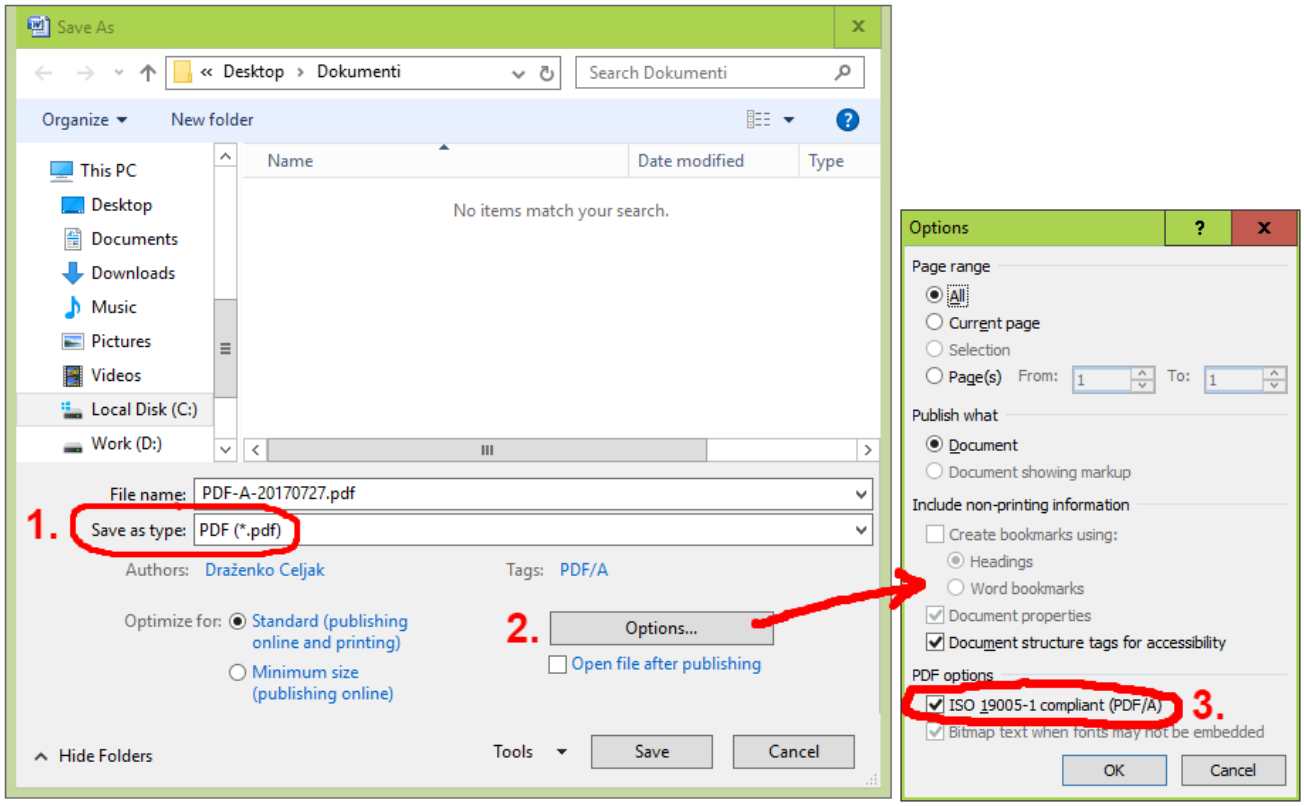
Smjernice za spremanje radova u PDF formatu

