

Praktikum iz programiranja 2

Materijal potreban za peti domaći zadatak

Detaljnija objašnjenja o postojećim formatima datoteka sa kontaktima (address book) se mogu pronaći uz pomoć Internet pretraživača i na Internet enciklopediji www.wikipedia.org.

U ovom dokumentu je priložen kratak i nepotpun opis značajnih formata datoteka za čuvanje i razmenu podataka o kontaktima (address book) potreban za izradu domaćeg zadatka, kao i primer jednog istog kontakta predstavljenog različitim formatima.

Sadržaj

Ukratko o formatima datoteka sa kontaktima	2
Formati datoteka sa kontaktima	2
Comma Separated Values (.CSV)	2
Mozilla Thunderbird CSV format	2
Microsoft Address Book CSV format.....	3
Eudora address book format (.TXT)	3
Mozilla Thunderbird LDAP Data Interchange format (.LDIF)	4
Internet Mail Consortium vCard format (.VCF).....	5

Ukratko o formatima datoteka sa kontaktima

Sa razvojem Interneta i elektronske pošte (e-mail) i njihovom sve širom dostupnošću velikom broju korisnika, došlo je do pojave velikog broja programa za razmenu elektronske pošte (MS Outlook i Outlook Express, Mozilla Thunderbird, Qualcomm Eudora...), kao i velikog broja web baziranih servisa za razmenu elektronske pošte (Yahoo! Mail, Gmail, MS Hotmail...). Neizostavno, sa razvojem ovih programa i širenjem broja korisnika elektronske pošte, razvijali su se i moduli/aplikacije za upravljanje kontaktima, kao podrška korisnicima prilikom korišćenja programa za razmenu pošte. Pored osnovne funkcionalnosti, čuvanja e-mail adresa osoba, ovi programi su omogućili i skladištenje različitih informacija o samom kontaktu, kao što su adresa stanovanja i posla, telefoni, različite zabeleške, te su time postali svojevrsni organizatori kontakata, slično papirnim adresarima. Stalna konkurencija između proizvođača programa za razmenu elektronske pošte, kao i sve veći napredak web baziranih servisa za razmenu pošte uzrokuju sve češće selidbe korisnika sa jedne aplikacije na drugu. Međutim, korisnici često imaju potrebu da prilikom promene aplikacije za razmenu pošte zadrže svoje stare poruke i kontakte. Kako su aplikacije za upravljanje kontaktima najčešće integrisane unutar programa za razmenu elektronske pošte, one obično koriste neki interni (proprietary) format za skladištenje podataka o kontaktima, tako da njihova razmena između aplikacija različitih proizvođača nije nimalo jednostavna. Tokom vremena se izdvojilo nekoliko tekstualnih formata za razmenu kontakata, od kojih su neki neformalni standardi, dok drugi imaju standardizovanu formu. Ipak, različite aplikacije za upravljanje kontaktima mogu da generišu datoteke sa kontaktima na različit način, čak za isti format datoteke, tako da može doći do nepravilnog uvoza kontakata u neku aplikaciju. Ovo se najčešće dešava zbog različitog broja informacija i različitog rasporeda informacija unutar datoteke koje generišu različite aplikacije. Zato je često potrebno namenskim programom uraditi konverziju datoteka iz jednog formata u drugi kako bi se uvoz kontakata izvršio sa što manjim gubitkom podataka.

Formati datoteka sa kontaktima

Postoji nekoliko značajnijih formata datoteka sa kontaktima, koji se razlikuju kako po internoj organizaciji (strukturi), tako i po broju i načinu predstavljanja pojedinih informacija u samom formatu. Većina formata ima promenljiv broj parametara koje sadrži, kao i promenljiv raspored informacija unutar formata. Zbog jednostavnosti, pri izradi domaćeg zadatka za svaki format će biti fiksiran format za raspored informacija unutar jednog kontakta.

