HRVATSKA AKADEMSKA I ISTRAŽIVAČKA MREŽA
CROATIAN ACADEMIC AND RESEARCH NETWORK

Uklanjanje povjerljivih informacija iz MS Word dokumenata

CCERT-PUBDOC-2003-09-42

CARNet CERT u suradnji s LS&S

Sigurnosni problemi u računalnim programima i operativnim sustavima područje je na kojem CARNet CERT kontinuirano radi.

Rezultat toga rada ovaj je dokument koji je nastao suradnjom CARNet CERT-a i LS&S-a, a za koji se nadamo se da će Vam koristiti u poboljšanju sigurnosti Vašeg sustava.

CARNet CERT, www.cert.hr - nacionalno središte za **sigurnost računalnih mreža** i sustava.

LS&S, www.lss.hr - laboratorij za sustave i signale pri Zavodu za elektroničke sustave i obradbu informacija Fakulteta elektrotehnike i računarstva Sveučilišta u Zagrebu.

Sadržaj

1. Uvod

Elektronički dokumenti napisani u MS Word programskom paketu gotovo uvijek, osim samog sadržaja vidljivog na zaslonu računala, sadrže i brojne druge informacije koje vrlo često nisu poželjne za daljnju distribuciju. Radi se o tzv. "*metadata*" podacima koje MS Word programski paket koristi u svojem radu, i od kojih velik broj korisniku nije vidljiv bez posebnih zahvata i alata.

Iako se u ovom slučaju ne radi o klasičnom sigurnosnom propustu koji je posljedica greške u implementaciji programskog koda, već o svojstvu namjerno ugrađenom u MS Word programski paket, kojim su implementirane njegove funkcionalnosti, ovakav način rada u određenim situacijama može predstavljati ozbiljan sigurnosni nedostatak, pogotovo ukoliko se radi o elektroničkoj razmjeni poslovnih ili sličnih drugih dokumenata.

U ovome dokumentu biti će opisan osnovni koncept *metadata* podataka, informacije koje oni otkrivaju te njihov utjecaj na sigurnost i privatnost korisnika. Također će biti opisani i načini na koje se može smanjiti količina *metadata* podataka u MS Word dokumentima. Treba napomenuti da se sličan koncept koristi i kod drugih programa koji dolaze unutar MS Office uredskog paketa (MS Excel, MS PowerPoint, ...) te se ovo razmatranje u velikoj mjeri može prenijeti i na njih.

2. Što su "*metadata*" podaci?

Metadata podaci najbolje se mogu opisati kao podaci o podacima (engl. *data about data*). To su strukturirani opisni zapisi, koji pobliže opisuju neki drugi podatak ili informaciju (sliku, tablicu, katalog i sl.).

Svakako treba napomenuti da koncept *metadata* podataka nije specifičan za Microsoftove proizvode, već da se radi o općenitom pojmu koji se koristi u različitim tehnologijama i djelatnostima (npr. Web stranice također koriste *metadata* podatke kako bi se pobliže opisao njihov sadržaj).

Koncept opisivanja podataka drugim podacima pokazao se vrlo praktičnim i korisnim u brojnim primjenama te je kao takav primijenjen i kod MS Word programskog paketa.

Primjer podataka koje MS Word programski paket u dokumentima pohranjuje u obliku *metadata* podataka su:

- ime i prezime autora,
- inicijali autora,
- ime računala na kojem je uređivan document,
- ime mrežnog poslužitelja na kojem je dokument pohranjen (ukoliko je to slučaj),
- ostala svojstva dokumenta,
- imena i prezimena svih ostalih autora koji su radili na dokumentu,
- verzija dokumenta,
- revizija dokumenta,
- podaci o korištenom predlošku,
- skriveni podaci (engl. hidden text),
- komentari.

