

PŘÍRODOVĚDECKÁ FAKULTA UNIVERZITY PALACKÉHO
KATEDRA INFORMATIKY

BAKALÁŘSKÁ PRÁCE

Konvertor datových formátů ve FCA
a veřejných repozitářích



2011

Jan Strnad

Anotace

Hlavním cílem této práce bylo vytvořit aplikaci s konzolovým i grafickým uživatelským rozhraním pro převod mezi různými datovými formáty vstupních dat ve formální konceptuální analýze a transformaci datových souborů z veřejných online databází.

Děkuji Mgr. Janovi Outratovi, Ph.D. bez kterého by tato práce nemohla vzniknout.

Obsah

1. Úvod	7
2. Formální konceptuální analýza	7
2.1. Typy datových souborů	8
2.1.1. Comma Separated Values	8
2.1.2. Attribute-Relation File Format	9
2.1.3. Burmeister Formal Context	10
2.1.4. Frequent Itemset Mining Implementations	11
2.1.5. The three column CSV	11
3. Vývojová platforma	12
3.1. Microsoft .NET Framework	13
3.1.1. Historie	13
3.1.2. Významné vlastnosti	14
3.2. Mono	15
3.2.1. Historie	15
3.2.2. Hlavní rysy projektu Mono	16
3.2.3. Mono na Windows	16
3.3. Vlastní realizace	16
3.4. USE CASE	17
4. Flereo User's guide	18
4.1. Prerequisites	18
4.2. Installation	18
4.3. Parameters	18
5. WinFlereo User's guide	20
5.1. Prerequisites	20
5.2. Installation	20
5.3. Main Application	20
5.3.1. Search for values	21
Závěr	22
Conclusions	23
Reference	24
A. Obsah přiloženého CD	25

Seznam obrázků

1.	Struktura .NET Frameworku	13
2.	Struktura .NET Frameworku	14
3.	Metody knihovny FceHelper	17
4.	Class diagram knihovny FceHelper	27
5.	Use Case WinFlereo	28
6.	Main screen	28
7.	Menu Option	28
8.	New Data Dialog	29
9.	Item Detail	30
10.	File Load 1	31
11.	Save data dialog	31
12.	Save Dialog	32
13.	Save Csv Dialog	32

Seznam tabulek

1.	Příklad vstupních dat	7
2.	Příklad výstupních dat 1	8
3.	Příklad výstupních dat 2.	8

1. Úvod

Data a datové soubory nás obklopují doslova na každém kroku. Každý výrobce software ukládá data tak, aby byly pro jeho software co nejefektivněji dosažitelné. Tímto vzniká velké množství druhů datových souborů. Některé druhy si jsou navzájem velmi podobné a lze říci že jsou nadstavbou již existujících a některé přichází se zcela originálním způsobem organizace dat v souboru. Používá-li uživatel dvě podobné aplikace, tak velmi brzy vznikne potřeba sdílet data mezi těmito aplikace a následně narazí na problém s nekompatibilitou datových souborů. Řešením pak bývá používat další aplikace pro převod takových dat. Takovou aplikace zde budeme popisovat. Budeme se specializovat na převod a zpracování textových souborů, které se používají ve formální konceptuální analýze (FCA). Nyní si stručně popíšeme co je FCA a představíme si možné vstupní a výstupní soubory.

2. Formální konceptuální analýza

Rozhlédnete-li se okolo sebe, uvidíte spousty předmětů – objektů. Pokud tyto objekty budeme chtít popsat začneme popisovat jejich vlastnosti. O autě můžeme říct, že jede a že je červené. Broskev je žlutá a sladká. Pokud ale řekneme: „Je to žluté a sladké, co to je?“ Takse nám určitě broskev nevybaví. Potřebuje popis rozšířit – popsat další vlastnosti. Při výčtu vlastností zjišťujeme, že by se nám hodilo vyjádřit, že nějakou vlastnost daný objekt nemá. A nebo, že ji má jen částečně. Pokud taková data zapíšeme do tabulky máme vstup pro formální konceptuální analýzu, která se zabývá zpracováním tabelárních dat a poskytuje netriviální pohled na data a je schopna zobrazit souvislosti, které nejsou na první pohled v datech vidět.

Mějme například jednoduchou tabulku s názvem obce a dalšími statistickými informacemi.

Název obce	Počet obyvatel	Hlavní město	Průměrný věk	Statut
Praha	1288696	Ano	41,5	Město
Brno	404345	Ne	42,2	Město
Jihlava	51234	Ne	40,3	Město
Herálec	1087	Ne	38,5	Městys
Kozlany	125	Ne	38,8	Obec

Tabulka 1. Příklad vstupních dat

Výsledek zpracování může vypadat jako v tabulce 2. nebo v tabulce 3.

.	.	.	X
.	X	.	X
X	X	.	X
X	X	X	X
X	X	X	.

Tabulka 2. Příklad výstupních dat 1

4			
2	4		
1	2	4	
1	2	3	4
1	2	3	

Tabulka 3. Příklad výstupních dat 2.

Aby bylo možné datové soubory jednoduše zpracovávat byly definovány datové soubory pro výměnu dat. Nyní si zde několik vybraných popíšeme.

2.1. Typy datových souborů

V tomto textu si popíšeme 5 datových typů, které program WinFlereo a Flereo podporuje. Jsou to CSV, 3COL, ARF, CXT a DAT. Ve všech případech se jedná o ASCII textové soubory, kde jsou jednotlivé řádky odděleny znaky pro konec řádku.

2.1.1. Comma Separated Values

Comma Separated Values zkráceně CSV je definován standardem RFC 4180. V tomto formátu jsou dle normy jednotlivé sloupce odděleny znakem „,”. Je-li potřeba zapsat do dat znak „,” musí být řetězec uzavřen v uvozovkách. Pro zápis znaku „” je třeba uvozovky zdvojit ””. Jistou komplikací je implementace CSV souboru v aplikaci Excel od firmy Microsoft, která používá v závislosti na regionálním nastavení znak „,” jako oddělovač v desetinných číslech. Pro oddělení sloupců pak používá znak „;“. Protože je tento formát pro svou jednoduchost velmi rozšířen, lze se často setkat i s jinými oddělovači.

Year,Make,Model,Description,Price


```

1997,Ford,E350,"ac, abs, moon",3000.00
1999,Chevy,"Venture ""Extended Edition""",",",4900.00
1999,Chevy,"Venture ""Extended Edition, Very Large""",",",5000.00
1996,Jeep,Grand Cherokee,"MUST SELL!
air, moon roof, loaded",4799.00

```

2.1.2. Attribute-Relation File Format

Formát Attribute-Relation File Format (ARF) vznikl v roce 2002 jako součást projektu Machine Learning Project na fakultě informatiky University of Waikato. Soubor ARF povinně obsahuje 2 části: záhlaví a data.

