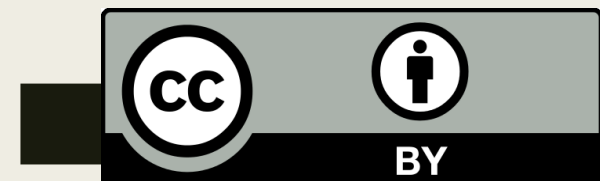


IZRADA FUNKCIONALNE PDF DATOTEKE: 1. DIO

izv. prof. dr. sc. Jadranka Stojanovski
Sveučilište u Zadru / Institut Ruđer Bošković

Webinar Hrvatske udruge za znanstvenu
komunikaciju ZNAK
Zagreb, 11.06.2021.



Agenda

1. dio

- PDF kao format – prednosti i nedostaci
- PDF u znanstvenoj komunikaciji – uloga
- primjeri hrvatskih časopisa

2. dio (Marko Ban)

- tehničke značajke
- korisni alati

Što je PDF?



- PDF je kratica za Portable Document Format
- koristi se za prikaz dokumenata u digitalnom obliku neovisno o softveru, hardveru ili operativnom sustavu – identični izgled!
- izvorno razvijen od strane Adobe® Systems 1991.g. kao univerzalni kompatibilni format datoteke zasnovan na formatu PostScript, vremenom je postao međunarodni standard za razmjenu dokumenata
- u osnovi namijenjen tiskanju i dizajnu
- 2008. godine Adobe je prepustio kontrolu razvoja PDF-a ISO-u (Međunarodnoj organizaciji za standardizaciju) i time je PDF postao je „otvoreni standard“
- specifikacije za trenutnu PDF verziju (2.0) dokumentirane su pod ISO 32000-2
- ISO je također zadužen za ažuriranje i razvoj budućih verzija

Zašto?

PDF Association

2016: 2.5 trilijuna PDF dokumenata (x4?)

<https://www.pdfa.org/>

Kako nastaje PDF dokument?

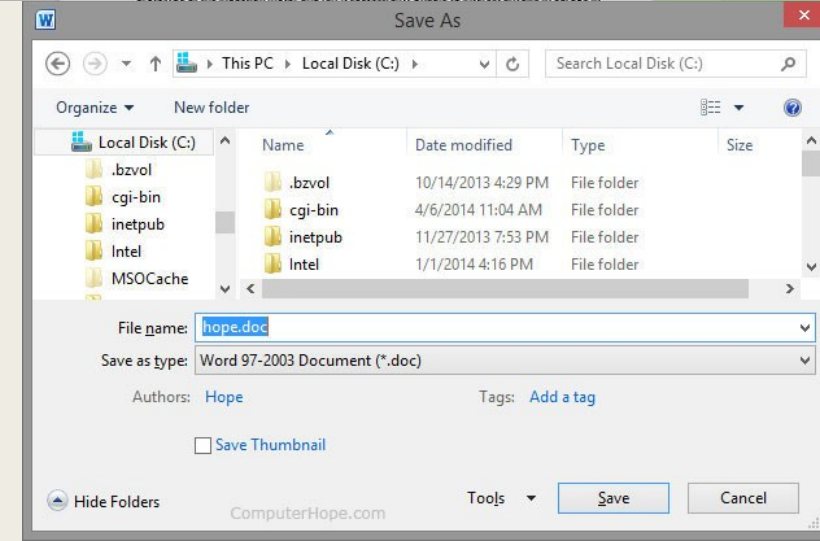
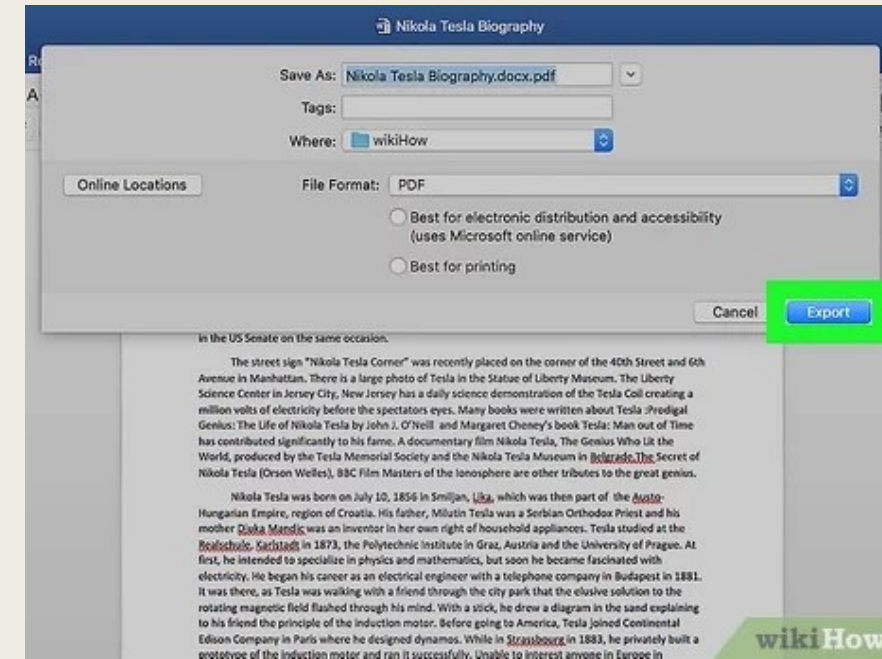
1. „Pravi” ili digitalno stvoreni PDF dokument