Comma Separated Values (.CSV)

Ovo je jedan od najčešćih formata za razmenu informacija. Nastao je kao format za razmenu informacija između programa za tabelarna izračunavanja. Našao je primenu i u aplikacijama za razmenu kontakata. Iako često korišćen, ne postoji njegova formalna specifikacija, već mnogobrojne različite implementacije, u zavisnosti od aplikacije. Ipak, postoje neke zajedničke karakteristika koje poštuju skoro sve implementacije:

- CSV je tekstualni format, koji sadrži zapise, koji se sastoje od polja razdvojenih znakom za razdvajanje, koji je najčešće *zarez (comma)*, dok su zapisi razdvojeni oznakom za kraj reda;
- svaki zapis (red datoteke) mora imati isti broj polja;
- svako polje može, ali najčešće ne mora biti, biti unutar znaka navoda "" (double quotes);
- polja koja sadrže znakove koji predstavljaju zarez, znake navoda i znak za novi red moraju biti stavljena unutar znaka navoda;
- u datoteci može postojati opciono zaglavlje u prvom redu datoteke, po istom formatu koji važi za polja unutar reda; zaglavlje obično sadrži imena odgovarajućih polja u ostatku datoteke;
- ukoliko za neko polje ne postoji vrednost, ono se ostavlja prazno, odnosno neće postojati nikakva informacija između dva znaka za razdvajanje;

Više detalja o CSV formatu se može pronaći u Request For Comment (RFC) dokumentu 4180, na Wikipedia enciklopediji i na Internetu.

Mozilla Thunderbird CSV format

Ovaj format najbolje odgovara opisanim karakteristikama CSV formata. Zaglavlje ne postoji u datoteci, a polja su razdvojena znakom *zarez (comma)*. Ukoliko postoji neka vrednost koja sadrži specijalne znake, biće stavljena pod znake navoda. Znak za razdvajanje postoji i nakon poslednjeg polja. Značenje odgovarajućih polja, po redosledu pojavljivanja u datoteci, je sledeće:

```
first name,last name,full name,nickname,e-mail,additional e-mail,work phone,home
phone,fax,pager,mobile,address,number,city,state/province,postal code,countru,web
page,work title,department,organization,org. address,org. number,org. city,org.
state/province,org. postal code,org. country,org. web.
page,custom1,custom2,custom3,custom4, custom5, custom6, custom7 notes,
```

Custom polja je potrebno zanemariti prilikom čitanja, ona će biti prazna, kao u primeru.

Primer kontakta:

```
Petar,Peric,Petar Peric,Peca,peraperic@etf.bg.ac.yu,pera@yahoo.com,011/222-
555,013/777-555,011/888-555,060/112,064/999-999,Programskog jezika
C,38,Pancevo,Vojvodina,26000,Srbija,Bulevar Kralja
Aleksandra,73,Beograd,Beograd,11000,Srbija,Dipl. ing.,RTI,Elektrotehnicki
fakultet,http://www.etf.bg.ac.yu,http://galeb.etf.bg.ac.yu/~pera,,,,,,,Pera je
mnogo dobar covek,
```

Microsoft Address Book CSV format

Ovaj format se u nekoliko stvari razlikuje od uobičajenog CSV formata. Znak za razdvajanje nije zarez, već *tačka-zarez (semi-colon)*, a značenje i broj odgovarajućih polja zavisi od toga koja su polja izabrana za izvoz u samoj aplikaciji. Zaglavlje postoji i svako polje zaglavlja predstavlja ime odgovarajućeg polja u datoteci. Iza poslednjeg polja ne postoji znak za razdvajanje. Zbog jednostavnosti, biće usvojeno sledeće zaglavlje:

```
First Name;Last Name;Nickname;E-mail Address;Home Street;Home City;Home Postal
Code;Home State;Home Country/Region;Home Phone;Mobile Phone;Personal Web
Page;Business Street;Business City;Business Postal Code;Business State;Business
Country/Region;Business Web Page;Business Phone;Business Fax;Pager;Company;Job
Title;Department;Office Location;Notes
```

Primer kontakta:

```
Petar;Peric;Pera;peraperic@etf.bg.ac.yu;Programskog jezika C
38;Pancevo;26000;Vojvodina;Srbija;013/555-555;064/999-
999;http://galeb.etf.bg.ac.yu/~pera;Bulevar Kralja
Aleksandra;Beograd;11000;Beograd;Srbija;http://www.etf.bg.a.cyu;011/555-
333;011/333-333;060/1112;ETF;Laborant;RTI;26;Pera je los laborant
```

