

Josip MIHALJEVIĆ,
mag. educ. inf. & mag. inf.
Institut za hrvatski jezik i jezikoslovlje

ZAŠTITA I UPRAVLJANJE METAPODATCIMA U PDF DOKUMENTIMA¹

SAŽETAK: U radu se proučava PDF format, koji se koristi kao standard za digitalno poslovanje te se sistematiziraju sve inačice toga formata i metapodatci koje pohranjuje. Spominju se i mogućnosti novog PDF 2.0 formata, koji je izišao godine 2017. Autor je također s pomoću programskoga jezika Python izradio program koji omogućuje šifriranje, dodavanje digitalnog žiga te pregled i izmjenu PDF dokumenata kako bi provjerio složenost procesa izrade takvih programa te načina na koje se može zaštititi dokumente. Poglavlje o izradi programa može služiti kao model za daljnju izradu sličnih ili naprednijih programa ili sustava koji obuhvaćaju iste opcije. Autor daje pregled drugih programa koji se mogu koristiti za upravljanje i izmjenu PDF dokumenata

KLJUČNE RIJEČI: metapodatci, mrežni programi, PDF format, Python, šifriranje dokumenata

Uvod

Digitalni dokumenti sve se više koriste u poslovanju zbog porasta informatičko-komunikacijske tehnologije. Svi radovi, potvrde, odluke, molbe te pravilnici računalno se izrađuju zbog urednosti i efikasnosti te se šalju i pohranjuju uglavnom korištenjem interneta. Razvoj računalne tehnologije također je omogućilo elektroničko potpisivanje² i pečačenje dokumenata, što omogućuje da se

¹ Ovaj je rad izrađen na istraživačkome projektu *Hrvatski mrežni rječnik – Mrežnik* (IP -2016--06-2141, <http://ihjj.hr/mreznik/>), koji u cijelosti financira Hrvatska zaklada za znanost. Mrežnik se izrađuje s pomoću programa TLex, ali se sva projektna dokumentacija izrađuje, pohranjuje i šalje Zakladi u PDF-u.

² Elektronički potpisi određuju se kao elektronički skup podataka logički povezanih na datoteku. Taj skup podataka dokazuje pristanak ili neki drugi oblik aktivnosti određene osobe nad tom datotekom. Služi kao zamjena za vlastoručni potpis na fizičkom dokumentu. Elektronički potpis može se dodati na PDF dokument s pomoću različitih metoda koje omogućuju različiti programi poput Adobe Acrobat Pro i Foxit Readera. Digitalizirani ručni potpis ili njegova imitacija može se unižeti na dokument dodavanjem slike skeniranoga rukopisa, unošenjem pina ili čak potpisivanje s pomoću tipkovnice. Elektronički potpis štiti podatke povezane s elektronički potpisanim dokumentom te pomaže u utvrđivanju autentičnosti potpisanog dokumenta. Dhérent, C. (2003). *Elektronički zapisi*. Zagreb: Hrvatski državni arhiv. str. 39-41.

većina formalnih dogovora sklopi elektroničkim putem, a da se potpisnici ne moraju fizički naći. Kod poslovanja na internetu najčešće se koristi PDF format. *Portable Document Format* (PDF) računalni je format datoteke koji je u devedesetima prošloga stoljeća stvorila tvrtka Adobe Systems. Format PDF izrađen je s pomoću jezika PostScript u cilju da se neovisno može učitati na različitim računalima i u različitim programima. Svaki PDF dokument ima u svojem opisu jasno određenu strukturu različitih elemenata poput slika, teksta, fonta i vektorske grafike te drugih tipova informacija koje treba prikazati. Upravo su to razlozi zašto je PDF jedan od najkorištenijih formata za razmjenu dokumenta na mreži kao i činjenica da se može izravno čitati i ispisati s pomoću mrežnoga preglednika (engl. *web browser*). U PDF dokumentima ne može se izravno uređivati tekst, ali se, za razliku od slikovne datoteke formata. JP(E)G, može tekst označiti i kopirati. Kod PDF-a također je moguće stvoriti sadržaj u obliku hiperveze sličan tablici sadržaja koji može, ako se odabere, odvesti korisnika do određenih dijelova dokumenta. Mogu se staviti hiperveze na tekstove i slike. Moguće je i pretraživati tekst PDF dokumenta s pomoću mrežnog preglednika. PDF format ne mijenja se često, pa se tako stariji PDF dokumenti mogu bez problema otvoriti na novim računalima s novijim inačicama PDF preglednika³. Godine 2000. Američki federalni sud izglasalo je da je PDF glavni format za elektroničke dokumente. Godine 2008. izglasano je da je PDF otvoreni format⁴ usklađen s normom ISO 32000. PDF dokument također omogućuje enkripciju, elektroničke potpise, digitalne vodene žigove te detaljne metapodatke za opis. Zbog toga se danas njime koriste različiti servisi poput *e-Gradani*, koji omogućuju državljanima Republike Hrvatske preuzimanje dokumenata poput domovnice, rodnog lista, potvrde o stažu itd. Svaki od tih dokumenata preuzima se kao PDF s elektroničkim potpisom. Mnogi digitalni repozitoriji poput *Digitalnog akademskog arhiva i repozitorija* (DABAR) pohranjuju i nude svoje sadržaje kao PDF dokumente. Na tim primjerima može se vidjeti zašto se PDF koristi kao glavni format elektroničkog dokumenta na mreži.