Iako se na prvi pogled ne čini kako bi neželjeno otkrivanje nekog od navedenih podataka moglo prouzročiti bilo kakav ozbiljniji incident, dosada je poznato je nekoliko slučajeva u kojima takvi podaci bili korišteni za pokretanje različitih tužbi i pravnih postupaka. U današnje vrijeme, kada privatnost korisnika te povjerljivost i integritet podataka imaju posebno važnu ulogu u području računalnih komunikacija i tehnologija, ovo postaje dodatno naglašen problem.

Dodatni problem predstavlja i činjenica da se većina ovih podataka u dokumente ugrađuje automatski te da korisnici nisu niti svjesni njihovog postojanja. Ukoliko se radi o dokumentima koji se razmjenjuju u okviru važnih poslovnih ili pravnih procesa, informatička pismenost korisnika u ovom pogledu vrlo je važna.

Poznat je i velik broj slučajeva gdje su *metadata* podaci unutar različitih MS Office dokumenata bili iskorišteni u različitim sigurnosnim testiranjima za prikupljanje informacija o organizaciji za koju se testiranje provodi. Prikupljeni podaci, osim što su otkrili prilično zanimljive informacije o samoj organizaciji, također su bili iskorišteni i kao podloga za provođenje tzv. *social engineering* napada, koji su dali vrlo poučne rezultate.

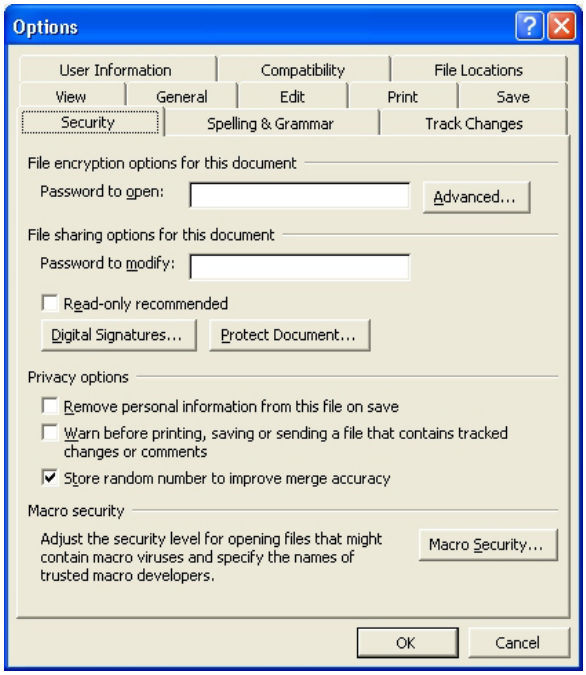
Budući da se različiti *metadata* podaci na različite načine stvaraju u MS Word dokumentima, ne postoji niti jedinstven način na koji ih je moguće sve ukloniti. U nastavku dokumenta biti će opisani neki od podataka koje MS Word programski paket ugrađuje u dokumente te načini na koje ih je moguće ukloniti.

3. Preporuke za uklanjanje "*metadata*" podataka

U ovom poglavlju opisane su neke od metoda kojima je moguće minimizirati broj *metadata* podataka ugrađenih u MS Word dokumente. Neke od navedenih metoda vezane su uz točno određene inačice MS Word programskog paketa te ih je kao takve moguće koristiti samo kod navedenih inačica.

3.1. Automatsko uklanjanje informacija iz MS Word dokumenata

2002 inačica MS Word programskog paketa sadrži opciju koja korisnicima omogućuje automatsko uklanjanje osobnih informacija iz dokumenata prilikom njihovog pohranjivanja. Radi se o "**Remove personal information from this file on save**" opciji, koja je dostupna unutar izbornika **Tools** (*Tools -> Options -> Security*), Slika 1.