- **PDF datoteka treba imati upisane ispravne metapodake** poput autora i naslova jer tražilice i baze podataka te podatke nerijetko koriste za indeksiranje. Metapodaci su u PDF čitaču *Adobe Reader* vidljivi kroz izbornik: *File > Properties*. U programu za obradu teksta *Microsoft Word* se metapodaci uređuju pomoću opcije *Properties*. Upute za ažuriranje metapodataka PDF datoteka pomoću besplatnih alata nalaze se niže na stranici.
- **PDF datoteka ne smije imati uključenu enkripciju** jer to značajno ograničava mogućnost budućih migracija na druge formate ili verzije PDF-a. Iz tog razloga i neke baze (npr. WoS) nisu sklone preuzimanju i korištenju PDF datoteka koje imaju uključenu bilo kakvu enkripciju. U PDF čitaču *Ad obe Reader* postavke sigurnosti vidljive su kroz izbornik: *File > Properties > Security*. Vrijednost opcije *Security Method* mora biti postavljena na *N o Security*.
- **PDF datoteku je potrebno kreirati koristeći PDF/A inačicu PDF formata**. PDF/A inačica namijenjena je za dugoročno čuvanje elektroničkih dokumenata. Specificirana je standardom ISO 19005 i osigurava samodostatnost dokumenta na način da sve vanjske resurse potrebne za reprodukciju dokumenta (npr. fontove) uključi u PDF datoteku kako bi se dokument mogao reproducirati bez obzira na promjene tehnologije i neovisno o platformi. PDF/A ujedno isključuje mogućnost enkripcije te pokretanja programskih skripti i izvršnih datoteka.



Spremanje PDF/A-1 pomoću programa za obradu teksta Microsoft Word 2007.

U većini slučajeva je za kreiranje PDF/A dovoljna verzija PDF/A-1 koja je podržana u programu za obradu teksta Microsoft Word, ali ovisno o verziji PDF/A postoje i neka ograničenja, npr. PDF/A-1 ne dozvoljavaju transparentne slike, pa je ponekad potrebno koristiti više verzije PDF /A (PDF/A-2 ili PDF/A-3) kako bi dokument bio ispravno prikazan.

- **PDF datoteka treba sadržavati tekst rada** kako bi se sadržaj mogao indeskirati, pretraživati, kopirati i rudariti. Kod kreiranja PDF datoteka iz *born digital* formata (npr. doc) uglavnom se tekst dobro prenosi u PDF, ali prisutnost i kvalitetu teksta treba pripaziti kad PDF datoteke nastaju digitalizacijom odnosno skeniranjem tiskanih radova (Smjernice za digitalizaciju tekstualnih sadržaja nalaze se u nastavku ovog teksta). Sadrži li PDF datoteka tekst rada i kakva je kvaliteta tog teksta možete provjeriti tako da pokušate označiti dio teksta u PDF pregledniku i kopirate ga u program za obradu teksta.

Smjernice za digitalizaciju tekstualnih sadržaja

Prilikom skeniranja tekstualnih sadržaja potrebno uključiti opciju OCR (*Optical Character Recognition*) ako ju alat za skeniranje ima ili je za OCR potrebno koristiti neki specijalizirani alat (npr. ABBY FineReader, Tesseract OCR Engine,…).

Proizvođači OCR softvera i literatura obično preporučaju da se pri digitalizaciji skenira s rezolucijom od 300 dpi za fontove veličine 10 točaka i više, a od 400-600 dpi s rezolucijom za fontove veličine 9 točaka i manje.

Zbog arhivske svrhe i pravne valjanosti dokumenta, tekst prepoznat u postupku OCR-a potrebno je spremiti u sloju ispod slike teksta kako bi se sačuvao izvorni izgled dokumenta i istovremeno dobila mogućnost označavanja teksta.