³ PDF Reference, Sixth edition. (2006). *Adobe Systems Incorporated*. Preuzeto 21. 5. 2018. s: - https://www.adobe.com/content/dam/acom/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf

⁴ Format za pohranu digitalnih podataka kojim se može bilo tko koristiti bez autorskih ograničenja.

Open Format Definition. (2015). *The Open Definition*. Preuzeto 21. 5. 2018. s: <https://opendefinition.org/ofd/>

Metapodatci unutar PDF-a

Metapodatci se često definiraju kao podatci o podacima⁵. Metapodatci daju dodatne informacije vezane za određene entitete poput dokumenata, datoteka, knjiga, predmeta, osoba itd. PDF datoteka može sadržavati metapodatke koji utvrđuju tko je njezin autor, kojeg je dana stvorena, kojeg je dana dokument posljednji put izmijenjen, tko ima autorska prava nad dokumentom te navodi opis sadržaja za dokument, ključne riječi i koji je program iskorišten za stvaranje datoteke.

Description	
File:	Behealthy-dojdiosmas.pdf
Title:	PowerPoint Presentation
Author:	JOE
Subject:	
Keywords:	
Created:	7.5.2018. 23:25:24
Modified:	7.5.2018. 23:25:24
Application:	Microsoft® PowerPoint® 2016
Additional Metadata...	
Advanced	
PDF Producer:	Microsoft® PowerPoint® 2016
PDF Version:	1.7 (Acrobat 8.x)
Location:	D:\be healthy\
File Size:	604,26 KB (618.767 Bytes)
Page Size:	11,00 x 8,50 in
Number of Pages:	2
Tagged PDF:	Yes
Fast Web View:	No

Slika 1. Prikaz osnovnih metapodataka unutar uobičajenog PDF dokumenta koji se mogu pročitati programom Adobe Acrobat Reader

⁵ Panian, Ž. (2005). *Informatički enciklopedijski rječnik: M – Z*. Zagreb: Europapers, str. 21

Pregled metapodataka unutar PDF-a može se napraviti s pomoću programa Adobe Acrobat Reader, ali besplatna inačica tog programa ne omogućuje pregled dodatnih podataka poput onih vezanih za autorsko prava niti se ne mogu izmijeniti osnovni metapodatci poput opisa, naziva dokumenta itd. To je moguće s komercijalnom inačicom programa koja se zove Adobe Acrobat Pro. Metapodatci PDF-a kao što su naziv i opis mogu se koristiti u različitim sustavima za upravljanje sadržajima (CMS – Content Management System) kad treba razvrstati, brisati ili pronaći dokumente koji sadržavaju određene metapodatke. Neki programi poput Zotera⁶ kad učitaju PDF uzimaju njegove metapodatke na temelju kojih stvaraju citat za dokument. Metapodatci za PDF dokumente proširili su se s razvojem novih PDF inačica te određene PDF inačice unutar ISO-ove norme imaju svoje odvojene metapodatke.

Različite inačice PDF

Otkad je 2008. ISO izglasao PDF kao vlastiti standard za pohranu digitalnih podataka, tvrtka Adobe nije smjela napraviti noviju inačicu od PDF 1.7 iz 2006. Adobe je umjesto toga napravio nekoliko vlastitih proširenja za posljednju inačicu programa poput 256 bitne AES enkripcije⁷ i nove XML arhitekture, koje jedino njihovi programi mogu pokrenuti. Nova inačica formata PDF 2.0 pojavila se 2017. s donošenjem ISO 32000–2 norme⁸. PDF 2.0 omogućuje ugradnju interaktivnih sadržaja poput 3D modela (Slika 2), videozapisa i geografskih lokacija. Te opcije još nisu podržane mrežnim preglednicima, ali se mogu pregledati s pomoću programa Adobe Reader.

