# ProgTest ► BI-PA2 (12/13 LS) ► Domácí úloha 01 ► Převody mezi soborovými formáty

Logout

## Převody mezi soborovými formáty

Termín odevzdání: 24.03.2013 23:59:59

Pozdní odevzdání s penalizací: 12.05.2013 23:59:59 (Penále za pozdní odevzdání: 100.0000 %)

Hodnocení: 4.4000

Max. hodnocení: 4.0000 (bez bonusů)

**Odevzdaná řešení:** 5 / 20 Volné pokusy + 20 Penalizované pokusy (-2 % penalizace za každé odevzdání)

Nápovědy: 3 / 2 Volné nápovědy + 2 Penalizované nápovědy (-10 % penalizace za každou

nápovědu)

Úkolem je realizovat funkce, která provedou konverzi mezi souborovými formáty.

Funkce budou provádět konverzi mezi formátem CSV a binární formou. CSV je souborový formát, který se používá pro výměnu dat mezi systémy různých výrobců. Jeho velkou výhodou je jednoduchost a snadná implementace. CSV soubor ukládá informace v řádcích, jedna řádka tvoří jeden záznam. Informace na řádce jsou oddělené pomocí oddělovačů (typicky tabulátor, čárka nebo středník, v naší implementaci bude použit středník).

Předpokládejme, že chceme ukládat informace o lidech - například jméno, příjmení a datum narození. Ve formátu CSV by to bylo uloženo např. následovně:

```
Al;Capone;1899-01-17
Vito;Corleone;1891-12-07
Genco;Abbandando;1895-08-06
```

Pokud jsou v CSV souboru ukládána pouze data bez odřádkování, středníků a uvozovek, je formát velmi přímočarý. Pokud bychom chtěli ukládat i tyto znaky, bude ukládání a hlavně načítání složitější - taková data musí být uložena v uvozovkách:

- data mohou být uložena přímo (bez uvozovek), pokud neobsahují odřádkování, uvozovku ani oddělovač (středník),
- data mohou být vždy uložena v uvozovkách, uvozovky slouží k vymezení počátku/konce dat. Při ukládání se uvozovky přidají, při načítání odstraní,
- data v uvozovkách mohou obsahovat libovolné znaky včetně středníku a odřádkování, tyto speciální uvnitř uvozovek nemají svůj speciální význam,
- pokud data uvnitř obsahují znak uvozovka, musí být uvozovka zdvojená.

Například toto je jeden záznam:

```
Vito;Corleone;"the ""Godfather""; the head of the ""Corleone"" family"
```

Tento záznam je tvořen následujícími třemi řetězci:

```
Vito
Corleone
```

the "Godfather"; the head of the "Corleone" family

Druhým formátem, se kterým bude Vaše aplikace pracovat, je binární reprezentace CSV. Tento souborový formát není nijak standardizovaný, byl vytvořen pouze pro účely této domácí úlohy. Formát je tvořen hlavičkou souboru za kterou následuje seznam záznamů. Každý záznam je pak tvořen několika složkami. Záznam odpovídá řádce CSV, složka pak jednomu řetězci v záznamu. Složky se rozlišují řetězcové a celočíselné.

```
Hlavička souboru:
```

```
Offset
        velikost
                    význam
+0
         1 B
                    znak H (identifikace hlavičky)
                    int hodnota - velikost obsahu (počet bajtů, počínaje
+1
         4 B
                    prvním bajtem za koncem hlavičky)
+5
         4 B
                    int hodnota - počet záznamů v souboru (počet řádek)
+9
         4 B
                    int hodnota - počet složek v každém záznamu
```

Za hlavičkou následuje seznam záznamů. Počet záznamů je udán v hlavičce souboru. Každý záznam je uvozen hlavičkou záznamu:

```
Hlavička záznamu:
```

```
Offset velikost význam
```

+0 1 B znak R (identifikace záznamu)

```
+1 4 B int hodnota - délka tohoto záznamu v bajtech, počínaje
prvním bajtem za touto hlavičkou záznamu
```

