

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 format (.CSV)	2
Mozilla Thunderbird CSV format.....	3
Tab Separated Values format (.TXT, .TAB ili .TSV).....	3
LDAP Data Interchange format (.LDIF)	4
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* i sl.), kao i velikog broja *web* baziranih servisa za razmenu elektronske pošte (*Yahoo! Mail*, *Gmail*, *MS Hotmail* i sl.).

Neizostavno, sa razvojem ovih programa i širenjem broja korisnika elektronske pošte, razvijali su se i moduli i 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 (binarni) format za skladištenje podataka o kontaktima, tako da njihova razmena između aplikacija različitih proizvođača nije sasvim 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 format (.CSV)

Ovo je jedan od najčešćih formata za razmenu informacija. Nastao je kao format za razmenu informacija između programa za tabelarna izračunavanje. 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 i biće korišćen u domaćem zadatku. Zaglavlje 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,Display Name,Nickname,Primary Email,Secondary Email,Screen
Name,Work Phone,Home Phone,Fax Number,Pager Number,Mobile Number,Home
Address,Home Address 2,Home City,Home State,Home ZipCode,Home Country,Work
Address,Work Address 2,Work City,Work State,Work ZipCode,Work Country,Job
Title,Department,Organization,Web Page 1,Web Page 2,Birth Year,Birth Month,Birth
Day,Custom 1,Custom 2,Custom 3,Custom 4,Notes,
```

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

Primer kontakta:

```
Petar,Peric,Petar Peric,Peca,peraperic@etf.rs,pera@gmail.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://home.etf.rs/~pera,www.panet.rs/~pera,1984,05,1,,,,,Pera je mnogo
dobar covek,
```

Tab Separated Values format (.TXT, .TAB ili .TSV)

Ovo je takođe čest format za razmenu informacija. Za razliku od CSV formata koji za razdvajanje podataka koristi zarez, kod ovog formata se za razdvajanje podataka koristi znak za tabulaciju \t. Znak za tabulaciju se ne sme pojaviti niti u okviru jednog polja. U kontekstu korišćenja ovog formata za razmenu podataka o kontaktima, može se smatrati da sve što važi za CSV format važi i za TSV format.

Zaglavlje postoji u datoteci, a polja su razdvojena tabulacijom. 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 Display Name Nickname Primary Email Secondary Email
Screen Name Work Phone Home Phone Fax Number Pager Number Mobile
Number Home Address Home Address 2 Home City Home State Home ZipCode
Home Country Work Address Work Address 2 Work City Work State Work
ZipCode Work Country Job Title Department Organization Web Page 1 Web
Page 2 Birth Year Birth Month Birth Day Custom 1 Custom 2 Custom 3
Custom 4 Notes
```

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

Više detalja o TSV formatu se može pronaći u IANA *Definition of tab-separated-values (tsv)*, na *Wikipedia* enciklopediji i na Internetu.

Primer kontakta:

```
Petar Peric Petar Peric Peca peraperic@etf.rs 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 Beograd Beograd 11000 Srbija Dipl. ing RTI
Elektrotehnicki fakultet http://home.etf.rs/~pera
http://www.panet.rs/~pera Pera je
mnogo dobar covek
```

LDAP Data Interchange format (.LDIF)

LDAP Data Interchange format (LDIF) je jedan od načina za skladištenje informacija u obliku tzv. direktorijuma definisanih 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. Za potrebe domaćeg zadatka, smatraće se da između dva kontakta postoji samo jedan prazan red. 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, u istom redosledu. Polja *objectclass* navesti istovetno kao u primeru, dok za polje *modifytimestamp* uvek navesti vrednost 0.

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.rs
objectclass: top
objectclass: person
objectclass: organizationalPerson
objectclass: inetOrgPerson
objectclass: mozillaAbPersonAlpha
givenName: Petar
sn: Peric
cn: Petar Peric
mozillaNickname: Peca
mail: peraperic@etf.rs
mozillaSecondEmail: peraperic@gmail.com
modifytimestamp: 0
telephoneNumber: 011/222-555
homePhone: 013/777-555
facsimiletelephonenumber: 011/888-555
pager: 060/112
mobile: 064/999-999
mozillaHomeStreet: Programskog jezika C 38
mozillaHomeLocalityName: Pancevo
mozillaHomeState: Vojvodina
mozillaHomePostalCode: 26000
mozillaHomeCountryName: Srbija
street: Bulevar Kralja Aleksandra
l: Beograd
st: Beograd
postalCode: 11000
c: Srbija
title: Dipl. ing
ou: RTI
o: Elektrotehnicki fakultet
mozillaWorkUrl: http://home.etf.rs/~pera
mozillaHomeUrl: http://www.panet.rs/~pera
description: Pera je mnogo dobar covek
```

vCard format (.VCF)

Format *vCard* (*VersitCard – Electronic Bussines Card*) je format za razmenu personalnih informacija definisan od strane *Internet Mail Consortium*-a. Koristi se često kao elektronska vizit karta, koja se može razmenjivati putem *e-mail*-a, *web*-a ili između pametnih 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 *Mozilla Thunderbird* 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 *Mozilla Thunderbird* aplikacijom, za potrebe ovog zadatka sve zapise o svim kontaktima smeštati u istu datoteku. Zapisi treba da budu odvojeni jednim praznim redom. Prilikom konverzije u *vCard* format treba podržati attribute kao u primeru, u istom redosledu. Svi atributi treba da budu napisani malim slovima.

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
fn:Petar Peric
n:Peric;Petar
org:Elektrotehnicki fakultet;RTI
adr;;;Bulevar Kralja Aleksandra;Beograd;Beograd;11000;Srbija
email;internet:peraperic@etf.rs
title:Dipl. ing
tel;work:011/222-555
tel;fax:011/888-555
tel;pager:060/112
tel;home:013/777-555
tel;cell:064/999-999
note:Pera je mnogo dobar covek
url:http://home.etf.rs/~pera
version:2.1
end:vcard
```