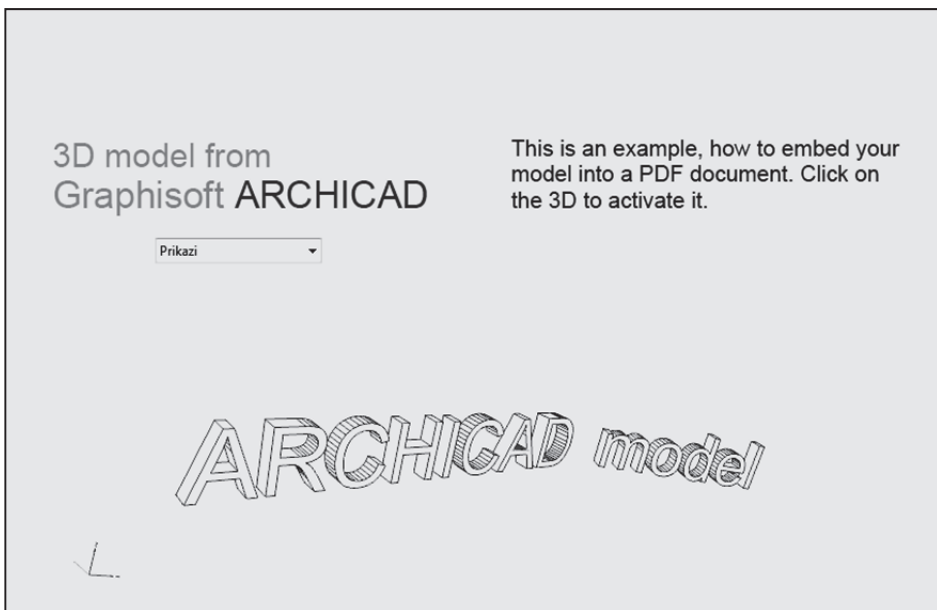
⁶ Zotero - Your personal research assistant. *Zotero*. Preuzeto 21. 7. 2018. s: <https://www.zotero.org/>

⁷ Advanced Encryption Standard (AES) standard je za simetričnu enkripciju, koji je uspostavio američki Nacionalni institut za standarde i tehnologiju godine 2001. Simetrična enkripcija podrazumijeva korištenje istoga ključa za šifriranje i dešifriranje podataka. AES blokovi za enkripciju imaju 128 bita, a ključevi za šifriranje mogu imati 128, 192 ili 256 bita. Djeluje na više razina OSI (engl. *Open System Interconnection*) referentnog modela.

Panian, Ž. (2005). Informatički enciklopedijski rječnik: @-L. Zagreb: Europapress holding d.o.o. str. 21.

⁸ ISO 32000-2 iz 2017. određuje digitalni oblik za prikaz elektroničkih dokumenata koji omogućuje korisnicima razmjenu i pregled elektroničkih dokumenata neovisan o okolini u kojoj su nastali. Norma je namijenjena proizvođačima softvera koji stvaraju PDF dokumente i programe za čitanje PDF-a te obradu i interpretaciju njihova podatkovnoga sadržaja.

ISO 32000-2:2017. *ISO - International Organization for Standardization*. Preuzeto 23. 5. 2018. s: <https://www.iso.org/standard/63534.html>



Slika 2. Primjerna 3D rotacije s mišem u PDF 2.0 dokumentu

Nažalost, u besplatnoj inačici programa Adobe Acrobat Reader nisu omogućene opcije za dodavanje interaktivnih sadržaja, digitalno potpisivanje dokumenata, uspoređivanje dvaju dokumenata, optičko prepoznavanje znakova iz slike (engl. Optical Character Recognition, OCR) i stvaranje bočnog kazala. Neke od tih opcija mogu se iskoristiti u besplatnom programu Foxit Reader za rad s PDF dokumentima⁹. Prije pojave PDF-a 2.0 postojalo je sedam glavnih inačica formata PDF od kojih je šest razvijeno unutar ISO-a, a jednu inačicu razvila je vanjska organizacija. PDF inačice unutar ISO-a su:

- PDF –format koji se najčešće koristi. Koristi se za osnovno uredsko poslovanje te objavljivanje sadržaja na mreži.
- PDF/A (PDF/Archival) – format koji je zamišljen za trajnu pohranu. Sadržaj dokumenta takav je da se svugdje reproducira jednako bez ikakve izmjene neovisno o tehnološkim promjenama. Font dokumenta zbog toga mora biti univerzalan za računala da ne bi ovisio o tome ima li korisnik instaliran font dokumenta na svojem računalu. Šifriranje dokumenta nemoguće je jer bi moglo onemogućiti da se dokument u budućnosti otvori te nije dopuštena ugradnja videozapisa i audiozapisa jer

⁹ Free PDF Reader & PDF Viewer Download. *Foxit Software*. Preuzeto 14. 5. 2018. s: - <https://www.foxitsoftware.com/pdf-reader/>

se njihov sadržaj može neispravno prikazati na drugim računalima. Zbog toga se tim formatom najčešće koriste arhivisti jer time dugoročno mogu autentično očuvati sadržaj originalnoga dokumenta.

