podrobnosti ohledně formátovacích funkcí, včetně příkladů, jsou uvedeny v uživatelské dokumentaci.

Konfigurační soubory umožňují vytvářet poměrně složitá pravidla, což se pozitivně projeví na kvalitě výstupních citací. Přestože pravidla neumožňují přímo testovat existenci obsahu položek v interním záznamu, určení pořadí jednotlivých částí citace a zápis do položky podle nastavených priorit dosahuje shodné vyjadřovací síly při rozhodování, kam se má položka v rámci citace umístit.

Pokud je ale nějaký údaj v záznamu uveden na dvou místech, například organizace jako autor a vydavatel, pak se tato chyba objeví i v citaci.

7.1.2 Příklady citací

Příklad citace standardu, nejprve uveden interní záznam, pak pravidla a nakonec výsledná citace:

```
14-STANDARD [NN]
27-identifier: ČSN ISO 690.
02-title: Bibliografické citace - Obsah, forma a struktura
09-edition: #
12-place_published: Praha
13-publisher: Český normalizační institut
14-date: 1996///
20-extent: 31
32-key: CSNISO690
34-label: ČSN ISO 690:1996
```

```

14-STANDARD [--]
1(27-identifier Y).
2(02-title Y @I).
5{(12-place_published N):[N (13-publisher N)(06-organisation_name
N)],(14-date Y -toDate "YYYY")}.
3{[N (04-editor N -editEditors "; |, et al.")(06-organisation_nam
e N)].(05-subsidiary_responsibility N)}.
4(09-edition Y).
6(20-extent N -labelAfter " s. ").
7{(23-note N) (28-URL N -labelFirst "Dostupné též na: ")}.

```

```

\bibitem{CSNIS0690}\v{C}SN ISO 690. \emph{Bibliografick\'{e} cita
ce - Obsah, forma a struktura}. Praha: \v{C}esk\'{y} normaliza\v{c}
n\{i} institut, 1996. 31 s.

```

7.2 Tvorba seznamů citací

Vytváření seznamů citací probíhá tak, že se nejprve načte ze vstupního souboru seznam klíčů citovaných záznamů a názvy souborů s příslušnými interními záznamy (dále jako zdroje).

Pak se procházejí jednotlivé zdroje a načítají se interní záznamy. Pokud je klíč záznamu nalezen v seznamu klíčů nebo se mají citovat všechny záznamy, vytvoří se ze záznamu citace. Výsledná citace obsahuje klíč, popisek a vlastní text citace. Způsob vytvoření textu citace je popsán výše. V případě citování všech záznamů se projdou všechny zdroje, jinak se prohledávání zdrojů ukončí, jakmile jsou nalezeny záznamy ke všem klíčům v seznamu.

Pokud se v citacích vyskytují znaky shodné s metaznaky výsledného formátu, jsou nahrazeny odpovídajícími sekvencemi. U typu .bbl se provede i nahrazení národních znaků, tj. všech kromě ASCII. Ostatní výstupní soubory zůstávají ve znakové sadě UTF-8.

Nakonec se podle nastavení v konfiguračním souboru citace setřídí podle pořadí klíčů ve vstupním souboru a doplněním značek pro označení položek, začátku a konce souboru, uvedených v konfiguračním souboru vytvoří výsledný seznam.

7.2.1 Formát vstupního souboru

Vstupní soubor má stejný formát jako soubor .aux generovaný LaTeXem.

Klíč se zapisuje příkazem `\citation{klíč}`. Názvy souborů s interními záznamy (dále zdroje) příkazem `\bibdata{soubor1, soubor2...}`. Pokud by chtěl uživatel citovat všechny záznamy ve zdrojích, uvede místo klíče v příkazu hvězdičku, `\citation{*}`.

Informace o tom, ve které položce interního záznamu je klíč uložen, se nachází v konfiguračním souboru „editFile.cfg“, tato položka má na místě načítací funkce písmeno K.

Příklad vstupního souboru:

```
\citation{openGL_guide}  
\citation{warren_weimer}  
\citation{CSNISO6902}  
\citation{mayr}  
\bibdata{./tests/sample.txt}
```

7.2.2 Formát konfiguračního souboru pro tvorbu seznamů

Konfigurační soubor obsahuje značky pro začátek a konec seznamu a také značky pro začátek a konec položky.