Za hlavičkou záznamu jsou uložené jednotlivé složky záznamu. Počet složek záznamu je určen v hlavičce souboru. Složkami záznamu jsou buď celá čísla nebo řetězce (pokud lze data reprezentovat jako celé číslo typu int, ukládají se jako celé číslo).

```
Složka typu celé číslo:
                    význam
Offset
        velikost
+0
         1 B
                    znak I (identifikace složky typu int)
                    int - hodnota čísla
+1
         4 B
Složka typu řetězec:
Offset
        velikost
                    význam
                    znak S (identifikace složky typu string)
+0
         1 B
+1
         4 B
                    int - počet znaků v řetězci
         ? B
+5
                    znaky řetězce
```

Je vidět, že binární formát je snazší na čtení, pokud data obsahují speciální znaky. Dále je vidět, že v binárním souboru lze snáze přeskakovat záznamy (není třeba je číst, stačí načíst jejich velikost v bajtech). Obtížnější bude zápis binárního souboru, kde bude potřeba správně uložit velikost složek a záznamů.

Vaším úkolem je realizovat funkce:

```
#ifndef _
          PROGTEST__
#include <iostream>
#include <iomanip>
#include <fstream>
#include <sstream>
#include <string>
#include <cstdio>
#include <cstdlib>
#include <cstring>
#include <cctype>
#include <stdint.h>
using namespace std;
#endif /* __PROGTEST__ */
bool BINtoCSV ( const char * inFileName, const char * outFileName )
{
   // todo
}
bool CSVtoBIN ( const char * inFileName, const char * outFileName )
{
   // todo
}
```

Funkce provádí konverzi vstupního souboru určeného jménem inFileName do souboru v cílovém formátu, jméno výstupního souboru je dané parametrem outFileName. Návratovou hodnotou je příznak úspěchu true nebo selhání false. Za chybu považujte následující:

- chyba při práci se souborem (nelze otevřít/číst/zapisovat/...),
- chyba ve struktuře dat v souboru (špatné formátování CSV, chybné/chybějící/poškozené hlavičky BIN, ...),
- chyba v obsahu CSV (počet složek se liší pro jednotlivé řádky, ...).

Odevzdávejte zdrojový kód, který obsahuje Vaši implementaci požadovaných funkcí. Do odevzdávaného souboru samozřejmě patří ještě další Vaše funkce, které jsou z těchto funkcí volané. Naopak, v odevzdávaném souboru nesmí být vkládání hlavičkových souborů a funkci main zabalíte do bloku podmíněného překladu, mohou zůstat). Za základ řešení použijte ukázku výše. Ukázku si zkopírujte a doplňte pouze implementaci požadované funkce (případně další Vaše funkce). Pokud zachováte bloky podmíněného překladu, můžete takový soubor odevzdávat Progtestu.

Hodnocení úlohy je rozdělení na závaznou a bonusovou část. V závazném testu se pracuje s CSV soubory, které neobsahují speciální znaky, tedy datový obsah nemusí být obalen uvozovkami. To zjednodušuje zejména čtení CSV souborů. V posledním (bonusovém) testu jsou pak přidané vstupy, které obalení uvozovkami potřebují.

Pro tuto úlohu nejsou k dispozici ukázky použití funkcí. Význam parametrů a volání funkcí je zřejmé. V přiloženém archivu máte k dispozici sadu testovacích souborů, které budou předkládány Vašim funkcím v základním testu. Zároveň si v nich můžete ověřit, že skutečně rozumíte způsobu ukládání dat v binárním souboru. Vstupní data in\_3.csv jsou neplatná, proto pro ně v archivu neexistuje odpovídající výstupní soubor.