- PDF/E (PDF/Engineering) – arhitekti, inženjeri i graditelji često se koriste ovim formatom za razmjenu informacija unutar projekta. Podržava interaktivne sadržaje poput 3D modela i animacija.
- PDF/X (PDF/eXchange) – format koji se koristi za ispis detaljnijih rješenja u području grafičkog dizajna. Omogućuje ispravan ispis za različite fontove, boje i slike.
- PDF/UA (PDF/Universal Accessibility) – format koji je namijenjen proizvođačima softvera radi stvaranja, čitanja, uređivanja i vrednovanja PDF-a čije informacije mogu svi pregledati. UA format određuje skup potreba za pristup PDF dokumentima i programima. On omogućuje korištenje dodatne programske podrške poput čitača teksta, povećala za dijelove dokumenta na ekranu i kontrolera. Ta podrška namijenjena je ljudima koji imaju određena fizička ograničenja kako bi lakše mogli pristupiti informacijama iz dokumenata.
- PDF/VT (PDF/Variable and Transactional printing) – ovaj format PDF-a razvijen je iz PDF/X-a te se njime može koristiti kad je potreban kvalitetan ispis dokumenata u boji. Omogućuje generiranje kvalitetnijih digitalnih slika od PDF/X¹⁰.

PDF inačica izvan ISO-a:

- PDF/H (PDF/Healthcare) – format je dokumenta proizveden za uporabu u zdravstvu. Sadržava standard za stvaranje, prikaz, očuvanje i razmjenu medicinskih podataka. Taj format može pohraniti osobne podatke, XML podatke, medicinske slike poput rendgenskih i ultrazvučnih slika, kliničke podatke te zapise s pregleda¹¹.

Spremanje u neke od tih formata poput PDF/A, PDF/X i PDF/E moguće je s pomoću komercijalnoga programa Adobe Acrobat Pro. Format PDF/UA trenutno podržava malo programa, ali je neke njegove funkcionalnosti moguće iskoristiti u programima poput axesPDF te Acrobat XI Pro¹². Napredne elek-

¹⁰ Willis, T. (2013). 8 Types of PDF Standards – Each Serves a Unique Purpose. *Marco's Blog*. Preuzeto 14. 5. 2018. s: <https://blog.marconet.com/blog/bid/326753/8-types-of-pdf-standards-each-serves-a-unique-purpose>

¹¹ Liang, L. (2016). The PDF/H standard for healthcare. *Foxit Blog - Foxit Software*. Preuzeto 16. 5. 2018. s: <https://www.foxitsoftware.com/blog/the-pdfh-standard-for-healthcare/>

¹² PDF/UA: Creation tools. *PDF Association*. Preuzeto 16. 5. 2018. s: - <https://www.pdfa.org/pdfua-creation-tools/>

troničke potpise poput potpisa PAdES¹³ za dokumente također je moguće stvoriti s pomoću Adobe Acrobat Pro te različitih knjižnica¹⁴ za programski jezik Java. Identifikaciju različitih inačica PDF-a može se provjeriti s pomoću programa DROID, koji može učitati različite dokumente te za njih prepoznati format i kad su posljednji put izmijenjeni.¹⁵ Većina opcija iz različitih PDF formata mogu se ugraditi u dokumente s pomoću programskih knjižnica za jezike poput Pythona i Jave.

U ovome će se radu nadalje analizirati opcije za izmjenu i stvaranje PDF-a koje nudi programski jezik Python¹⁶.

Prikaz mogućnosti zaštite PDF dokumenata te upravljanja njegovim metapodacima na temelju izrađenog programa u Pythonu

Python¹⁷ omogućuje zaštitu PDF dokumenata s pomoću različitih postupaka šifriranja koje mu omogućuje PyPDF2 modul. Osim šifriranja, modul omo-

¹³ PDF s PAdES (PDF Advanced Electronic Signatures) potpisom – godine 2009. Europski telekomunikacijski institut za standarde (engl. European Telecommunications Standard Institute) izdao je novi standard na temelju kojega je izrađen e-potpis koji podržava sigurnu i autentičnu transakciju dokumenata unutar Europe. Podržava napredne elektroničke potpise (eIDAS) određene EU regulativom za elektroničku identifikaciju te transakciju na internetskom tržištu. Basics of PAdES (PDF Advanced Electronic Signatures). *SecureBlackbox*. Preuzeto 16. 5. 2018. s: <https://www.secureblackbox.com/kb/articles/10-BasicsOfPAdES.rst>

Više o tome vidi u Brzica, H.; Herceg, B.; Stančić, H.. *Long-term Preservation of Validity of Electronically Signed Records* // Information Governance / Gilliland, A. ; McKemmish, S. ; Stančić, H.; Seljan, S. ; Lasić-Lazić, J. (ur.). Zagreb : Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2013. 147-158. Preuzeto 15. 7. 2018. s: <http://infoz.ffzg.hr/INFuture/2013/papers/4-03%20Brzica,%20Herceg,%20Stancic,%20LTP%20of%20Validity%20of%20Electronically%20Signed%20Records.pdf>

¹⁴ U programiranju knjižnica je skup kompiliranih rutina ili naredbi koje koristi neki program. Ponekad se te rutine nazivaju i modulima.