Záhlaví obsahuje jméno relace, dále jména jednotlivých atributů a jejich datové typy.

```

% 1. Title: Iris Plants Database
%
%
% 2. Sources:
%      (a) Creator: R.A. Fisher
%      (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
%      (c) Date: July, 1988
%
@RELATION iris

@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class       {Iris-setosa,Iris-versicolor,Iris-virginica}

```

Sekce data může vypadat třeba takto:

```

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa

```

Rádek začínající znakem "%" je komentář. @RELATION, @ATTRIBUTE a @DATA jsou klíčová slova a nerozlišují se u nich malá a velká písmena. Definice atributu má následující syntaxi: @attribute <název-atributu> <datový-typ>

Datové typy:

numeric - celé nebo desetinné číslo

<výčet> - hodnoty oddělené čárkou uzavřené ve {}

String - libovolný řetězec. Pokud obsahuje mezeru měl by být uzavřený v uvozovkách

Date - datum ve formátu dle ISO ISO-8601 tj. "yyyy-MM-dd'T'HH:mm:ss".

Název atributu musí začínat písmenem a nesmí obsahovat mezery. Formát ARF má i další rozšíření např. Sparse ARFF. Tyto rozšíření nejsou implementovány.

2.1.3. Burmeister Formal Context

Burmeister Formal Context (CXT) je oblíbený a rozšířený formát ve FCA. CXT soubor se skládá ze dvou částí: záhlaví a data. Záhlaví začíná písmenem "B", pak následuje počet objektů a počet atributů vždy na samostatném řádku. Následuje výpis objektů a atributů. Sekce data obsahuje tabulku obsahující pouze znaky "X" nebo ".". Znak "X" vyjadřuje pravdivost výroku a znak "." nepravdivost. Následující příklad popisuje 8 objektů a jejich 15 atributů.

```
B
8
15

Sam
Jenny
John
Andrew
Mary
Laura
Alice
Tim
education-Bachelors
education-HS-grad
education-11th
education-Masters
```

```

education-?
employment-Clerical
employment-Managerial
employment-Unskilled
employment-Professional
sex-Male
sex-Female
US-citizen-Yes
US-citizen-No
class-<=50K
class->50K
X...X...X.X.X.
X....X...XX.X.
.X....X.X.X.X.
..X...X.X.X.X.
X.....X.XX..X
...X..X...X.XX.
....XX....X.XX.
.X...X..X.X..X

```

2.1.4. Frequent Itemset Mining Implementations

Frequent Itemset Mining Implementations (DAT) je datový formát, který se používá v Data Miningu a při testování efektivity algoritmů. Tento formát neobsahuje názvy objektů ani atributů. Každý řádek obsahuje data pro jeden vstupní objekt oddělené mezerou. Data jsou celé číslo reprezentující pořadí atributu při vyhodnocení začínající od 1.

```

2 5 9 13 15 17
4 5 10 14 15 17
2 8 12 13 15 17
4 7 12 13 15 17
1 5 11 14 15 16
2 6 10 14 17
3 9 14 17
4 8 10 13 15 16

```

2.1.5. The three column CSV

Jak z názvu vyplývá, jedná se o tři sloupcový CSV soubor (3COL). Každý řádek reprezentuje trojici objekt, atribut a hodnota oddělenou znakem ”,”. Tento formát se hojně používá pro Resource Description Format (RDF).

Sam, age, 39

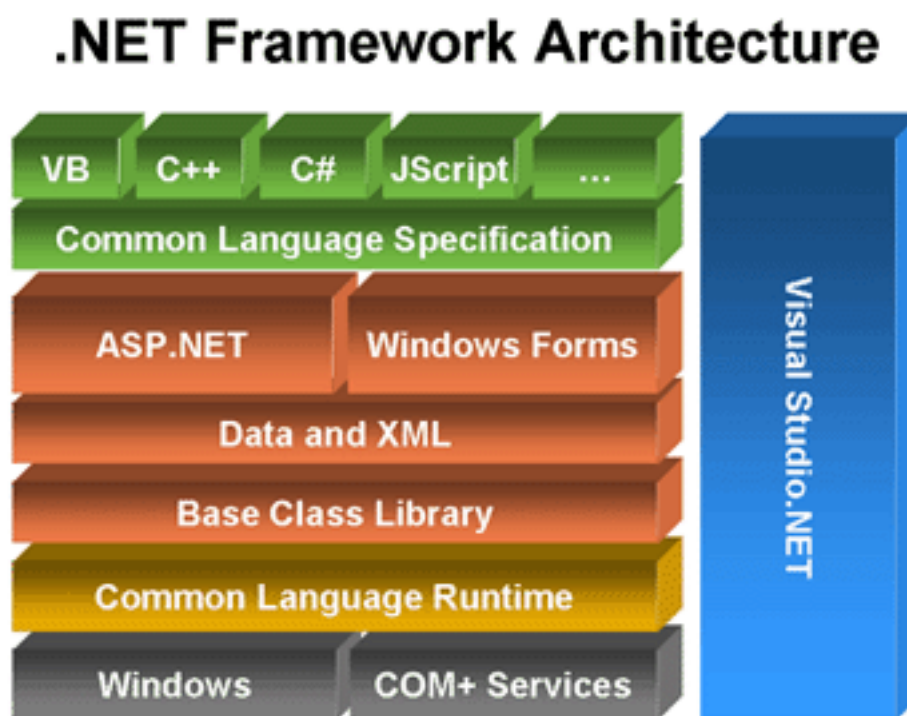
Jenny, age, 50
John, age, 38
Andrew, age, 53
Mary, age, 28
Laura, age, 37
Alice, age, 49
Tim, age, 52
Sam, employment, Clerical
Jenny, employment, Managerial
John, employment, Unskilled
Andrew, employment, Unskilled
Mary, employment, Professional
Laura, employment, Managerial
Alice, employment, Clerical
Tim, employment, Managerial
Sam, sex, Male
Jenny, sex, Female
John, sex, Male
Andrew, sex, Male
Mary, sex, Female
Laura, sex, Female
Alice, sex, Female
Tim, sex, Male
Sam, class, <=50K
Jenny, class, <=50K
John, class, <=50K
Andrew, class, <=50K
Mary, class, >50K
Laura, class, <=50K
Alice, class, <=50K
Tim, class, >50K

3. Vývojová platforma

Na začátku vývoje každého software je třeba dobře promyslet ve kterém vývojovém prostředí budeme vyvíjet. Pozdější změna v podstatě znamená kompletní přepracování zdrojového kódu pro jinou platformu. V dnešní době naštěstí existuje jedno téměř univerzální multiplatformní prostředí. Je to projekt Mono a .NET Framework (.NET).