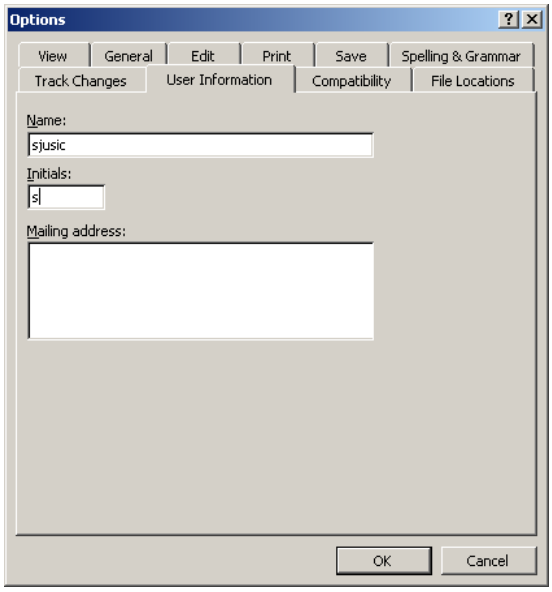


Slika 1: Security sekcija MS Word 2002 programskog paketa

Unutar sekcije **Security** dostupne su i druge opcije koje korisniku omogućuju unaprjeđenje sigurnosnih karakteristika MS Word dokumenata (digitalno potpisivanje, zaštita dokumenata zaporkom i sl.).
Ova opcija dostupna je samo kod MS Word 2002 programskog paketa.

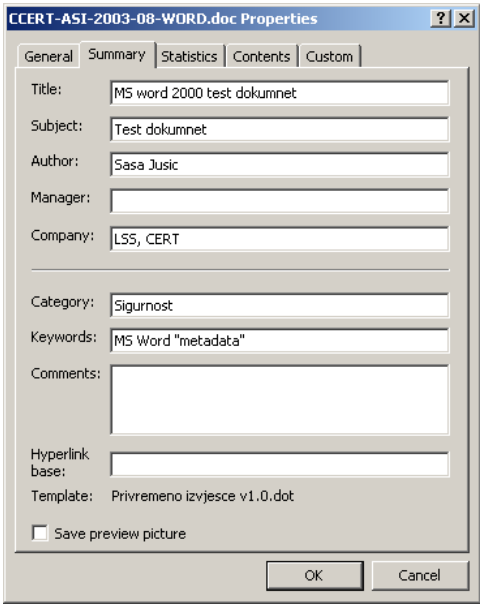
3.2. Ručno uklanjanje *metadata* podataka

Prilikom kreiranja dokumenata u MS Word programskom paketu korisniku je stavljeno na raspolaganje da u dokument upiše različite informacije koje pobliže opisuju karakter samog dokumenta (npr. ime autora, kratki opis, revizija itd.).
Ove informacije moguće je uređivati na nekoliko mjesta unutar dokumenta i korisnik svakako treba biti upoznat sa njihovim postojanjem i mogućnošću uređivanja.
Podatke o autoru dokumenta moguće je uređivati pritiskom na polje **User Information** unutar *Tools -> Options* padajućeg izbornika (Slika 2).



Slika 2: Uređivanje podataka o autoru dokumenta

Detaljnije podatke o dokumentu moguće je pronaći pritiskom na karticu *File -> Properties*, a isto sučelje dostupno je i putem desnog klika miša na sam dokument te pritiskom na polje **Properties**. Napredniji korisnici istim ovim podacima mogu pristupiti putem Microsoft Visual Basic for Applications (VBA) makro funkcija ili nekog drugog sličnog programskog koda sličnih mogućnosti.
Unutar ovoga sučelja dostupne su mnogo detaljnije informacije o dokumentu i ukoliko su iste navedene treba pažljivo razmotriti mogućnost njihovog javnog objavljivanja ili daljnjeg prosljeđivanja poslovnim partnerima ili drugim suradnicima.
Ovi podaci mogu biti posebno nezgodni ukoliko su naslijeđeni iz prethodnih dokumenata. Ukoliko se za novi dokument kao predložak odabere neki od dokumenata iz drugih projekata, ovi podaci mogu sadržavati posve neprikladne informacije te njihova javna objava može biti upitna.