Upotreba alata Abby FineReader – smjernice za naknadno OCR-iranje PDF dokumenata

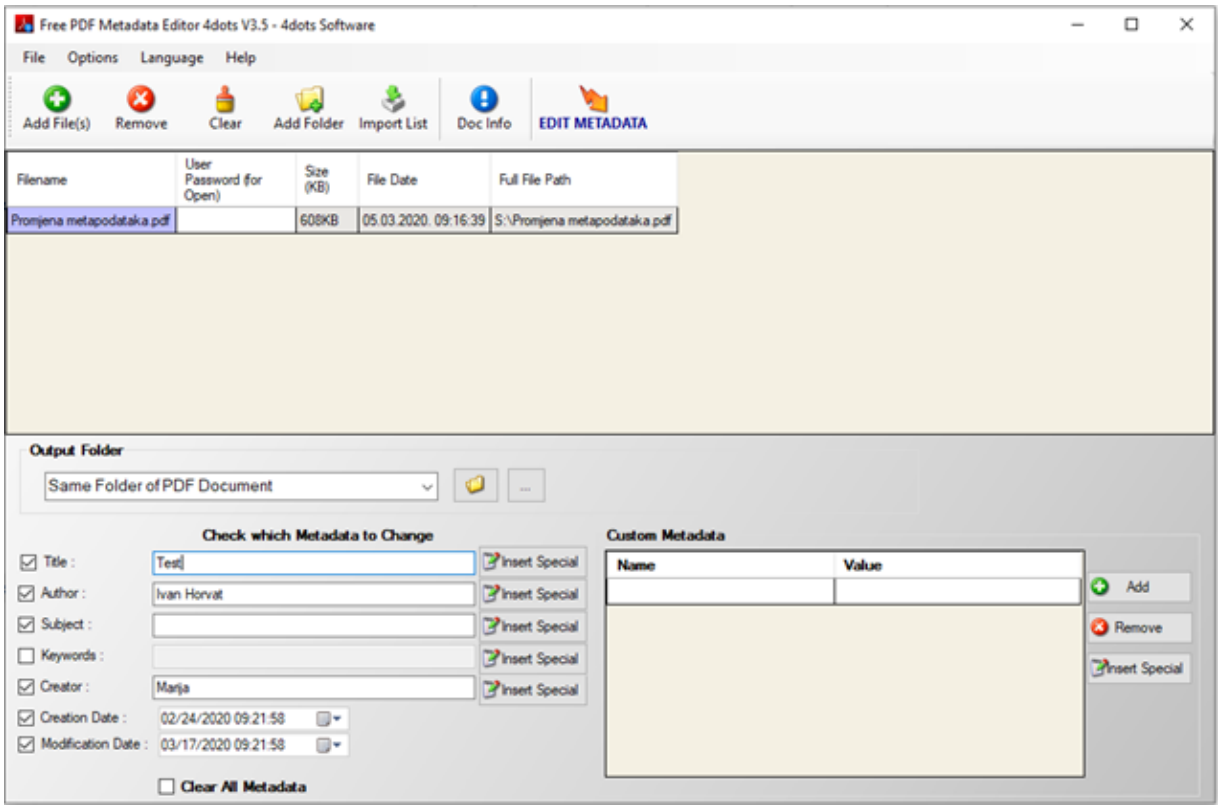
U nastavku se nalazi prikaz koraka za naknadno OCR-iranje PDF dokumenata u alatu AbbyFineReader (komercijalni alat):

- Za naknadno OCR-iranje skenirane PDF datoteke u alatu Abby FineReader (PDF pohranjen kao slika, ne kao tekst), klikom na **Tools** u alatnoj traci potrebno je odabrati **OCR Editor**.
- Otvara se novi prozor za izradu projekta.
- Ponuđena je mogućnost odabira jezika koje(g) će alat koristiti za prepoznavanje/OCR-iranje te mogućnost automatskog prepoznavanja jezika prilikom OCR-a.
- Za otvaranje PDF datoteke unutar prozora za izradu projekta, moguće je koristiti metodu **drag and drop** pri čemu će alat automatski početi s optičkim prepoznavanjem stranica PDF-a, odnosno s postupkom OCR-a.
- Po završetku OCR-a, prikazat će se obavijest da je proces završen.
- Moguće je pregledati svaki dio PDF datoteke pri čemu korisnik može provjeriti je li alat ispravno prepoznao dijelove PDF-a; je li tekstualni sadržaj prepoznao kao tekst, slikovni sadržaj kao sliku, tablične podatke kao tablicu itd.
- Po završetku projekta, potrebno je odabrati opciju **Save** – **Save as Searchable PDF Document**.
- Prikazat će se prozor s odabirom gumba **Options**. Klikom na **Options**, bit će ponuđena mogućnost spremanja PDF/A inačice dokumenta.