Panian, Ž. (2005). Informatički enciklopedijski rječnik: @-L. Zagreb: Europapress holding d.o.o. str. 323-324.

¹⁵ DROID: file format identification tool. *The National Archives*. Preuzeto 1. 11. 2018. S: - <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

¹⁶ Odabran je programski jezik Python jer je jednostavniji za upotrebu od Jave.

¹⁷ Python je proceduralni programski jezik treće generacije koji je nastao 1990. Njegov stvaratelj Guido van Rossum osmislio ga je s namjerom da bude sintaksno jednostavan jezik namijenjen složenom programiranju. Sintaksa Pythona mnogo je jednostavnija od većine programskih jezika treće generacije poput C-a i Jave te zahtijeva manje linija koda za izvršavanje naredbi. Primjena Pythona također je vrlo raznolika jer podržava mnoge programske module

gućuje spajanje i razdvajanje PDF stranica, vađenje sadržaja dokumenata s određenih stranica, postavljanje vodenih žigova te pregled i izmjenu metapodataka. Šifriranje može biti 40 bitna ili 128 bitna te se omogućuje zadavanje više lozinka za pristup dokumentu. Onaj koji poznaje lozinku koja štiti dokument može i trajno maknuti šifriranje iz dokumenta. Za napredniji rad sa šifriranjem može se koristiti PyCrypto modul, koji podržava različite algoritme za šifriranje poput algoritama AES, DES, RSA i ElGamal¹⁸. Moguće je podesiti ograničenja poput pisanja i čitanja dokumenata za obične korisnike koji nisu njihovi vlasnici koristeći se naredbom *chmod*. Ona se koristi za podešavanje ograničenja rada sa zapisima vezanima za njihovo čitanje, pisanje te izvršavanje. Ako se podesi da je zapis dostupan jedino za čitanje, onda nijedan drugi korisnik neće moći izmijeniti sadržaj zapisa, ali može napraviti izmjene te ih spremiti kao kopije originala. S PyPDF2 modulom moguće je postaviti vodeni žig koji mora biti u PDF formatu te prelaziti preko glavnoga teksta. Zbog toga se preporučuje da slika koja se koristi za žig bude dijelom prozirna. Modul može izvući i osnovne metapodatke koji daju informacije s pomoću kojega je programa zapis nastao. Na primjer ako je program skeniran s programom AbbyFineReader, onda se taj program navodi kao alat za izradu, a ako je dokument spremljen kao PDF iz Worda, onda se navodi Office program. Ako napravimo kompresiju PDF dokumenta s nekim ponuđenim alatom ili servisom, katkad se taj program navodi umjesto originalnog programa koji je služio za proizvod PDF datoteke. Moguće je dodati vlastite metapodatke za izrađene kopije originala kojima se može zamijeniti izvorne metapodatke¹⁹.

Kako bi se prikazale spomenute opcije u Pythonu, napravljen je program s grafičkim sučeljem za Windowse koji omogućuje učitavanje PDF dokumenta za koji se može dodati šifra za zaštitu te se mogu izmijeniti metapodatci (Slika 3). Taj program može poslužiti kao razvojni model za buduće programe ili sustav koji žele obuhvatiti osnovne opcije zaštite te upravljanja metapodacima PDF dokumenta. Program s pomoću grafičkog sučelja omogućuje korisniku da učita PDF dokumente te nakon učitavanja dokumenta korisnik može postaviti šifru za dokument, postaviti sliku kao vodeni žig i pročitati te izmijeniti osnovne metapodatke dokumenta. Ako su dokumenti zaštićeni šifrom, za njih se ne mogu pročitati metapodatci. Korisnik se na to upozorava kratkom porukom u skoč-

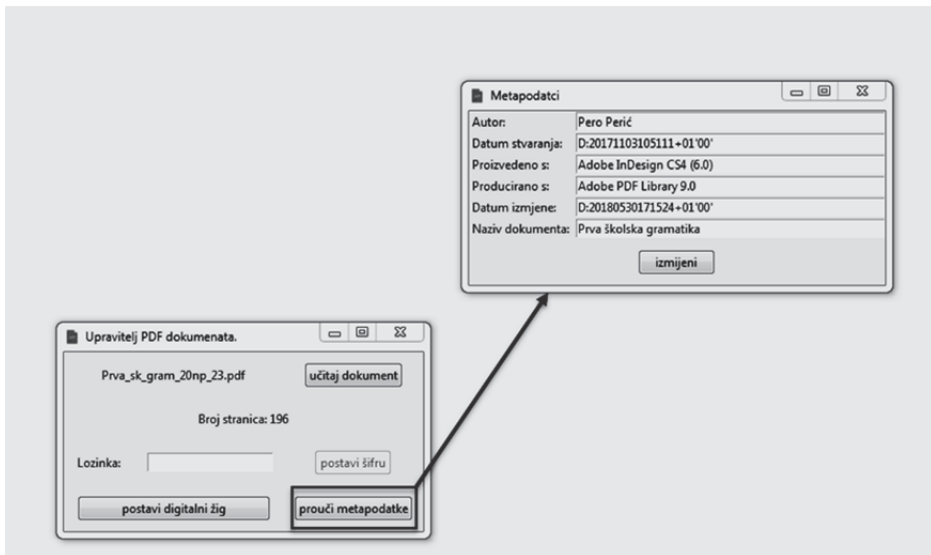
poput Django za izradu mrežnih stranica, Beautiful Soup za parsiranje HTML-a i XML-a, SQLAlchemy za dizajniranje baza podataka, Pygame za izradu 2D videoigara itd.