Slika 3: Detaljniji podaci o dokumentu

Kako bi se smanjila mogućnost neželjenog prosljeđivanja povjerljivih informacija putem MS Word dokumenata, prije pohranjivanja svakog novog dokumenta korisnicima se preporučuje ili potpuno uklanjanje ili prikladno modificiranje navedenih podataka.

3.3. Uklanjanje *metadata* podataka kod rada na mreži

U slučaju mrežnog prijavljivanja u sustav (engl. *Network logon*) uklanjanje informacija iz dokumenta može biti nešto kompleksnije, budući da program automatski ubacuje podatke o korisniku prilikom pohranjivanja dokumenta.

Prilikom kreiranja novog dokumenta, MS Word programski paket automatski upisuje podatke o korisničkom imenu pod kojim je korisnik prijavljen u sustav, kao i druge podatke o korisniku.

Kako bi se iz dokumenata u ovom slučaju uklonili automatski generirani *metadata* podaci, korisnik se mora u sustav prijaviti lokalno, ukloniti sve podatke koji se smatraju neprikladnim za daljnje prosljeđivanje te pohraniti dokument. Ukoliko se radi o dokumentu pohranjenom na mrežnom poslužitelju, dokument je prije uređivanja potrebno pohraniti lokalno.

Posebno je važno da korisnik nakon ponovnog mrežnog prijavljivanja u sustav ne otvara pohranjeni dokument, jer će u tom slučaju program automatski upisati ranije spomenute podatke o korisniku.

3.4. Ostale metode

Osim upravo navedenih podataka o autoru dokumenta, kratkom opisu, oznakama revizija i sl., postoje i druge funkcionalnosti MS Word programa čije korištenje može utjecati na ugrađivanje *metadata* podataka. U nastavku je dano nekoliko primjera te načini na koji se mogu ukloniti pripadajući *metadata* podaci.

- **Uklanjanje komentara iz dokumenata**

  MS Word program korisnicima omogućuje dodavanje komentara u dokumente, što se pokazalo kao vrlo praktična funkcionalnost kod višestrukih revizija i ispravaka na dokumentima. Tekst na koji je dodan komentar prepoznatljiv je po tome što je označen žutom bojom i uz njega stoji oznaka komentara.

  S obzirom da su komentari svojstvo programa koje se najčešće koristi u fazi razvoja dokumentacije, prilikom finalnog prosljeđivanja ili javnog objavljivanja dokumenta njih svakako treba ukloniti. Komentari, osim što sadrže podatke o ranijim razvojnim fazama dokumenta, također sadrže i podatke o korisniku koji ih je generirao. Komentare je iz dokumenta moguće ukloniti desnim klikom miša na tekst označen kao komentar te odabirom opcije **Delete Comment**.

  Uklanjanje podataka iz zaglavlja dokumenta te ispravljenih sadržaja

  Sličan problem, kao onaj opisan kod komentara, postoji i kod zaglavlja dokumenta i grešaka ispravljanih pomoću **Track Changes** funkcionalnosti. Površna revizija i nedovoljno obavljene promjene i na ovaj način mogu rezultirati da se zajedno sa dokumentom proslijede povjerljive informacije, koje nisu poželjne za daljnju distribuciju.

  U tom smislu može se preporučiti provođenje temeljitih provjera nad završnom verzijom dokumenta, prije nego što se donese konačna odluka o njegovom prosljeđivanju ili javnoj objavi.