3.1. Microsoft .NET Framework

Microsoft .NET Framework (.NET) je softwarový balík od firmy Microsoft, který obsahuje nejen velké množství hotových knihoven, ale i několik programovacích jazyků, které umožňují jazykovou interoperabilitu. To znamená, že můžeme výsledný kód z jednoho jazyku použít v jiných jazycích. Aby toto fungovalo není při kompilaci program překládám do strojového kódu, ale do speciálního jazyka MSIL (Microsoft Intermediate Language), který je interpretován prostřednictvím CLR (Common Language Runtime). CLR je v podstatě virtuální stroj, který se mimo jiné stará o bezpečnost, správu paměti a zpracování výjimek. Nad CLR je vystavěna základní knihovna Base Class Library, nad kterou jsou vystavěny ostatní knihovny pro přístup k datům, práce s XML, šifrování, knihovny pro tvorbu uživatelského rozhraní a další.

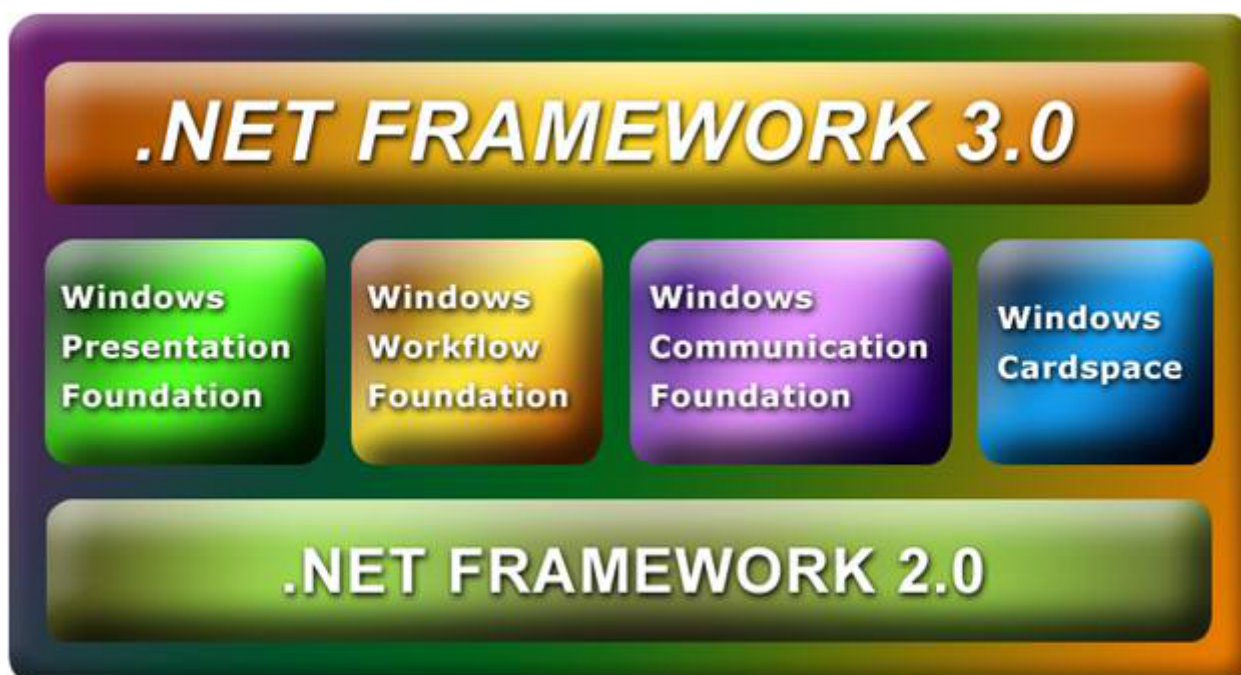


Obrázek 1. Struktura .NET Frameworku

3.1.1. Historie

Firma Microsoft začala na .NET pracovat v devadesátých letech pod názvem Next Generation Windows Services. Kolem roku 2000 byla zveřejněna první beta verze. O dva roky později přichází první oficiální verze .NET 1.0. Microsoft současně uvedl novou verzi vývojového prostředí Visual Studio .NET. V dubnu roku 2003 byla uvedena verze 1.1. Tato verze jednak přinesla přepracování vývojového

prostředí (Visual Studio 2003), opravu mnoha chyb a velké rozšíření knihoven. Poprvé je .NET implementován jako součást operačního systému Windows 2003 server. Masovému rozvoji už nestálo nic v cestě. Další významný milník přichází v dubnu 2005. Microsoft uvádí verzi 2.0 a Visual Studio 2005. Celé prostředí je přepracované a v jedné verzi Visual Studia, lze kompilovat kód pro různé verze .NET - multitargeting. Tato novinka spočívá v tom, že verze 3, 3.5 a 4 jsou pouze nadstavby nad jádrem verze 2.0.



Obrázek 2. Struktura .NET Frameworku

3.1.2. Významné vlastnosti

Typová bezpečnost Program má přístup pouze k paměti, která je alokována objekty programu. Toto platí za předpokladu, že je využíván přístup přes definované rozhraní. Při kompilaci probíhají verifikační procesy, které ověří jestli je kód typově bezpečný. V takovém případě je zajištěna úplná izolace assembly a její běh je zcela bezpečně.

Management paměti Pro správu paměti obsahuje .NET nástroj Garbage Collection, který kontroluje paměť a pokud v ní najde objekty, které už nebudou v budoucnu použity, tak tuto paměť uvolní a zároveň vrátí všechny zdroje, které byly použity těmito objekty. Prakticky to znamená, že se programátor nemusí starat o uvolňování paměti. Nutno podotknout, že by tento nástroj funguje

velmi dobře, je vhodné, aby programátor paměť uvolňoval přímo v kódu. V C# je od verze 2.0 k tomuto účelu speciální sekce `using`.

Bezpečnost Další silným rysem .NET je bezpečnost. NET pokrývá pod tímto pojmem několik oblastí. V kódu lze např. pomocí atributů určit, jaké systémové prostředky bude aplikace vyžadovat. Dále lze používat bezpečnostní role již při tvorbě programu. Lze tak efektivně vymezit, která funkcionalita bude přístupná které roli. Je k dispozici několik možností ověření uživatele a kryptografická služba, která pokrývá většinu standardů pro hashování, krytování a digitální podpisy. Tyto prostředky .NET i sám aktivně využívá např. v Global Assembly Cache (GAC) kam se ukládají pouze digitálně podepsané assembly.

3.2. Mono

Projekt Mono zahrnuje několik open source nástrojů pro vývoj a provoz aplikací založených na .NET platformě. Hlavní myšlenkou je umožnit vývojářům rychle a jednoduše vyvíjet multiplatformní aplikace. Mono je dnes dostupné pro systémy Unix, Linux, OS X, BSD a Windows. V Mono lze provozovat krom desktopových aplikací i aplikace serverové, protože pokrývá kompletně platformu ASP.NET. Mono je postaveno na standardech European Computer Manufacturers Association (ECMA) pro C# a CLR. Konkrétně jsou to standard ECMA 334, který specifikuje jazyk C# - syntaxi, pravidla pro interpretaci, a omezující vlastnosti a standard ECMA 335 Common Language Infrastructure (CLI), což je mezinárodní standard definující základní jazykovou infrastrukturu pro aplikace napsané v několika násobně vyšších jazycích, které mohou být spuštěny v různých systémových prostředích bez potřeby jakkoliv tyto aplikace upravovat.