Eudora address book format (.TXT)

Aplikacija za razmenu elektronske pošte Eudora, za skladištenje podataka o kontaktima koristi isključivo tekstualni format datoteka. Format datoteke je sledeći:

```
alias neko_ime e-mail1,e-mail2,...,e-mailN
note neko_ime <polje1:vrednost><polje2:vrednost>...<poljeN:vrednost>Comment
```

Ključna reč alias mora postojati i ona označava jedan kontakt kome je pridružen proizvoljan broj e-mail adresa. Ključna reč note je opciona, i ako postoji označava postojanje dodatnih informacija za već definisani alias *neko_ime*. Dodatne informacije se zadaju u obliku <polje:vrednost> i može ih biti proizvoljan broj i u proizvoljnom redosledu. Komentar vezan za kontakt, ako postoji, se nalazi na kraju sekcije note. Zbog jednostavnosti, relevantna polja će biti samo ona koja su navedena u primeru i po datom redosledu, s tim što, ako za neko polje ne postoji vrednost ono neće biti navedeno u note sekciji.

Primer kontakta:

```
alias Pera pera@pera.yu
note Pera <web2:www.etf.bg.ac.yu><fax2:011/333-333><phone2:011/444-
444><zip2:11000><country2:Srbija><state2:Beograd><city2:Beograd><address2:Bulevar
Kralja Aleksandra><company:Elektrotehnicki
fakultet><title:Laborant><web:http://galeb.etf.bg.ac.yu/~pera><mobile:064/999-
999><phone:013/555-
```

```
555><zip:26000><country:Srbija><state:Vojvodina><city:Pancevo><address:Programsko  
g jezika C><name:Pera Peric><last:Peric><first:Pera>Pera je u Eudori
```

Polja sa istim imenom, označena sa indeksom 2 (kao npr. `web2`), predstavljaju informacije o poslu, pa ih kao takve treba tretirati prilikom konverzije. Više informacija o ovom formatu se može naći na www.eudora.com i na Internetu.

Mozilla Thunderbird LDAP Data Interchange format (.LDIF)

LDAP Data Interchange format (LDIF) je jedan od načina za skladištenje informacija u obliku tzv. direktorijuma definisnih LDAP protokolom. Pod direktorijumom se podrazumeva skup informacija sa sličnim atributima, organizovanih na logičan i hijerarhijski način. LDIF format je način za smeštanje LDAP direktorijuma u tekstualnom formatu. Ovo je iskorišćeno kod nekih aplikacija za razmenu pošte, kao Mozilla Thunderbird, za smeštanje informacija o kontaktima. Direktorijum je u stvari celokupan adresar, a pojedini kontakti predstavljaju zapise u okviru direktorijuma. Svaki zapis se sastoji od grupe atributa (to su informacije vezane za svaki kontakt), a zapisi su međusobno razdvojeni proizvoljnim brojem praznih redova. Pojedinačni atributi u okviru zapisa se smeštaju u uzastopne redove, po formatu *ime:vrednost*. Ukoliko za neki atribut ne postoji vrednost, on se izostavlja. Svaki od zapisa mora sadržati tzv. *Distinguished Name* polje (dn) koje predstavlja jedinstveni identifikator zapisa unutar direktorijuma, i formira se od određenih polja iz zapisa. U slučaju datoteke sa kontaktima, to su polja cn (Common Name) i mail. Prilikom konverzije u LDIF format treba podržati attribute kao u primeru.

Više informacija o ovom formatu se može naći u Request For Comment (RFC) dokumentu 2849, na Wikipedia enciklopediji i na Internetu.

Primer kontakta:

```
dn: cn=Petar Peric,mail=peraperic@etf.bg.ac.yu  
givenName: Petar  
sn: Peric  
cn: Petar Peric  
mozillaNickname: Peca  
mail: peraperic@etf.bg.ac.yu  
mozillaSecondEmail: pera@yahoo.com  
modifytimestamp: 0Z  
telephoneNumber: 011/222-555  
homePhone: 013/777-555  
fax: 011/888-555  
pager: 060/112  
mobile: 064/999-999  
homeStreet: Programskog jezika C  
mozillaHomeStreet2: 38  
mozillaHomeLocalityName: Pancevo  
mozillaHomeState: Vojvodina  
mozillaHomePostalCode: 26000  
mozillaHomeCountryName: Srbija  
street: Bulevar Kralja Aleksandra  
mozillaWorkStreet2: 73  
l: Beograd  
st: Beograd  
postalCode: 11000  
c: Srbija  
title: Dipl. ing.  
department: RTI  
company: Elektrotehnicki fakultet  
mozillaWorkUrl: http://www.etf.bg.ac.yu  
mozillaHomeUrl: http://galeb.etf.bg.ac.yu/~pera  
description: Pera je mnogo dobar covek
```