Downwy, Allen B. (2014). *Naučite Python*. Zagreb: Dobar plan, str. 1-3.

¹⁸ Taj modul još nije podržan za novu Python 3 inačicu, stoga je zamijenjen modulom PyCryptodome koji nudi slične opcije.

¹⁹ PyPDF2 Documentation. *PythonHosted.org*. Preuzeto 11. 4. 2018. s: <https://pythonhosted.org/PyPDF2/>

nom prozorčiću. Korisnik se pita gdje želi spremiti kopiju dokumenta zaštićeno-ga šifrom. Kod izmjene metapodataka mogu se izmijeniti prethodno spomenuta polja poput autora, naziva dokumenta itd. Na primjeru izrade ovoga programa s pomoću jedne Python skripte može se vidjeti da jedna osoba uz osnovno programersko znanje može napraviti jednostavan program za zaštitu digitalnih dokumenata i upravljanje njima. Izrađeni program s razvijenim kodom (PDFmanager.py) moguće je preuzeti kao zip datoteku te isprobati s pomoću javnog repozitorija otvorenog na GitHub stranici (). U direktoriju *program* napravljena je i 32-bitna izvršna datoteka za operacijske sustave Windows, pa nije potrebno imati instaliran Python kako bi se testirao program.

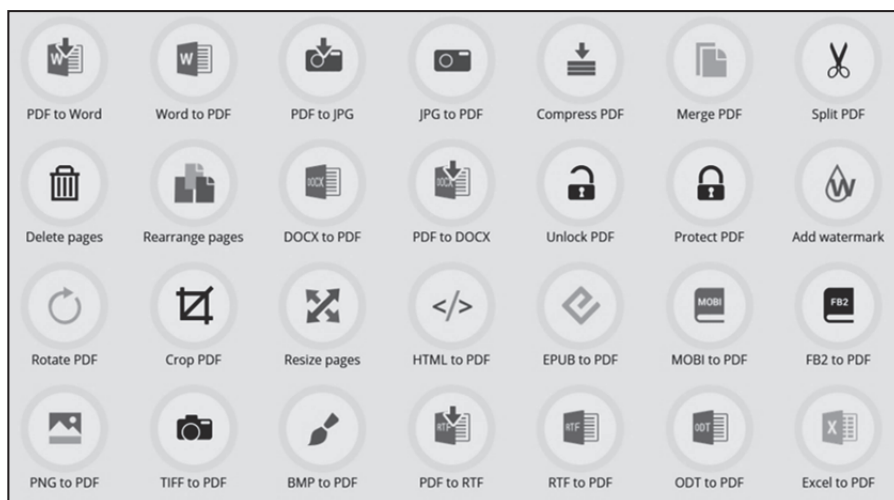


Slika 3. Primjer funkcionalnosti izrađenog programa u Pythonu za upravljanje s PDF dokumentima

Besplatni programi za izmjenu PDF dokumenata

Postoje mnogi besplatni programi koji mogu izvršiti izmjene u PDF dokumentima te poslužiti kao zamjena za komercijalne programe poput Adobe Acrobat Pro. Mnogo tih programa dostupno je na mreži te nude mnogo opcija za rad bez potrebe da se korisnik registrira (Slika 5). Od besplatnih programa koje je potrebno instalirati na računalo (engl. *desktop program*) često se spominju Foxit Reader i Nitro PDF Reader, koji omogućuju zaštitu dokumenata šifrom, dodavanje vodenog žiga te dodavanje elektroničkog potpisa. Nitro PDF Reader²⁰ omogućuje izmjenu i brisanje metapodataka, a Fox Reader to ne omogućuje. Nažalost, besplatna inačica Nitro PDF Reader nakon 14 dana onemogućuje rad s nekim spomenutim opcijama poput opcija za zaštitu i uređivanje dokumenta te program jedino ima osnovne opcije za čitanje, slanje i ispisivanje PDF-a. S pomoću programa LibreOffice Draw unutar besplatnoga programskoga paketa LibreOffice moguće je otvoriti i izmijeniti sadržaje nezaštićenih PDF dokumenata te izmijeniti osnovne metapodatke unutar dokumenta (autor, naziv dokumenta, ključne riječi). Također može se dodati digitalni potpis te zaštitu pristupa dokumentu s pomoću lozinke. Opcije za izmjenu metapodataka PDF dokumenta, njegovu zaštitu, vađenje određenih sadržaja te konverziju u druge formate poput Worda ili. JP(E)G slika uglavnom imaju mrežni programi za upravljanje PDF poput PDF Candy, Smallpdf i iLovePDF. Nedostatak je nekih od tih programa da ne mogu paralelno obraditi više od jednog PDF dokumenta te Small PDF ima ograničenje da može obraditi samo tri dokumenta te nakon toga korisnik treba čekati oko sat vremena kako bi se mogao ponovno služiti programom. Ta ograničenja ne vrijede ako se plati korištenje naprednih opcija. Stoga su ti besplatni programi dobri za kratkoročnu uporabu ako imamo jedan ili dva dokumenta koje treba obraditi, ali nisu dobri kad treba odjednom obraditi više PDF dokumenata.