3.2.1. Historie

Původním autorem projektu Mono je Miguel de Icaza, který se mimo jiné podílel na vývoji známého souborového manageru Midnight Commander, linuxového jádra a projektu GNOME. Sám je též autorem tabulkového procesoru Gnumeric. V roce 2001 přejmenoval svou firmu na Ximian a zahájil vývoj projektu Mono. Nejdříve zkoumal interpret bitového kódu a zjistil, že nikde nejsou dostupné informace o specifikaci metadat. V únoru roku 2001 se poptával po chybějících informacích a zhruba ve stejné době začal pracovat na kompilátoru jazyka C#, který psal rovněž v C#. V dubnu 2001 společnost ECMA publikovala chybějící informace. Po zveřejnění informací předvedl de Icaza na konferenci GU-ADEC svůj kompilátor, který mimo jiné uměl parsovat sám sebe. Projekt Mono byl oficiálně oznámen na konferenci O'Reilly 19. června 2001. V roce 2003 kupuje firmu Ximian firma Novell a De Icaza je v současné době viceprezident vývojářské sekce v Novellu.

3.2.2. Hlavní rysy projektu Mono

Nejdůležitější vlastnost již byla uvedena, je to podpora standardů ECMA. Ta zaručuje, že můžeme vzít libovolný program zkompilovaný jakýmkoliv jiným kompilátorem podporující tento standard a budeme mít jistotu, že náš program bude fungovat. Další významnou vlastností je plná podpora ASP.NET a WinForms aplikací. Mono zahrnuje krom runtime prostředí i vlastní kompilátor a vývojové prostředí MonoDevelop. K dispozici je i komerční nadstavba na Visual Studio.

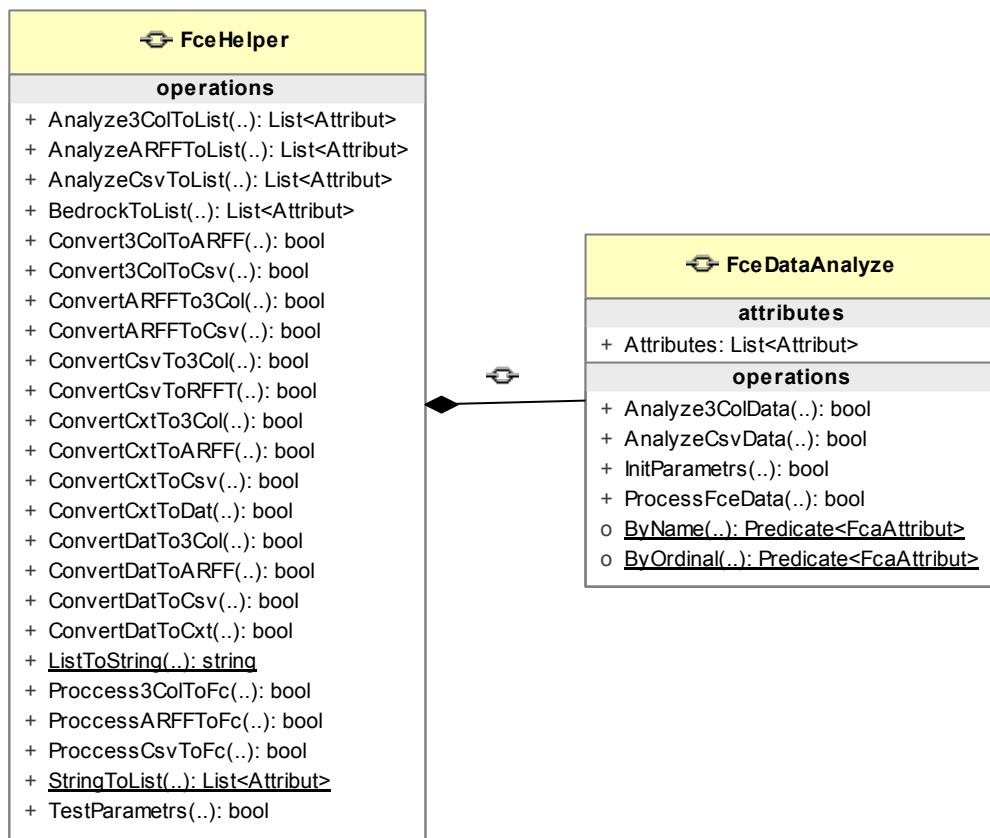
3.2.3. Mono na Windows

Zajímavostí je i existence projektu Mono pro platformu Windows, která je pokryta .NET Frameworkem od firmy Microsoft. V současné době si tam může uživatel vybrat zda-li jeho kód poběží v prostředí Mono či .NET. Vývojář si navíc může vybrat, který kompilátor využije a i které vývojové prostředí. Pokud máme zdrojový soubor pro konzolovou aplikaci v C#, tak ho můžeme zkompileovat buď příkazem `csc /out:program_net.exe zdroj.cs` (použije .NET C# kompilátor) a nebo `mcs {out:program_mono.exe zdroj.cs` (použije Mono C# kompilátor). Výsledkem jsou 2 spustitelné soubory stejné velikosti. Jednotlivé byty souborů jsou ale různé, přesto program pracuje stejně. V grafických aplikacích je situace horší. Mono neimplementuje CLI úplně. Grafické .NET aplikace využívají jmenný prostor `System.Windows.Forms` a ten obsahuje sadu prvků, které jsou navrženy speciálně pro Windows. Tento problém se dá řešit použitím knihovny `Gtk` či `gtksharp`. Knihovna `Gtk` je přenositelná a obsahuje téměř totožné komponenty a události jako knihovna `Forms` pro .NET a lze ji používat i v .NET. Druhou variantou je použít knihovnu `Windows.Forms`, což je Mono implementace původní knihovny `System.Windows.Forms` pro .NET, která je též multiplatformní. Projekt Mono nyní plně podporuje .NET verze 4.0 a WinForm aplikace verze 2.0.

3.3. Vlastní realizace

Na základě výše popsaných vlastností jednotlivých prostředí jsem zvolil pro vývoj Visual Studio 2010 a použil jsem .NET verze 3.5, která oproti verzi 2.0 obsahuje velmi užitečný objekt `HashSet`. Jako vývojovou platformu jsem zvolil též MS Windows a finální aplikaci jsem testoval v OpenSuse 11 s Mono 2.2. Dle předpokladů funguje na obou systémech nejen konzolová verze, ale i grafická verze. Dle zadání měla být vytvořena aplikace, která půjde používat z příkazové řádky a druhá, která obsahuje GUI.

Díky takto nastaveným požadavkům jsem vyrobil knihovnu `FceHelper.dll`, která zapouzdřuje veškerou funkcionalitu. Grafická verze (`WinFlereo`) a konzolová verze (`Flereo`) tedy pouze volají funkce z této knihovny. Class diagram je na obrázku 3. a na obrázku 4.