6/12/2017

Vzorová data: Download

## Referenční řešení

5	18.03.2013 23:16:04	Download
Stav odevzdání:	Ohodnoceno	
Hodnocení:	3.2400	

#### · Hodnotitel: automat

- Program zkompilován
- Test 'Zakladni test se soubory dle ukazky': Úspěch
  - Dosaženo: 100.00 %, požadováno: 100.00 %
  - Celková doba běhu: 0.001 s (limit: 2.000 s)
  - Využití paměti: 12592 KiB (limit: 14337 KiB)
  - Úspěch v závazném testu, hodnocení: 100.00 %
- Test 'Nahodny test (do csv)': Úspěch
  - Dosaženo: 90.00 %, požadováno: 50.00 %
  - Celková doba běhu: 0.032 s (limit: 1.999 s)
  - Využití paměti: 12592 KiB (limit: 14337 KiB)
  - Úspěch v závazném testu, hodnocení: 90.00 %
  - Nesprávný výstup [Zpřístupnit nápovědu (54.78 KiB)]
- Test 'Nahodny test (do bin)': Úspěch
  - Dosaženo: 100.00 %, požadováno: 50.00 %
  - Celková doba běhu: 0.021 s (limit: 1.967 s)
  - Využití paměti: 12592 KiB (limit: 14337 KiB)
  - Úspěch v závazném testu, hodnocení: 100.00 %
- Test 'Nespravne vstupy (csv)': Úspěch
  - Dosaženo: 100.00 %, požadováno: 75.00 %
  - Celková doba běhu: 0.000 s (limit: 2.000 s)
  - Využití paměti: 12592 KiB (limit: 28009 KiB)
  - Úspěch v nepovinném testu, hodnocení: 100.00 %
- Test 'Nespravny vstup (bin)': Úspěch
  - Dosaženo: 100.00 %, požadováno: 75.00 %
  - Celková doba běhu: 0.001 s (limit: 2.000 s)
  - Využití paměti: 12592 KiB (limit: 28009 KiB)
  - Úspěch v nepovinném testu, hodnocení: 100.00 %
- Test 'Test osetreni chyb pri I/O': Úspěch
  - Dosaženo: 100.00 %, požadováno: 25.00 %
  - Celková doba běhu: 0.278 s (limit: 1.999 s)
  - Využití paměti: 12592 KiB (limit: 28009 KiB)
  - Úspěch v nepovinném testu, hodnocení: 100.00 %
- Test 'CSV s retezci v uvozovkach': Neúspěch
  - Dosaženo: 46.67 %, požadováno: 100.00 %
  - Celková doba běhu: 0.047 s (limit: 2.000 s)
  - Využití paměti: 12592 KiB (limit: 14337 KiB)
  - Neúspěch v bonusovém testu, hodnocení: Bonus nebude udělen
  - Nesprávný výstup [Zpřístupnit nápovědu (63 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (59 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (56 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (83 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (32.04 KiB)]
  - Nesprávný výstup [Zpřístupnit nápovědu (8.42 KiB)]
  - Nesprávný výstup [Zpřístupnit nápovědu (2.54 KiB)]
  - Nesprávný výstup [Zpřístupnit nápovědu (4.44 KiB)]
- Celkové hodnocení: 90.00 % (= 1.00 \* 0.90 \* 1.00 \* 1.00 \* 1.00 \* 1.00)
- Použité nápovědy: 3
- Penalizace za vyčerpané nápovědy: 10.00 % (= (3 2) \* 10 %)
- Celkové procentní hodnocení: 81.00 % (= 0.90 \* 0.90)
- Celkem bodů: 0.81 \* 4.00 = 3.24

SW metriky:		Celkem	Průměr	Maximum	Jméno funkce
	Funkce:	19			
	Řádek kódu:	332	17.47 ± 24.56	108	<pre>CSVtoBIN(const char *,const char *)</pre>