- **Isključivanje Fast Save opcije**

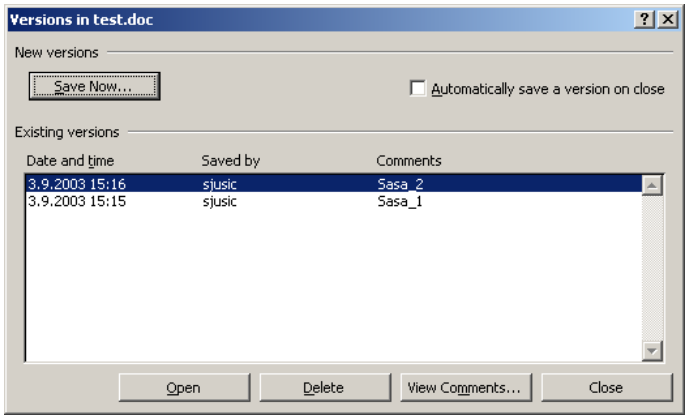
  Fast Save je opcija MS Word programskog paketa koja ubrzava proces pohranjivanja MS Word dokumenata. Ukoliko je prilikom revizije dokumenta ova opcija uključena, pohranjivati će se samo razlike u odnosu na prethodnu verziju. Modificirani ili izbrisani dijelovi dokumenta, iako nisu vidljivi na zaslonu računala, pritom ostaju zapisani u dokument kao nevidljivi tekst (engl. *hidden text*), što predstavlja potencijalni sigurnosni rizik.

  Ukoliko se želi voditi računa o povjerljivosti i vjerodostojnosti dokumenata, prilikom njihovog uređivanja preporučuje se onemogućavanje **Fast Save** opcije.

  Uklanjanje podataka o prethodnim revizijama dokumenta

  Još jedna od karakteristika MS Word programskog paketa, koja olakšava ispravke i revizije dokumenata između više članova tima, je mogućnost pohranjivanja više verzija dokumenata u jednu datoteku. Ovakav način razvoja dokumentacije omogućuje jednostavno pregledavanje ranijih faza razvoja i vrlo je praktičan za praćenje promjena na dokumentu.

  Trenutnu verziju dokumenta moguće je pohraniti pod poljem **Version** unutar padajućeg izbornika File (Slika 4). Uz svaku verziju pohranjuje se radni naziv, korisničko ime korisnika koji ju je pohranio te komentar koji pobliže opisuje inačicu dokumenta.



Slika 4: Pohranjivanje različitih verzija dokumenata

Iako vrlo praktična, ova mogućnost također predstavlja potencijalni sigurnosni rizik kod daljnjeg prosljeđivanja ili objavljivanja dokumenta. Podaci o ranijim fazama projekta mogu sadržavati informacije koje nisu predviđene za objavljivanje.

Iz tog razloga svim se korisnicima koji upotrebljavaju ovu funkcionalnost preporučuje uklanjanje svih podataka o ranijim inačicama prije daljnjeg prosljeđivanja dokumenta.

– **Uklanjanje teksta označenog kao "*hidden text*"**

MS Word programski paket korisnicima nudi mogućnost da određene dijelove teksta označe nevidljivima (opcija "**Hidden text**"). To je moguće postići tako da se željeni tekst selektira te se unutar sučelja za podešavanje fontova (*Format -> Font*) tekst označi skrivenim (*Hidden text*).

Budući da se ovako označeni dijelovi teksta mogu "naslijediti" iz prethodnih dokumenata, korisnicima se svakako preporučuje pretraživanje ovako označenih dijelova dokumenta te njihovo uklanjanje, ovisno o sadržaju i namjeni dokumenta.

Da li će tekst označen kao skriven biti vidljiv ili ne, ovisi o globalnim postavkama programa (*Tools->Optons*). Unutar sekcije **View** nalazi se checkbox polje **Hidden text**, čijim se uključivanjem i isključivanjem može utjecati na prikaz skrivenog teksta.

Skriveni tekst moguće je iz dokumenta ukloniti korištenjem **Find-Replace** sučelja (*Edit-> Replace*). U tu svrhu potrebno je unutar polja **Find what** navesti da se pretražuje skriveni tekst (*More->Format->Font*, **Hidden text** checkbox) te pritisnuti gumb **Replace All**.