Obrázek 3. Metody knihovny FceHelper

3.4. USE CASE

Use Case k programu WinbFlereo je na obrázku 5.

Scénáře:

Načti data Uživatel vybere v menu položku „Načti data“. Objeví se standardní dialog pro výběr vstupního souboru. Pokud uživatel soubor vybere pokračuje se na další krok, pokud ne akce končí. V dalším kroku se uživateli zobrazí náhled na vstupní soubor a je požádán o kontrolu automaticky zjištěného typu souboru. Pokud se jedná o typ CSV pokračuje uživatel na další krok, kde je požádán o kontrolu automaticky zjištěného separátoru csv a má možnost nastavit zpracování prvního řádku souboru. Uživatel je vrácen na formulář ze kterého přišel. Pokud vybral typ souboru ARF, CSV, 3COL je uživatel dotázán jestli chce pokračovat viz. scénář „Otevřít data“. V ostatních případech je dotázán jestli chce pokračovat viz scénář „Ulož data“.

Otevři data Okno obsahuje tlačítka pro Načtení konfigurace ze souboru, Ulo-

žení konfigurace do souboru, Přidání definice atributu. Tlačítko Odebrání definice atributu je aktivní pouze v případě, že formulář obsahuje více jak jednu definici atributu. Je-li vybrán vstupní soubor typu ARF, CSV, 3COL je uživatel dotázán, na automatickou analýzu vstupního souboru. Uživatel je vrácen na formulář ze kterého přišel.

Načtení konfigurace ze soubor Objeví se standardní dialog pro výběr vstupního souboru. Pokud uživatel soubor vybere, předvyplní se data, dle načtené konfigurace. Uživatel zůstává na okně, ze kterého vyšel.

Uložení konfigurace ze soubor Objeví se standardní dialog pro výběr výstupního souboru. Data jsou uložena a uživatel zůstává na okně, ze kterého vyšel.

Ulož data Objeví se standardní dialog pro výběr výstupního souboru. Je-li výstupní typ CSV je uživatel dotázán na oddělovač a má možnost nastavit zpracování prvního řádku. Byl-navíc vstupní soubor typu CXT nebo 3col má uživatel možnost nastavit zpracování prvního sloupce. Uživatel se vrací na okno ze kterého přišel.

Otevři data

4. Flereo User's guide

Flereo is command line tool for converting and analysing CSV, 3COL, ARF, CXT and DAT files.

4.1. Prerequisites

For unix based systems is required Mono version 2.10 and for windows system is required Microsoft .NET Framewrok 3.5.

4.2. Installation

Copy file Flereo.exe and FceHelper.dll to destination folder.

4.3. Parameters

Scénáře:

help or -h Write this help.

inputfile or -i Full path to input file. Parameter is required

outputfile or -o Full path to output file. Parameter is required for INPUTTYPE ARF, 3COL, CSV, DAT, CXT

inputtype or -it ARF, 3COL, CSV, DAT, CXT, CSV?, 3COL?, ARFQ?, ARF? Type CSV?, 3COL?, ARF? and ARFQ? only analyzes the input file This will help to prepare parameters for next work. CSV?, 3COL?, ARF? analyze the entire file and ARFQ? only read heads from ARF file.

outputtype or -ot ARF, 3COL, CSV, DAT, CXT. Parameter is required for INPUTTYPE ARF, 3COL, CSV, DAT.

csvseparator or -s Regular expression for the line separator in csv. Parameter is optional for INPUTTYPE CSV or OUTPUTTYPE CSV. Default value is , (comma)

csvrowsheader or -rh First row contains headers. Possible values are 1 - true or 0 - false. Parameter is optional for INPUTTYPE CSV or OUTPUTTYPE CSV. Default value is false.

csvcolumnsheader or -ch First column contains the names of objects. Possible values are 1 = true or 0 = false. Parameter is optional for OUTPUTTYPE CSV. Default value is false

fcabedrock or -b Full path to FcaBedrock settings file.

parameters or -p Special string describing the input data and their transformation. Parameter is required for INPUTTYPE ARF, 3COL, CSV with combination OUTPUTTYPE DAT, CXT For each column in input file is defined this structure: a[@b]—c—d[@e]—f#

a Position in the input file starting from zero.

@b Name of parameter. This parameter is optional and must start with @.

c: Date type of attribute - c Category, b Binary, n Number, u Notype, ! The column will not be processed

d Attribute values. More values are separated by :. For Category and Notype include search category name. For Binary include true value. If specified multiple values, accepts only the first. For Number include number or interval. Closed interval ;LowerBound;UpperBound; or open interval (LowerBound;UpperBound).

@e Output name of value. This parameter is optional and must start with @.

f Position in the output file starting from zero. This parameter is optional and can include value ! - The column will not be processed.

Example:

```
0@Color—c—r@Red:b@Blue:g@Green—1#1—n—(-  
5;30);j30;50):70.5—0#2@IsMerid—b—yes:no—2
```

Input file include 3 columns. First is Category type and include r,b,g value. In the output file will be renamed the Red, Blue and Green. Second column is Number type and will be evaluated for three conditions: Is larger than 30 and smaller than -5? Is greater than or equal to 30 and smaller than 50? Is 70.5? Third parametr is Binary type. Include two possible value value (yes, no) and in he output file will contain only the value for yes.

Flereo is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

Flereo is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

5. WinFlereo User's guide

WinFlereo is software for converting and analysing CSV, 3COL, ARF, CXT and DAT files.

5.1. Prerequisites

For unix based systems is required Mono version 2.10 and for windows system is required Microsoft .NET Framework 3.5.

5.2. Installation

Copy file WinFlereo.exe and FceHelper.dll to destination folder.

5.3. Main Application

Converting and transforming data with WinFlereo is easy.

1. Start Menu, selecting "Load data..." Picture 7.
2. Select an appropriate input file Picture 10.
3. Confirm data type Picture 11.
4. Explain attributes for transformation Picture 8. and 9.
5. Select the type of output file Picture 12. and 13.

6. Finish

5.3.1. Search for values

In the form you can specify attributes, their names and the order in which the export shed. For each attribute you can add any number of values ??for the search. Values ??can be any sort and name as attributes. In addition to type Number, you can specify the interval for evaluation. Closed interval $[LowerBound;UpperBound]$ or open interval $(LowerBound;UpperBound)$.

Závěr

Během práce jsem si rozšířil znalosti programování a softwarové architektury a naučil jsem se mnoho nových věcí.

Conclusions

While working to expand my knowledge of programming and software architecture and I learned many new things.

Reference

- [1] Uta Priss. *FCA file format conversion and interoperability software* Elektronická publikace., 2011
- [2] ANDREWS, Simon J. *Data conversion and interoperability for FCA* Elektronická publikace, 2009.

A. Obsah příloženého CD

V samotném závěru práce je uveden stručný popis obsahu příloženého CD/DVD, tj. závazné adresářové struktury, důležitých souborů apod.