²⁰ The Best Free PDF Reader and Viewer. *Nitro PDF Reader – Nitro Pro*. Preuzeto 22. 7. 2018. s <https://www.gonitro.com/pdf-reader>



Slika 4. Neke od mogućnosti izmjene učitanoa PDF dokumenta koje nudi mrežna stranica PDF Candy
Zaključak

PDF je i dalje glavni standard za digitalne dokumente na mreži te se redovito unaprjeđuje tako da se dodaju nove mogućnosti zaštite i prikaza sadržaja. Taj je format čitljiv na većini računala. Ima mnogo mrežnih i instalacijskih programa za zaštitu PDF dokumenta. U radu je na primjeru programa izrađenoga u Pythonu pokazano da programe za šifriranje i dešifriranje informacija te izmjenu metapodataka unutar dokumenata nije teško napraviti. Arhivisti bi se više trebali upoznati s opcijama koje PDF formati omogućuju poput elektroničkoga potpisa za provjeru autentičnosti dokumenta, različitih multimedija koji se mogu ubaciti u PDF poput videozapisa ili 3D modela u novom formatu PDF 2.0, različitih metapodataka koji se mogu izmijeniti te koristiti za upravljanje dokumentima u različitim sustavima i programima. Sve te opcije mogu pomoći u poslovanju i arhivskoj praksi radi boljeg prijenosa informacija i organizacije i zaštite dokumenata.

LITERATURA I IZVORI

1. Brzica, H.; Herceg, B.; Katulić, T.; Stančić, H. Analiza utjecaja hrvatskoga zakonodavnog okvira na elektroničko poslovanje i dugoročno očuvanje elektronički potpisanih dokumenata, *Arhivski vjesnik*, 57(2014), str. 130-131.
2. Brzica, H.; Herceg, B.; Stančić, H.. *Long-term Preservation of Validity of Electronically Signed Records // Information Governance / Gilliland, A. ; McKemmish, S. ; Stančić, H.; Seljan, S. ; Lasić-Lazić, J. (ur.)*. Zagreb : Department of Information and Communication Sciences, Faculty of Humanities and Social Sciences, University of Zagreb, 2013. 147-158. Preuzeto 15. 7. 2018. s:
3. Downwy, Allen B. (2014). *Naučite Python*. Zagreb: Dobar plan, str. 1-3.
4. DROID: file format identification tool. *The National Archives*. Preuzeto 1. 11. 2018. S: <http://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>
5. Free PDF Reader & PDF Viewer Download. *Foxit Software*. Preuzeto 14. 5. 2018. s: <https://www.foxitsoftware.com/pdf-reader/>
6. iLovePDF – Online PDF tools for PDF lovers. *iLovePDF*. Preuzeto 12. 4. 2018. s: <https://www.ilovepdf.com/> (12.4.2018.)
7. ISO 32000-2:2017. *ISO - International Organization for Standardization*. Preuzeto 23. 5. 2018. s: <https://www.iso.org/standard/63534.html>
8. Liang, L. (2016). The PDF/H standard for healthcare. *Foxit Blog - Foxit Software*. Preuzeto 16. 5. 2018. s: <https://www.foxitsoftware.com/blog/the-pdfh-standard-for-healthcare/>
9. Ministarstvo uprave. (2014). *Projekt e-Građani-učestala pitanja i odgovori* Zagreb: Vlada Republike Hrvatske. str: 17-18.
10. Open Format Definition. (2015). The Open Definition. Pruzeto 21. 5. 2018. s: <https://opendefinition.org/ofd/>
11. Panian, Ž. (2005). *Informatički enciklopedijski rječnik: @-L*. Zagreb: Europapress holding d.o.o. str. 21, 323-324
12. Panian, Ž. (2005). *Informatički enciklopedijski rječnik: M - Z*. Zagreb: Europapress holding d.o.o. str. 21
13. PDF Candy – Edit PDF free with online PDF editor. *PDF Candy*. Preuzeto 12. 4. 2018. s: <https://pdfcandy.com>