Obecný tvar konfiguračního souboru:

```
text před seznamem  
%  
značka začátku seznamu  
značka začátku položky  
značka začátku zvýraznění textu  
značka konce zvýraznění textu  
značka konce položky  
značka konce seznamu  
text za seznamem
```

Text před seznamem může být víceřádkový, jeho konec je označen řádkem s procentem na začátku. Všechny značky jsou jednořádkové a jsou povinné, nicméně každá značka může být prázdná, což se zapíše jako prázdný řádek. Text za seznamem je volitelný.

Značka začátku seznamu může obsahovat znak plus. Tento symbol se při výpisu do cílového souboru nahradí tolika znaky „x“, kolik znaků má číslo poslední položky. To je užitečné především při citování v LaTeXu, kde je nutné předat do seznamu maximální šířku popisku položky.

Značka začátku položky může obsahovat dva symboly, „@“ a „*“. První se nahradí popiskem položky a druhý klíčem položky. Značka začátku položky může každý z těchto znaků obsahovat nejvýše jedenkrát, což ovšem postačí k tvorbě různých odkazů.

Příklad konfiguračního souboru a seznamu, který vznikl podle pravidel, v něm uvedených.

```
%  
\begin{thebibliography}{+}  
\bibitem{*}  
  
\emph{  
}  
\end{thebibliography}
```

Odsazené řádky jsou pokračováním řádku předchozího, text je zalomen z důvodu omezené šířky stránky.

```
\begin{thebibliography}{x}  
\bibitem{openGL_guide}Shreiner, Dave, et al. \emph{OpenGL Programming  
Guide : The Official Guide to Learning OpenGL, Version 1.4}.  
OpenGL Architecture Review Board. 4th ed. Boston: Addison  
-Wesley, c2004. xliii, 759 s. ISBN 0-321-17348-1.  
\bibitem{CSNISO690}\v{C}SN ISO 690. \emph{Bibliografick\'{e} citace  
- Obsah, forma a struktura}. Praha: \v{C}esk\'{y} normaliza  
\v{c}n\'{\i} institut, 1996. 31 s.  
\end{thebibliography}
```

Kapitola 8

Závěr

Aplikace je schopna při konverzi poměrně dobře pracovat se záznamy, které se skládají z elementárních, vzájemně nezávislých položek. Velkou výhodou je možnost přidávat nové položky i celé typy vnějších záznamů pouhou změnou pravidel v konfiguračním souboru. Práce se záznamy, jejichž pole jsou provázaná, je omezena vnitřní implementací. Přidávat nové formáty možné není.

Množství ukládaných údajů a též kvalita konverze se odvíjí od implementace interního formátu. Obsah položek a jejich zpracování jsou nastavitelné pomocí konfiguračního souboru, což umožňuje uživateli ukládat údaje dle jeho potřeby. Jemnost rozlišení typů interního záznamu je ale omezena maximálním počtem typů záznamů. Stejně tak je omezen maximální počet položek a typ jejich obsahu (jména, datum, klíč, popisek, obyčejná položka).

Z výše jmenovaných omezení se nabízí několik možností pro budoucí rozšíření aplikace, například v rámci diplomové práce nebo softwarového projektu. Mírně lze kvalitu na úkor rychlosti zvýšit i tím, že se místo záznamů budou při konverzi procházet pravidla, čímž se umožnilo spojování položek v určitém pořadí.

Mnohem výrazněji se však kvalita zvýší, pokud dovolíme, aby pravidla pro konverzi v konfiguračních souborech mohla obsahovat podmínky na existenci a obsah položek. Dalším zajímavým rozšířením by byla opakovatelnost polí u záznamů vnitřního formátu, resp. opakovatelnost sekvence polí jako je u formátu MARC 21.

Část pro tvorbu citací je narozdíl od konverzí poměrně flexibilní. Konfigurační soubory umožňují téměř jakýkoli formát seznamů a možnosti pravidel pro tvorbu citací dovolují přidávat nové citační styly napsáním příslušného konfiguračního souboru. Na druhou stranu nelze přidávat vlastní formátovací funkce, jejich výčet je omezen implementací, takže v tomto směru také by bylo možné aplikaci ještě rozšířit. Uživatelé by také jistě ocenili možnost vytváření vnořených seznamů citací.