`bin/`

Instalátor `INSTALATOR` programu a další program `PROGRAM` spustitelné přímo z CD/DVD. / Kompletní adresářová struktura webové aplikace `WEBOVKA` (v ZIP archivu) pro zkopírování na webový server. Adresář obsahuje i všechny potřebné knihovny a další soubory pro bezproblémové spuštění programu / pro bezproblémový provoz na webovém serveru.

`doc/`

Dokumentace práce ve formátu PDF, vytvořená dle závazného stylu KI PřF pro diplomové práce, včetně všech příloh, a všechny soubory nutné pro bezproblémové vygenerování PDF souboru dokumentace (v ZIP archivu), tj. zdrojový text dokumentace, vložené obrázky, apod.

`src/`

Kompletní zdrojové texty programu `PROGRAM` / webové aplikace `WEBOVKA` se všemi potřebnými (převzatými) zdrojovými texty, knihovnamy a dalšími soubory pro bezproblémové vytvoření spustitelných verzí programu / adresářové struktury pro zkopírování na webový server (v ZIP archivu).

`readme.txt`

Instrukce pro instalaci a spuštění programu `PROGRAM`, včetně požadavků pro jeho provoz. / Instrukce pro nasazení webové aplikace `WEBOVKA` na webový server, včetně požadavků pro její provoz, a webová adresa, na které je aplikace nasazena pro testovací účely a pro účel obhajoby práce.

Navíc CD/DVD obsahuje:

`data/`

Ukázková a testovací data použitá v práci a pro potřeby obhajoby práce.

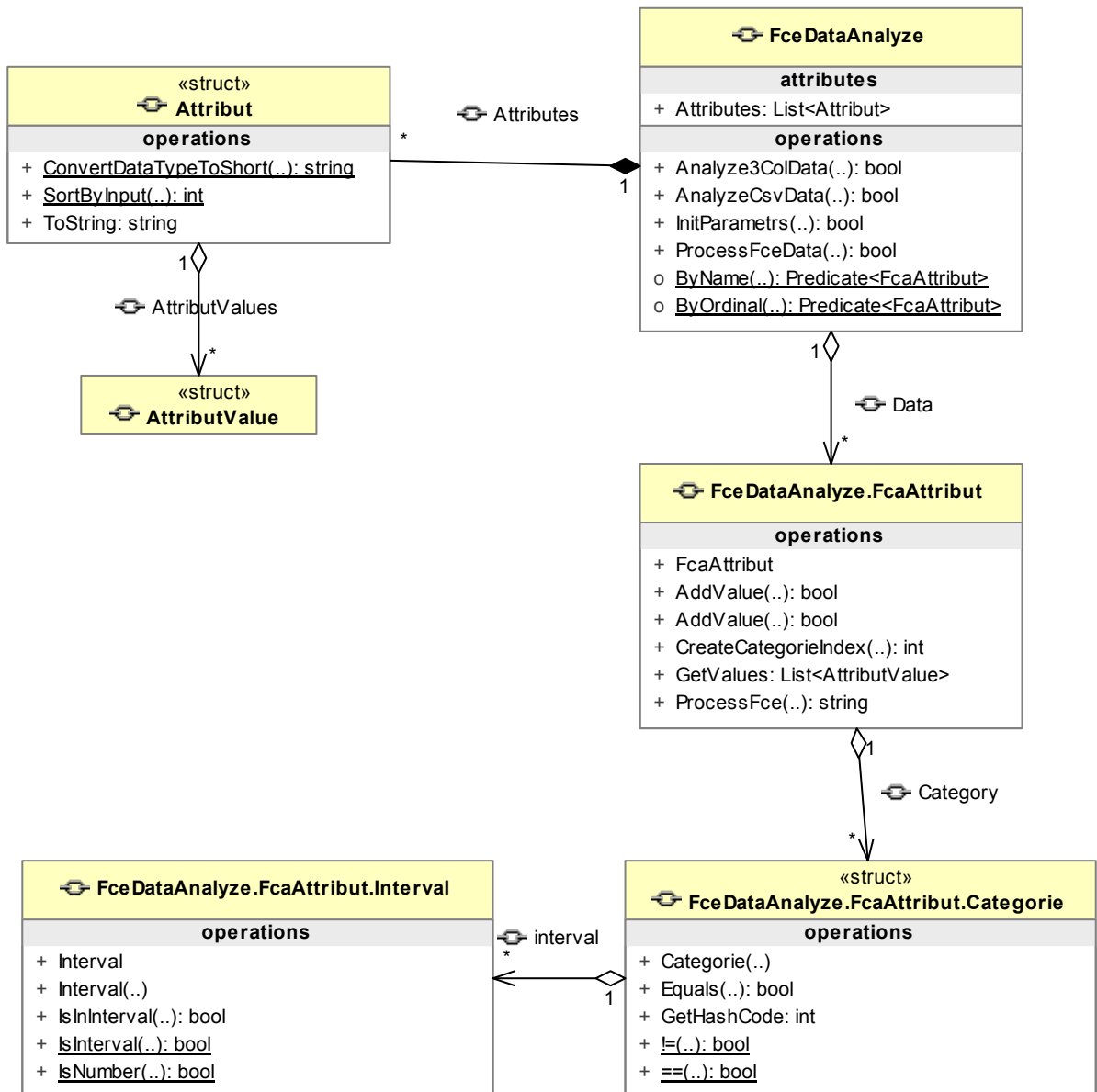
`install/`

Instalátory aplikací, knihoven a jiných souborů nutných pro provoz programu / webové aplikace, které nejsou standardní součástí operačního systému.

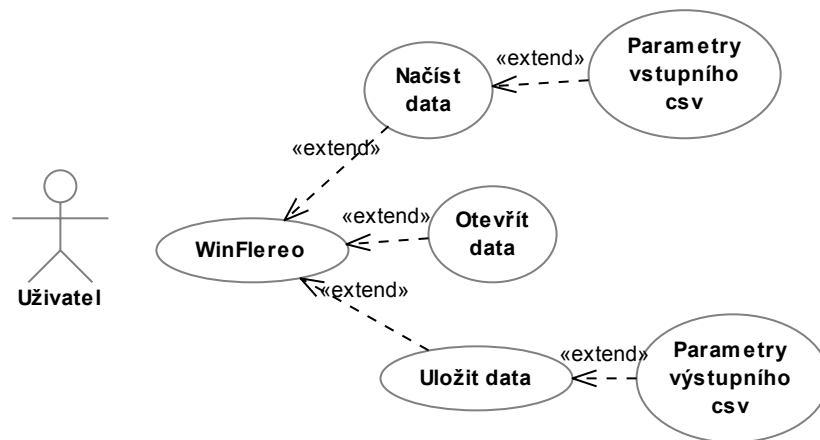
`literature/`

Některé položky literatury odkazované z dokumentace práce.