Cyklomatická složitost: 121  $6.37 \pm 8.30$ 

32 CSVtoBIN(const char \*,const
 char \*)

4	14.03.2013 10:58:19	Download
Stav odevzdání:	Ohodnoceno	
Hodnocení:	3.9600	
Stav odevzdání: Hodnocení:  Hodnotitel: autor Program zko Test 'Zaklad Dosa Celk Využ Využ Využ Využ Vspě Nesp Test 'Nahod Dosa Celk Využ Využ Vspě Test 'Nspra Dosa Celk Využ Využ Vspě Test 'Nespra Dosa Celk Využ Využ Vspě Test 'Nespra Dosa Celk Využ Vspě Test 'Nespra Dosa Celk Využ Vspě Test 'Nespra Dosa Celk Využ Vspě Test 'Test o Dosa Celk Využ	Ohodnoceno 3.9600  mat  mat  mi test se soubory dle ukazky': Úspěch nženo: 100.00 %, požadováno: 100.00 %  pová doba běhu: 0.001 s (limit: 2.000 s)  mi pilován  mi test se soubory dle ukazky': Úspěch nženo: 100.00 %, požadováno: 100.00 %  pová doba běhu: 0.001 s (limit: 14337 KiB)  mi test (do csy)': Úspěch nženo: 90.00 %, požadováno: 50.00 %  pová doba běhu: 0.028 s (limit: 1.999 s)  mi ti paměti: 12592 KiB (limit: 14337 KiB)  mi test (do bin)': Úspěch nženo: 100.00 %, požadováno: 50.00 %  mi test (do bin)': Úspěch nženo: 100.00 %, požadováno: 50.00 %  mi ti paměti: 12592 KiB (limit: 14337 KiB)  mi to province vstupy (csv)': Úspěch nženo: 100.00 %, požadováno: 75.00 %  mi to province vstupy (csv)': Úspěch nženo: 100.00 %, požadováno: 75.00 %  mi paměti: 12592 KiB (limit: 28009 KiB)  mi paměti: 12592 KiB (limit: 28009 KiB)	Download
<ul><li>Celk</li><li>Využ</li><li>Neús</li><li>Nesp</li><li>Nesp</li></ul>	ová doba běhu: 0.073 s (limit: 2.000 s) dití paměti: 12592 KiB (limit: 14337 KiB) spěch v bonusovém testu, hodnocení: Bonus nebude udělen právný výstup [ <b>Zpřístupnit nápovědu (83 B)]</b> právný výstup [ <b>Zpřístupnit nápovědu (82 B)</b> ]	
<ul> <li>Nesp</li> <li>Nesp</li> <li>Nesp</li> <li>Nesp</li> <li>Celkové hoc</li> <li>Použité nápovědy:</li> <li>Penalizace za vyčer</li> </ul>	Nesprávný výstup Nesprávný výstup orávný výstup [ <b>Zpřístupnit nápovědu (32.04 KiB)</b> ] orávný výstup [ <b>Zpřístupnit nápovědu (10.24 KiB)</b> ] orávný výstup [ <b>Zpřístupnit nápovědu (32.04 KiB)</b> ] orávný výstup [ <b>Zpřístupnit nápovědu (32.04 KiB)</b> ] Inocení: 90.00 % (= 1.00 * 0.90 * 1.00 * 1.00 * 1.00 * 1 pané nápovědy: Není (1 <= 2 limit) nodnocení: 90.00 %	.00)
<ul> <li>Bonus za včasné od</li> </ul>		

SW metriky:		Celkem	Průměr	Maximum	Jméno funkce
	Funkce:	19			
	Řádek kódu:	321	16.89 ± 22.46	97	<pre>CSVtoBIN(const char *,const char *)</pre>
	Cyklomatická	119 6.	26 ± 7.99	30	CSVtoBIN(const char *,const

složitost: char \*)