14. PDF Reference, Sixth edition. (2006). *Adobe Systems Incorporated*. Preuzeto 21. 5. 2018. s: https://www.adobe.com/content/dam/acom/en/devnet/acrobat/pdfs/pdf_reference_1-7.pdf
15. PDF/UA: Creation tools. *PDF Association*. Preuzeto 16. 5. 2018. s: <https://www.pdfa.org/pdfua-creation-tools/>
16. PyPDF2 Documentation. *PythonHosted.org* . Preuzeto 11. 4. 2018. s: <https://pythonhosted.org/PyPDF2/>
17. Rouse, M. (2016). HIPAA (Health Insurance Portability and Accountability Act). *TechTarget*. Preuzeto 20 5. 2018. s
18. The Best Free PDF Reader and Viewer. *Nitro PDF Reader – Nitro Pro*. Preuzeto 22. 7. 2018. s <https://www.gonitro.com/pdf-reader>
19. Willis, T. (2013). 8 Types of PDF Standards – Each Serves a Unique Purpose. *Marco's Blog*. Preuzeto 14. 5. 2018. s: <https://blog.marconet.com/blog/bid/326753/8-types-of-pdf-standards-each-serves-a-unique-purpose>
20. Zotero - Your personal research assistant. *Zotero*. Preuzeto 21. 7. 2018. s: <https://www.zotero.org/>

Josip MIHALJEVIĆ

PROTECTING AND MANIPULATING
METADATA IN PDF DOCUMENTS

Summary

The focus of the paper is on the PDF format which is most often used as a digital business standard. All versions of PDF files and metadata they store are analyzed as well as the new PDF 2.0 format which was released in 2017. Using the programming language Python, the author created a PDF management program which can encrypt documents, add watermarks and change files metadata. The explanation on the creation of this program can serve as a simple model for creating similar or advanced programs or systems for manipulating PDF files. The author also gives an overview of other available programs which can be used for manipulating PDF files.

Dr Bojan CVELFAR, direktor
Arhiv Republike Slovenije

USPOSTAVLJANJE SLOVENAČKOG ELEKTRONSKOG ARHIVA

SAŽETAK: Elektronsko arhiviranje provodi se u elektronskom arhivu i sačinjavaju ga sledeći postupci: predaja elektronske arhivske građe (eAG) u nadležne arhive, stručna obrada, dugoročno skladištenje sukladno važećoj legislativi, efikasno upravljanje i upotreba eAG. Prilog osvetljava prva praktička iskustva po pitanju preuzimanja eAG od strane Arhiva Republike Slovenije te same početke razvoja slovenačkog elektronskog arhiva, vezane na studiju izvodivosti, strategiju i plan implementacije razvoja slovenačkog javnog e-arhiva kao i međunarodne standarde (gde treba ispostaviti ISO 14721, koji je inače poznatiji kao referentni model OAIS te ISO 20625, koji detaljnije definiše sam proces predaje arhivske građe). Predstavljena su također i prva važnija dostignuća projekta e-ARH.si, među kojima treba istaći razvoj softverskih alata za neke tipove građe, uspostavljanje adekvatne informatičke sredine na tri međusobno udaljene lokacije, predlog za nacionalni standard digitalizacije arhivske građe (kontrolna lista), uspostavljanje zajedničkog web portala javne arhivske službe... Među strateškim zadacima prvenstvenog značaja daljeg razvoja e-ARH.si u godinama od 2016. i 2020. treba istaći sledeće: brigu za eAG kod stvaralaca i preuzimanje iste u e-ARH.si, brigu za dugoročno očuvanje eAG e-ARH.si, brigu za dalje korišćenje i povećanje dostupnosti arhivske građe, aktivnu saradnju sa drugim srodnim institucijama u zemlji i u inostranstvu; podržavanje stručnog razvoja zaposlenih kod stvaralaca arhivske građe i u arhivima te garantovanje kvaliteta e-ARH.si sukladno modelu CAF.

KLJUČNE REČI: elektronska arhivska građa (eAG), elektronski arhiv, e-ARH.si, digitalizacija

Nekoliko polaznih tačaka za razumevanje sistema e-arhiva

U uvodu priloga upozorio bih na nekolicinu polaznih tačaka za lakše razumevanje sistema elektronskog arhiviranja. Pritom već je u osnovu potrebno striktno razlikovati elektronsko skladištenje od elektronskog arhiviranja.

Elektronsko skladištenje naime provodi se kod stvaralaca građe ili njegovih pravnih sljedbenika te može biti kratkoročno ili dugoročno. Za arhivsku građu važan je ovaj drugi tip, koji u biti znači uskladištenje građe u digitalnoj formi na duži vremenski period te se odnosi na građu, čiji je rok čuvanja duži od pet godina. Slovenačko zakonodavstvo uz to propisuje da takva digitalna građa mora biti konvertovana u adekvatnu formu za dugoročno čuvanje, a sigurno konvertovanje mora sačuvati upotrebljivost i autentičnost sadržaja izvorne građe