U veškerých odjinud převzatých materiálů obsažených na CD/DVD jejich zahrnutí dovoluují podmínky pro jejich šíření nebo přiložený souhlas držitele copyrightu. Pro materiály, u kterých toto není splněno, je uveden jejich zdroj (webová adresa) v textu dokumentace práce nebo v souboru `readme.txt`.



Obrázek 4. Class diagram knihovny FceHelper



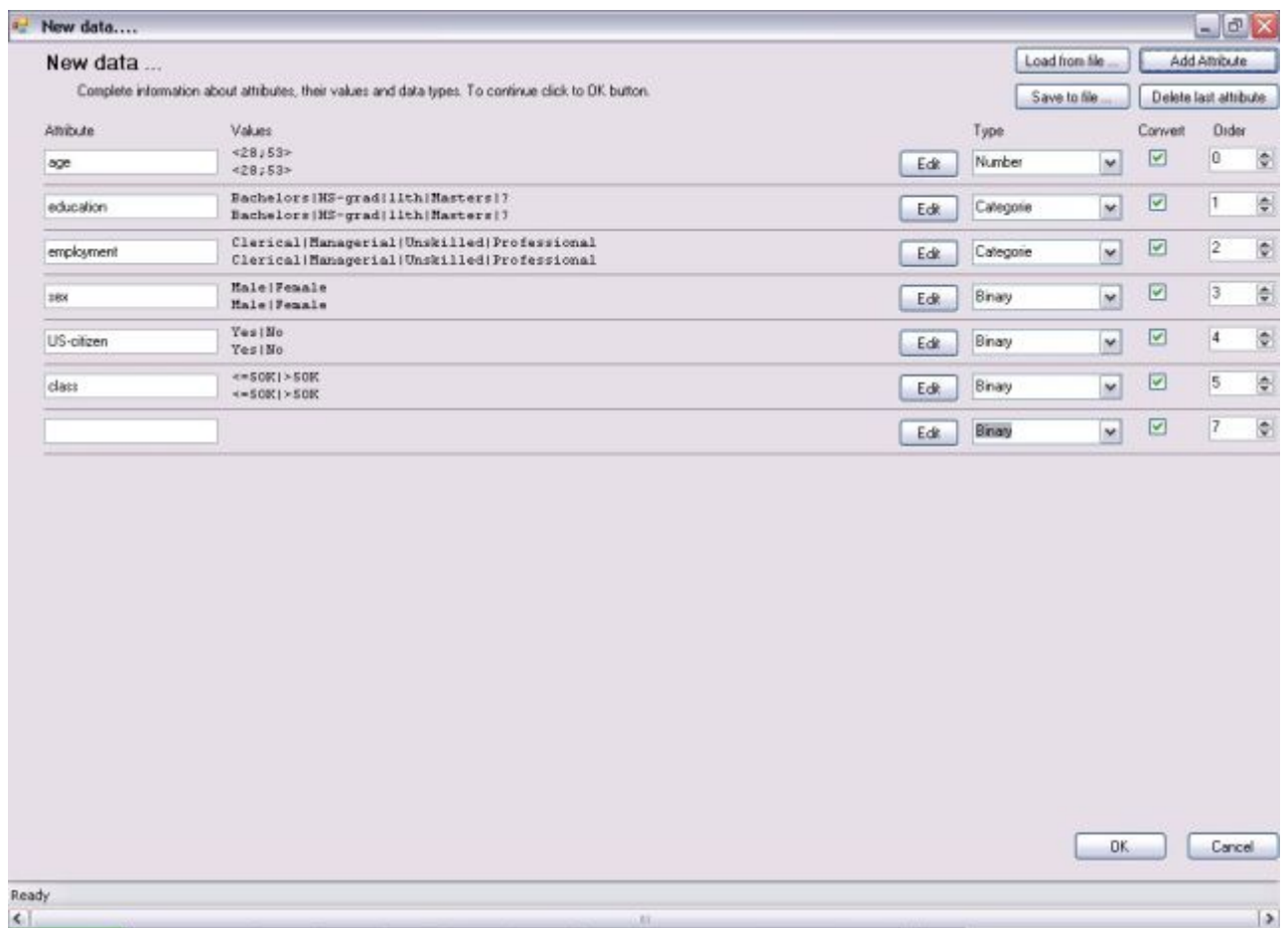