3 14.03.2013 10:55:35 **Download** Stav odevzdání: Ohodnoceno Hodnocení: 4.4000 · Hodnotitel: automat Program zkompilován Test 'Zakladni test se soubory dle ukazky': Úspěch Dosaženo: 100.00 %, požadováno: 100.00 % Celková doba běhu: 0.001 s (limit: 2.000 s) Využití paměti: 12592 KiB (limit: 14337 KiB) Úspěch v závazném testu, hodnocení: 100.00 % Test 'Nahodny test (do csv)': Úspěch Dosaženo: 100.00 %, požadováno: 50.00 % Celková doba běhu: 0.027 s (limit: 1.999 s) Využití paměti: 12592 KiB (limit: 14337 KiB) Úspěch v závazném testu, hodnocení: 100.00 % Test 'Nahodny test (do bin)': Úspěch Dosaženo: 100.00 %, požadováno: 50.00 % Celková doba běhu: 0.019 s (limit: 1.972 s) Využití paměti: 12592 KiB (limit: 14337 KiB) Úspěch v závazném testu, hodnocení: 100.00 % Test 'Nespravne vstupy (csv)': Úspěch Dosaženo: 100.00 %, požadováno: 75.00 % Celková doba běhu: 0.000 s (limit: 2.000 s) Využití paměti: 12592 KiB (limit: 28009 KiB) Úspěch v nepovinném testu, hodnocení: 100.00 % Test 'Nespravny vstup (bin)': Úspěch Dosaženo: 100.00 %, požadováno: 75.00 % Celková doba běhu: 0.001 s (limit: 2.000 s) Využití paměti: 12592 KiB (limit: 28009 KiB) Úspěch v nepovinném testu, hodnocení: 100.00 % Test 'Test osetreni chyb pri I/O': Úspěch Dosaženo: 100.00 %, požadováno: 25.00 % Celková doba běhu: 0.290 s (limit: 1.999 s) Využití paměti: 12592 KiB (limit: 28009 KiB) Úspěch v nepovinném testu, hodnocení: 100.00 % Test 'CSV s retezci v uvozovkach': Neúspěch Dosaženo: 50.00 %, požadováno: 100.00 % Celková doba běhu: 0.070 s (limit: 2.000 s) Využití paměti: 12592 KiB (limit: 14337 KiB) Neúspěch v bonusovém testu, hodnocení: Bonus nebude udělen Nesprávný výstup [Zpřístupnit nápovědu (83 B)] Nesprávný výstup [Zpřístupnit nápovědu (82 B)] Nesprávný výstup Nesprávný výstup [Zpřístupnit nápovědu (41 B)] Nesprávný výstup [Zpřístupnit nápovědu (32.04 KiB)] Nesprávný výstup [Zpřístupnit nápovědu (20.94 KiB)] Nesprávný výstup [Zpřístupnit nápovědu (32.04 KiB)] Nesprávný výstup [Zpřístupnit nápovědu (32.04 KiB)] Celkové hodnocení: 100.00 % (= 1.00 \* 1.00 \* 1.00 \* 1.00 \* 1.00 \* 1.00) Celkové procentní hodnocení: 100.00 % Bonus za včasné odevzdání: 0.40 Celkem bodů: 1.00 \* ( 4.00 + 0.40 ) = 4.40

	C	Celkem	Průměr	Maximum	Jméno funkce
	Funkce:	19			
SW metriky:	Řádek kódu:	337	17.74 ± 21.70	93	CSVtoBIN(const char *,const char *)
	Cyklomatická složitost:	122 6	.42 ± 7.62	28	CSVtoBIN(const char *,const char *)