Pokud jde o přívětivost k uživateli, obsluha aplikace je jednoduchá, ale o to obtížnější je správa konfiguračních souborů. Změny v nich je třeba provádět citlivě s ohledem na formáty, pravidla citačních stylů a hlavně s ohledem na vzájemnou konzistenci konfiguračních souborů.

Možnosti rozšíření ve směru uživatelské přívětivosti aplikace jsou velké. Asi nejzajímavější by byla implementace vyhledávání v souborech se záznamy a možnost vyhledávání duplicit. S tím by pravděpodobně souviselo ovládání aplikace pomocí okenního aplikačního rozhraní, na což by se dalo navázat vytvořením plug-in modulu pro textový editor.

Literatura

- [1] ČSN ISO 690. *Bibliografické citace – Obsah, forma a struktura*. Praha : Český normalizační institut, 1996. 31 s.
- [2] ČSN ISO 690-2. *Informace a dokumentace – Bibliografické citace – Část 2 : Elektronické dokumenty nebo jejich části*. Praha : Český normalizační institut, 2000. 22 s.
- [3] ISO 690:1987. *Information and documentation – Bibliographic references - Content, form and structure*. 2nd ed. ISO, 1987. Vybrané části dostupné na: <<http://www.collectionscanada.ca/iso/tc46sc9/standard/690-1e.htm>>.
- [4] ISO 690-2. *Information and documentation – Bibliographic references – Part 2 : Electronic documents or parts thereof*. ISO, 1997. Vybrané části dostupné na: <<http://www.collectionscanada.ca/iso/tc46sc9/standard/690-2e.htm>>.
- [5] LAMPORT, Leslie. *Latex : A Document Preparation System : User's Guide and Reference Manual*. Illustrations by Duane Bibby. 2nd ed. Boston : Addison-Wesley, c1994. xiii, 272 s. ISBN 0-201-52983-1.
- [6] LAMPORT, Leslie. *Latex : A Document Preparation System : User's Guide and Reference Manual*. Illustrations by Duane Bibby. 2nd ed. Boston : Addison-Wesley, c1994. Appendix B, The Bibliography Database, s. 155–164. ISBN 0-201-52983-1.
- [7] *Reference Manager : RIS Format Specifications* [online]. The Thomson Corporation, c2009, last modified on February 14, 2001 [cit. 2009-07-20]. Dostupné na: <http://www.refman.com/support/risformat_intro.asp>.
- [8] *Reference Manager 11* [online]. Version 11, for Windows. The Thomson Corporation, c2004-2005. Appendix C, Table of Default Reference Types and Fields, s. 523–534. Dostupné na: <www.refman.com/support/docs/ReferenceManager11.pdf>.

- [9] *Endnote X2: User's Guide. Windows Version X2* [online]. The Thomson Corporation, c2008. Dostupné na: <<http://www.endnote.com/support/helpdocs/EndNoteX2WinManual.pdf>>
- [10] Library of Congress. Network Development and MARC Standards Office. *MARC 21 Format for Bibliographic Data* [online]. 1999 ed., Update No.1(October 2001) through Update No.9(October 2008). 2008-12-03 [cit. 2009-07-20]. Dostupné na: <<http://www.loc.gov/marc/bibliographic/ecbdhome.html>>.
- [11] Library of Congress. Network Development and MARC Standards Office. *MARC 21 Specifications for Record Structure, Character Sets, and Exchange Media : Record Structure* [online]. January 2000, last update 2007-05-17 [cit. 2009-07-20]. Dostupné na: <<http://www.loc.gov/marc/specifications/specrecstruc.html>>.

Příloha A

Obsah přiloženého CD

- Elektronická verze bakalářské práce (pdf)
- Program BibTools
 - Uživatelská dokumentace (pdf)
 - Soubory ke spuštění
 - Konfigurační soubory
 - Ukázky záznamů
 - Programátorská dokumentace (pdf)
 - Zdrojový kód aplikace
- Návod na zpracování bibliografických citací dle normy ČSN ISO 690 a ČSN ISO 690-2 (pdf)