Obrázek 5. Use Case WinFlereo



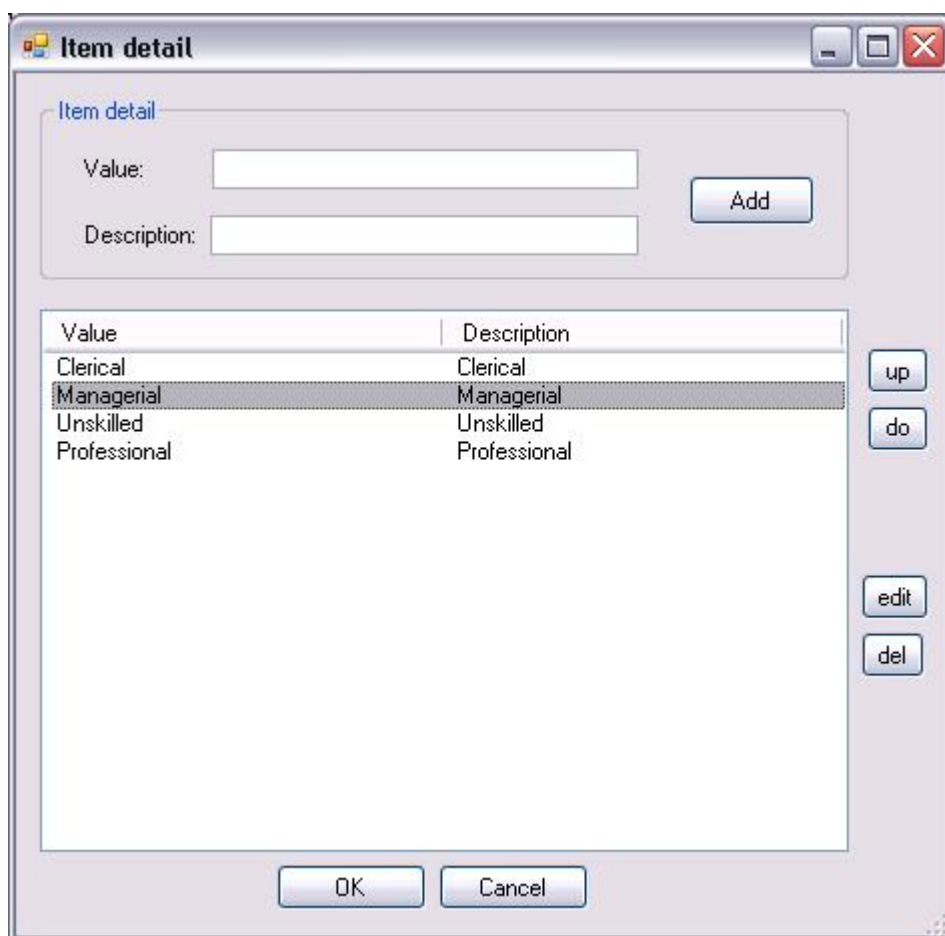
Obrázek 6. Main screen



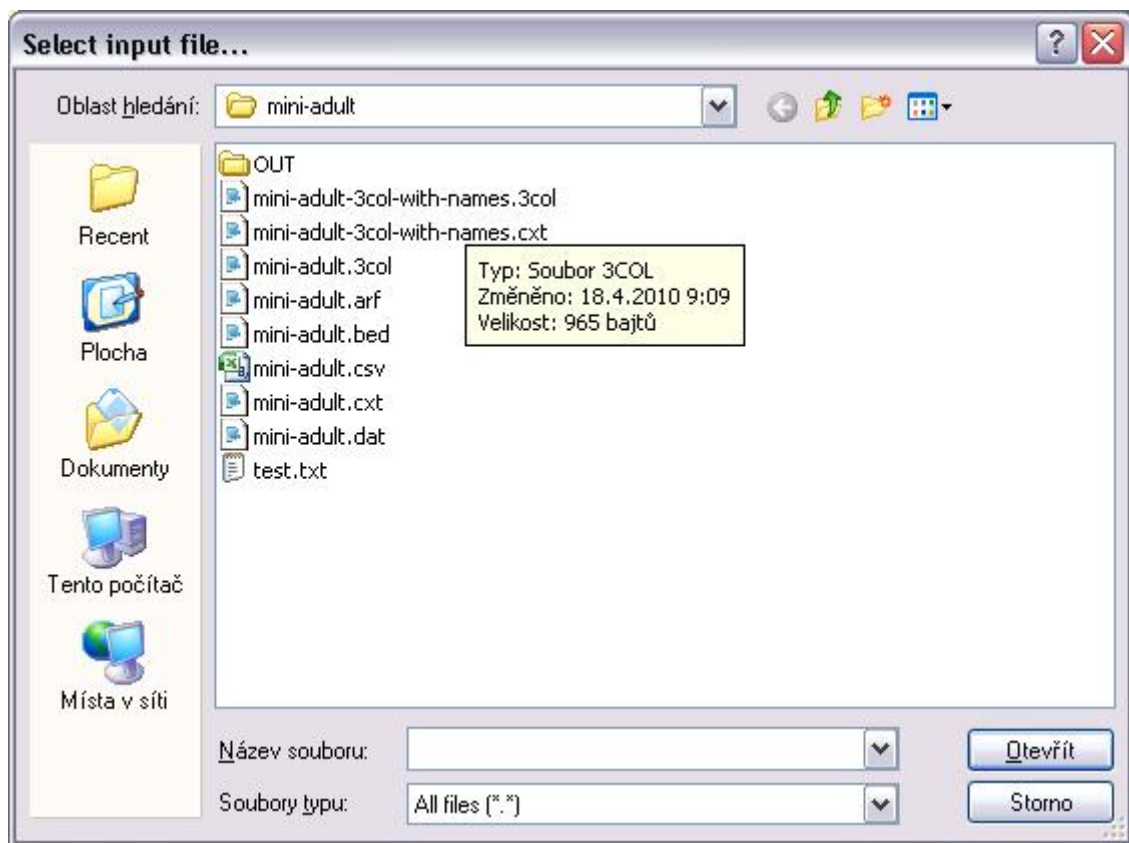
Obrázek 7. Menu Option



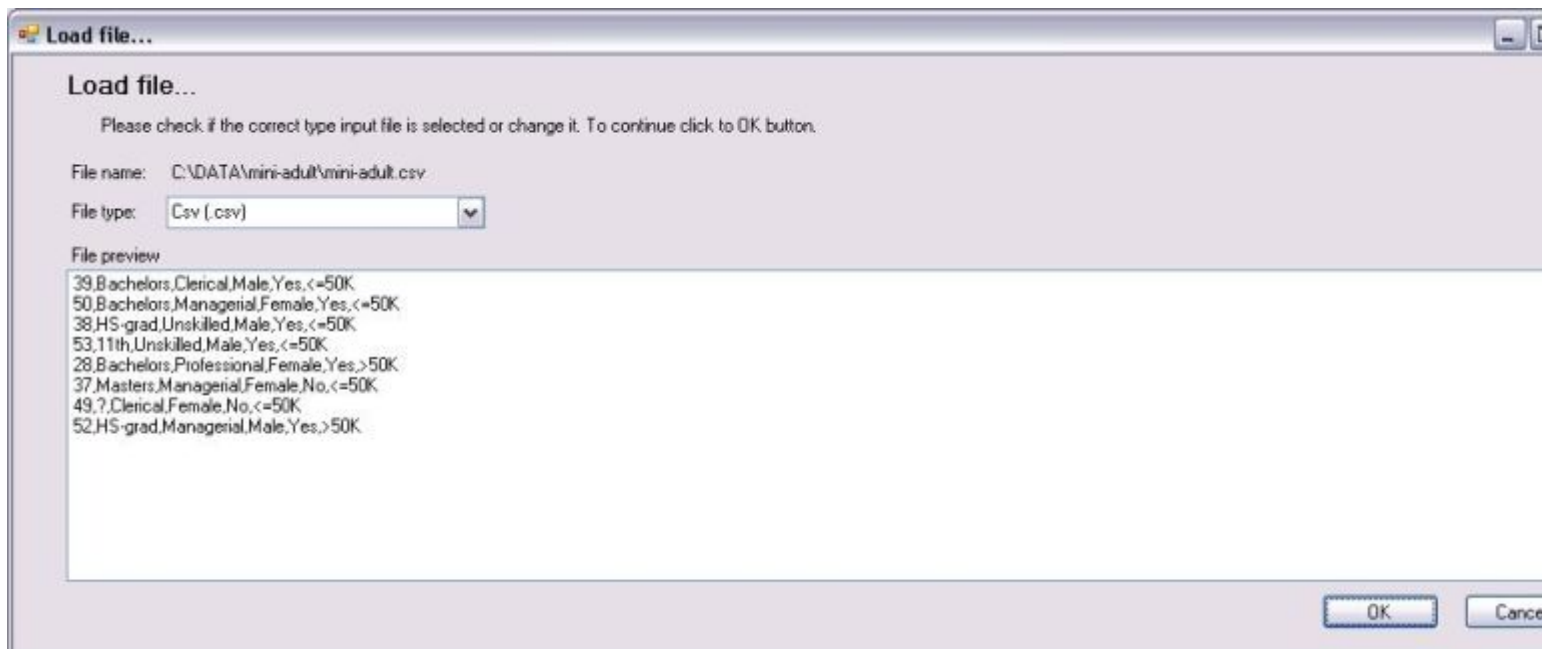
Obrázek 8. New Data Dialog



Obrázek 9. Item Detail



Obrázek 10. File Load 1



Obrázek 11. Save data dialog



Obrázek 12. Save Dialog



Obrázek 13. Save Csv Dialog