2 14.03.2013 10:42:10 Download

Stav odevzdání: Ohodnoceno
Hodnocení: 3.5640

#### · Hodnotitel: automat

- Program zkompilován
- Test 'Zakladni test se soubory dle ukazky': Úspěch
  - Dosaženo: 100.00 %, požadováno: 100.00 %
  - Celková doba běhu: 0.001 s (limit: 2.000 s)
  - Využití paměti: 12588 KiB (limit: 14337 KiB)
  - Úspěch v závazném testu, hodnocení: 100.00 %
- Test 'Nahodny test (do csv)': Úspěch
  - Dosaženo: 90.00 %, požadováno: 50.00 %
  - Celková doba běhu: 0.030 s (limit: 1.999 s)
  - Využití paměti: 12588 KiB (limit: 14337 KiB)
  - Úspěch v závazném testu, hodnocení: 90.00 %
  - Nesprávný výstup [Zpřístupnit nápovědu (38.39 KiB)]
- Test 'Nahodny test (do bin)': Úspěch
  - Dosaženo: 100.00 %, požadováno: 50.00 %
  - Celková doba běhu: 0.013 s (limit: 1.969 s)
  - Využití paměti: 12588 KiB (limit: 14337 KiB)
  - Úspěch v závazném testu, hodnocení: 100.00 %
- Test 'Nespravne vstupy (csv)': Úspěch
  - Dosaženo: 100.00 %, požadováno: 75.00 %
  - Celková doba běhu: 0.000 s (limit: 2.000 s)
  - Využití paměti: 12588 KiB (limit: 28009 KiB)
  - Úspěch v nepovinném testu, hodnocení: 100.00 %
- Test 'Nespravny vstup (bin)': Úspěch
  - Dosaženo: 90.00 %, požadováno: 75.00 %
  - Celková doba běhu: 0.001 s (limit: 2.000 s)
  - Využití paměti: 12588 KiB (limit: 28009 KiB)
  - Úspěch v nepovinném testu, hodnocení: 90.00 %
  - Nesprávný výstup [Zpřístupnit nápovědu (225 B)]
- Test 'Test osetreni chyb pri I/O': Úspěch
  - Dosaženo: 100.00 %, požadováno: 25.00 %
  - Celková doba běhu: 0.291 s (limit: 1.999 s)
  - Využití paměti: 12588 KiB (limit: 28009 KiB)
  - Úspěch v nepovinném testu, hodnocení: 100.00 %
- Test 'CSV s retezci v uvozovkach': Neúspěch
  - Dosaženo: 50.00 %, požadováno: 100.00 %
  - Celková doba běhu: 0.059 s (limit: 2.000 s)
  - Využití paměti: 12588 KiB (limit: 14337 KiB)
  - Neúspěch v bonusovém testu, hodnocení: Bonus nebude udělen
  - Nesprávný výstup [Zpřístupnit nápovědu (83 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (82 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (41 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (41 B)]
  - Nesprávný výstup [Zpřístupnit nápovědu (32.04 KiB)]
  - Nesprávný výstup [Zpřístupnit nápovědu (25.61 KiB)]
  - Nesprávný výstup [Zpřístupnit nápovědu (32.04 KiB)]
  - Nesprávný výstup [Zpřístupnit nápovědu (14.84 KiB)]
- Celkové hodnocení: 81.00 % (= 1.00 \* 0.90 \* 1.00 \* 1.00 \* 0.90 \* 1.00)
- Celkové procentní hodnocení: 81.00 %
- Bonus za včasné odevzdání: 0.40
- Celkem bodů: 0.81 \* ( 4.00 + 0.40 ) = 3.56

		Celkem	Průměr	Maximum	Jméno funkce
	Funkce:	19			
SW metriky:	Řádek kódu:	316	16.63 ± 20.90	93	CSVtoBIN(const char *,const char *)
	Cyklomatická složitost:	103 5	.42 ± 6.71	28	CSVtoBIN(const char *,const char *)

Stav odevzdání: Ohodnoceno	
Hodnocení: 0.0000	

### • Hodnotitel: automat

- Program zkompilován
- Test 'Zakladni test se soubory dle ukazky': Neúspěch
- Test Zakladní test se soubory dle ukazky: Neuspech
   Dosaženo: 83.33 %, požadováno: 100.00 %
   Celková doba běhu: 0.001 s (limit: 2.000 s)
   Využití paměti: 12588 KiB (limit: 14337 KiB)
   Neúspěch v závazném testu, hodnocení: 0.00 %
   Nesprávný výstup [Zpřístupnit nápovědu (50 B)]
   Celkové hodnocení: 0.00 %
- Celkové procentní hodnocení: 0.00 %
- Bonus za včasné odevzdání: 0.40
   Celkem bodů: 0.00 \* (4.00 + 0.40) = 0.00

/ <del></del>					
		Celkem	Průměr	Maximum	Jméno funkce
	Funkce:	19			
SW metriky:	Řádek kódu:	292	15.37 ± 20.31	90	CSVtoBIN(const char *,const char *)
	Cyklomatická složitost:	96 5.	05 ± 6.39	26	CSVtoBIN(const char *,const char *)