Internet Mail Consortium vCard format (.VCF)

Format vCard (VersitCard – Electronic Business Card) je format za razmenu personalnih informacija, koji se često koristi kao elektronska vizit karta, koja se može razmenjivati putem e-mail, web-a ili između PDA uređaja i mobilnih telefona. vCard datoteka može sadržati različite informacije o kontaktu, od osnovnih (imena i prezimena), do slika i audio i video zapisa. U zavisnosti od implementacije, može biti zapisan samo jedan ili više kontakata po datoteci. Ako ih je više, zapisi su odvojeni jednom ili više praznih linija. Svaki zapis počinje ključnom reči *begin:vcard*, a završava sa *end:vcard*, između kojih se nalazi proizvoljan broj atributa, imenovanih prema standardu za vCard. U svakom vCard dokumentu su obavezni atributi VERSION, N I FN. U ovom slučaju će se razmatrati verzija 2.1 vCard standarda, pošto je ona u punoj meri podržana od strane MS Address Book aplikacije, koja će se koristiti za testiranje. Atributi se zadaju po formatu *tip:vrednost*. Tip atributa može biti složen, po formatu *osnovni_tip;spec1;spec2*, gde je *osnovni_tip* neki tip, kao npr. adr ili tel, a *spec* predstavljaju parametre tipa, kao što su HOME, WORK, VOICE, FAX. Vrednost tipa mora biti u skladu sa očekivanim formatom, a on je sledeći za relevantne tipove:

- N (name tip) mora imati vrednost po formatu: *Family Name;Given Name;Additional Names;Honorific Prefixes;Honorific Suffixes*
- ORG (organization tip) može imati format: *Organization;Org. Unit1.;Org. Unit2*, ako postoji podatak u kojoj organizacionoj jedinici neke organizacije osoba iz kontakta radi;
- ADR (address tip) mora imati format: *post office box;the extended address;the street address;the locality (e.g., city);the region (e.g., state or province);the postal code;the country name*

Ukoliko za tipove N i ADR nedostaje neki podatak za neko polje (npr. Additional Names, State/Province), ono se ostavlja prazno, uz navođenje zankova tačka-zarez. Ukoliko za neki atribut uopšte ne postoje podaci, on se ne navodi (npr. ne postoji podatak o NICKNAME). Zbog kompatibilnosti sa MS Address Book aplikacijom, poželjno je prilikom konverzije zapis o svakom kontaktu smestiti u zasebnu datoteku.

Više informacija o ovom formatu se može naći u Request For Comment (RFC) dokumentima 2425 i 2426, na Wikipedia enciklopediji i na Internetu.

Primer kontakta:

```
BEGIN:VCARD
VERSION:2.1
N:Peric;Petar;;Dipl. ing.
FN:Petar Peric
NICKNAME:Pera
ORG:ETF;RTI
TITLE:Laborant
NOTE:Pera je los laborant
TEL;WORK;VOICE:011/555-333
TEL;HOME;VOICE:013/555-555
TEL;CELL;VOICE:064/999-999
TEL;PAGER;VOICE:060/1112
TEL;WORK;FAX:011/333-333
ADR;WORK:;26;Bulevar Kralja Aleksandra;Beograd;Beograd;11000;Srbija
ADR;HOME:;;Programskog jezika C 38;Pancevo;Vojvodina;26000;Srbija
URL;HOME:http://galeb.etf.bg.ac.yu/~pera
URL;WORK:http://www.etf.bg.ac.yu
EMAIL;PREF;INTERNET:peraperic@etf.bg.ac.yu
EMAIL;INTERNET:pera@yahoo.com
END:VCARD
```