Treba napomenuti da su u ovom poglavlju opisani samo neki od podataka koje MS Word programski paket zapisuje u dokumente prilikom njihovog uređivanja. Detaljna analiza svih *metadata* podataka i načina njihovog generiranja zahtijevala bi mnogo više prostora i vremena, čime se izgubio osnovni smisao i ideja dokumenta.

Ovdje opisani elementi iskorišteni su isključivo kao primjer kojim se korisnicima željelo ukazati na prisutnost *metadata* podataka, kao i njihov utjecaj na privatnost korisnika, odnosno organizacije u kojoj je dokument kreiran.

Kao rezultat razmatranja, korisnici MS Word programskog paketa trebali bi biti svjesni da korištenje većine naprednih mogućnosti za sobom povlači ugradnju određenih *metadata* podataka, o kojima posebno treba voditi računa prilikom javnog objavljivanja ili daljnjeg prosljeđivanja dokumenata.

4. Mogućnosti zaštite

Iako se na prvi pogled opisana problematika sa stanovišta sigurnosti ne čini posebno atraktivnom, poznati su brojni slučajevi u kojima je nepažljivo i brzopleto objavljivanje dokumenata rezultiralo otkrivanjem povjerljivih informacija (engl. *Information leakage*). Kao jedan od interesantnijih primjera može se spomenuti dokument koji je engleska vlada javno objavila na svojim Web stranicama, u kojem su pronađene brojne povjerljive informacije o prethodnim revizijama i autorima dokumenta, što je kasnije izazvalo razne probleme i afere. Više informacija o ovom događaju moguće je naći na adresi http://www.computerbytesman.com/privacy/blair.htm.

Također su poznata i istraživanja koja su uključivala pretraživanje javno objavljenih MS Word dokumenata na Internetu te njihovu analizu s obzirom na ovaj problem. Dobiveni rezultati bili su prilično zabrinjavajući, čime se dodatno potaklo pitanje sigurnosti dokumenata napisanih u MS Word

programu (i drugim programima iz MS Office uredskog paketa). Od 100,000 dokumenata koji su prikupljeni na Internetu, svi do jednog su sadržavali određene *metadata* podatke koji su omogućili dolazak do različitih informacija. Dodatne analize pokazale su da je oko polovica prikupljenih dokumenata sadržavala do 50 *metadata* podataka, trećina dokumenata sadržavala je do 500 elemenata, dok je 10% dokumenata sadržavalo preko 500 *metadata* podataka. Većina prikupljenih podataka sadržavala je različite podatke o autorima, organizaciji, revizijama dokumenta i sl., a neki od njih sadržavali su i prilično povjerljive podatke o korisnicima ili poslovnim ugovorima.
Upravo zbog ovog problema, većina svjetskih vladinih organizacija i velikih kompanija je odustala od korištenja MS Word programskog paketa kao službenog programa za distribuciju dokumenata (npr. Engleska).
Na temelju dosadašnjih slučajeva i iskustava vezanih uz ovaj problem, može se zaključiti da okruženja, u kojima se posebna važnost pridaje privatnosti podataka, trebaju ozbiljno razmisliti o alternativnim rješenjima umjesto MS Word programskog paketa, pogotovo kada se radi o razmjeni povjerljivih poslovnih dokumenata.
Jedno od mogućih rješenja je korištenje Acrobat programskog paketa, odnosno PDF (engl. ***P**ortable **D**ata **F**ormat*) formata. Korištenje Acrobat programskog paketa, kao alternative MS Word programu, u velikoj mjeri uklanja opisane probleme, i njegova primjena sve je popularnija i zbog činjenica da su preglednici prisutni gotovo za sve operacijske sustave i platforme.
Dodatna mogućnost, o kojoj svakako treba voditi računa, bez obzira o korištenom programskom paketu, je edukacija korisnika. Korisnicima treba naglasiti važnost i prioritet pojedinih podataka, važnost koju oni predstavljaju za organizaciju, te osmisliti odgovarajuće programe koji će korisnicima omogućiti da svoje znanje primijene u obavljanju svakodnevnih poslovnih zadataka.
U ovom smislu bitno može pridonijeti i donošenje odgovarajućih sigurnosnih politika i procedura, kojima će se definirati pravila ponašanja i odgovornosti pojedinih korisnika.
Kao jednu od mogućnosti treba navesti i korištenje OpenOffice programskog paketa, besplatne alternative MS Office programskom paketu (http://www.openoffice.org/). Program je potpuno besplatan i razvijen je sa primarnim ciljem da zamijeni MS Office programski paket. Popularnost mu vrtoglavo raste i svakim danom sve se veći broj korisnika odlučuje na ovo rješenje.