Besplatni alati za ažuriranje metapodataka PDF datoteka

Ažuriranje metapodataka PDF datoteka moguće je pomoću različitih programa, npr. *PDF Metadata Editor*, *LibreOffice Draw* i drugi.

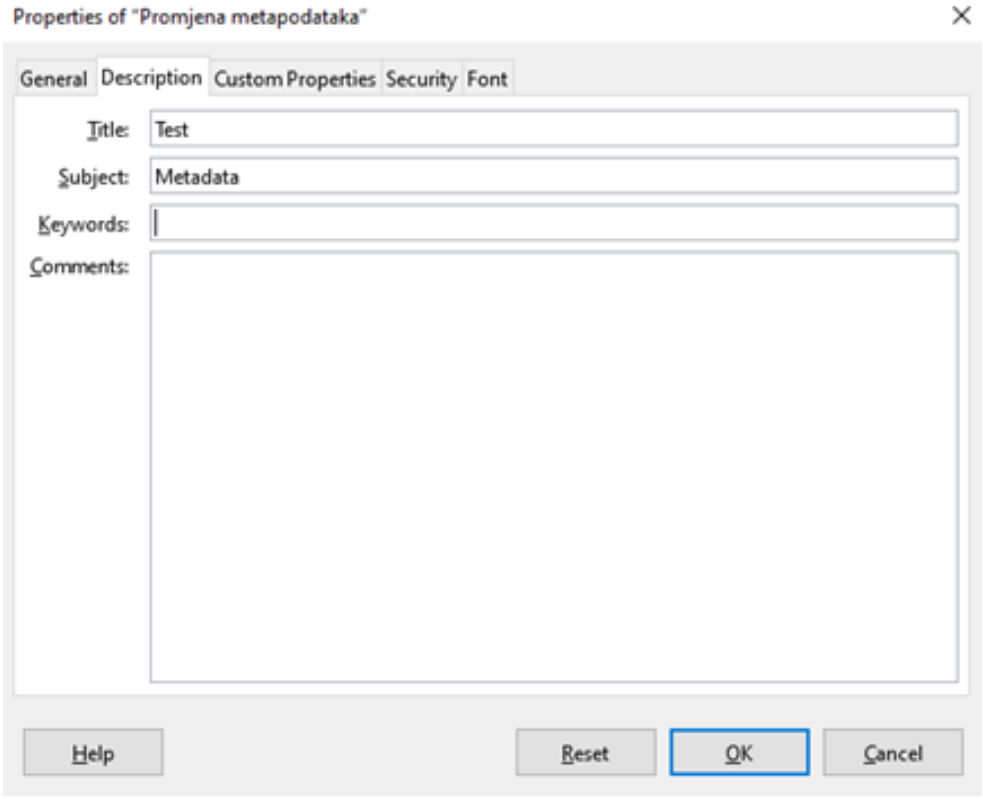
- *PDF Metadata Editor*
  - *Add File(s)* - dodajte datoteku koju želite urediti.
  - Odabrati folder u koji želimo spremiti pdf datoteku.
  - Dolje označiti polja koja želite urediti: npr. *Title*, *Author*… te upisati željene vrijednosti.
  - Kliknuti *EDIT METADATA* te spremiti PDF datoteku.



Uređivanje metapodataka pomoću programa PDF Metadata Editor.

- *LibreOffice Draw*

- Učitati PDF datoteku za koju želimo ažurirati metapodatke.
- Kliknuti *File* => *Properties…* => *Description*
- Upisati željene vrijednosti u polja koja želimo urediti.
- Kliknuti *OK* te spremiti PDF datoteku.
- Moguće je ažurirati i podatke o korisniku. Potrebno je kliknuti na *Tools* => *Options* => *User Data* te upisati željene vrijednosti.



Uređivanje metapodataka pomoću programa LibreOffice Draw.