5. Alati

U sklopu ovog razmatranja treba spomenuti dva programska paketa koja su usko vezana uz tematiku. Radi se o:

- catdoc (http://www.45.free.net/~vitus/ice/catdoc/ver-0.9.html) i
- antiword (http://www.winfield.demon.nl/index.html)

programskim paketima koji omogućuju čitanje MS Word dokumenata na Linux i drugim platformama, a ujedno i omogućuju prikaz podataka ugrađenih u MS Word dokumente. Oba programa potpuno su besplatna i dostupne su verzije za različite operacijske sustave (Linux, DOS, Mac OS...).
Korištenje oba programa vrlo je jednostavno. Program kao argumente prima niz imena MS Word dokumenata koji se žele analizirati, kao i dodatne opcije kojima je moguće preciznije definirati način rada. U nastavku je dan primjer korištenja `antiword` programa (opcija `-s` omogućuje ispisivanje skrivenih podataka):

```
# antiword -s test.doc
      Naslov 1

Ovo je primjer prvog poglavlja.

|Polje 1     |Polje 2     |Polje 3     |Polje 4     |Polje 5     |
|            |            |            |            |            |
|            |            |            |            |            |
|            |            |            |            |            |

1 Naslov 1.1

      Naslov 2
```

```
Ovo je primjer drugog poglavlja.

      Naslov 3

Ovo je primjer trećeg poglavlja.

Ovo je hidden tekst.

Ovo je primjer linka.
```

Za analizu skrivenih *metadata* podataka ugrađenih u MS Word dokumente izvrsno može poslužiti i `strings` naredba na Linux operacijskim sustavima, koja omogućuje ispis znakova iz datoteka različitih formata. `Strings` naredba svoju primjenu vrlo često nalazi u poslovima forenzičke analize, a odlično može poslužiti i u ovu svrhu. U nastavku je dan rezultat izvršavanja `strings` naredbe nad istom `test.doc` datotekom:

```
# strings test.doc
bjbj
 fLjI
K "http://www.lss.hr/"
linka.
Sasa Jusic
 PAGE
 DATE
30.9.2003
PAGE \# "'Page: '#'
'"
 Komentar 1.
sjusic
Normal.dot
Sasa
Microsoft Word 9.0
6eeI
6eeI
LS&S
Title
_PID_HLINKS
_kjI
_kjI
_kjI
Microsoft Word Document
MSWordDoc
Word.Document.8
```

Jasno se može vidjeti da su ovim putem dostupne dodatne informacije o samom dokumentu, koje nisu vidljive u prethodnom primjeru.

6. Zaključak

U ovome dokumentu opisani su sigurnosni problemi vezni uz mogućnost otkrivanja povjerljivih podataka ugrađenih u MS Word dokumente. Opisan je osnovni koncept i smisao *metadata* podataka, dane su preporuke za njihovo uklanjanje te općenite napomene vezane uz ovaj problem. Također su ukratko opisani i neki od alata koji omogućuju analizu ovakvih podataka kao i primjeri njihovog korištenja.