Save As, Export is MS Word-a

Adobe PDFMaker (komercijalni alat)

PrimoPDF ili slični besplatni alat

Print to PDF (odabir virtualnog „pisača”)



Kako nastaje PDF dokument?

2. „Slikovni”, skenirani, digitalizirani (tiskani) dokument
skeniranje

konverzija slikovnog (grafičkog) formata (JPEG, TIFF...)

needitabilni

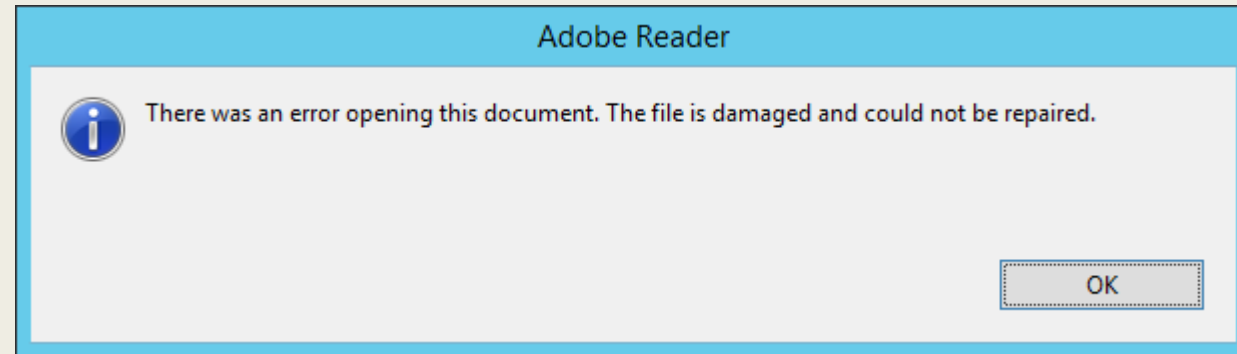
postaju editabilni tek uz primjenu OCR softvera (različite vrste)



Prednosti

- popularnost formata - PDF se može otvoriti na bilo kojem uređaju s bilo kojim operativnim sustavom u istom obliku u kojem je stvoren (ipak nespretno za čitanje na manjim ekranima)
- relativno lagano se otvara - Adobe Acrobat Reader (jedini besplatni Adobe alat!) lagano se instalira (ugradnja u web prebirkike)
- zauzima relativno malo memorijskog prostora - algoritmi za kompresiju (više od HTML-a)
- onaj tko ga izrađuje može prilagoditi sigurnosne postavke, na primjer, zabraniti ispis ili uređivanje, regulirati razine pristupa (vodeni žigovi, lozinke)
- omogućuje upotrebu digitalnih potpisa za pristup ili utvrđivanje vjerodostojnosti dokumenta
- otporan na viruse
- omogućuje integraciju različitih vrsta sadržaja u jednoj PDF datoteci - teksta, slika i vektorske grafike, videozapisa, animacija, audio datoteka, 3D modela, interaktivnih polja, hiperveza i dr.
- pogodan za obrasce (?)

Nedostaci



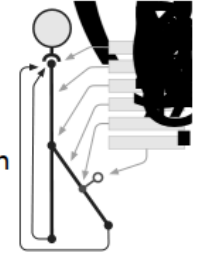
- u fokusu je izgled dokumenta, a ne korištenje teksta (sadržaja)
- datoteke je teško ili nemoguće uređivati (editirati) bez korištenja komercijalnih softvera
- čak i kad je PDF editabilan, teško je ili nemoguće selektirati upotrebljivi segment teksta
- teško je ili nemoguće je konverziju napraviti „unatrag” (jednosmjerna ulica)
- prvenstveno za Windows i Mac
- neodgovarajući format za dugoročno arhiviranje teksta
- različite vrste PDF-a zahtijevaju različite načine rada s njima, na primjer prilikom pretraživanja ili izdvajanja podataka

Nedostaci (nast.)

PDF unreadable rectangular text - graphics ok.



Quimplementatio... extend...
 onlythe side...
 exit...
 atrace...
 wantocoe...



3.3 Blacklisting

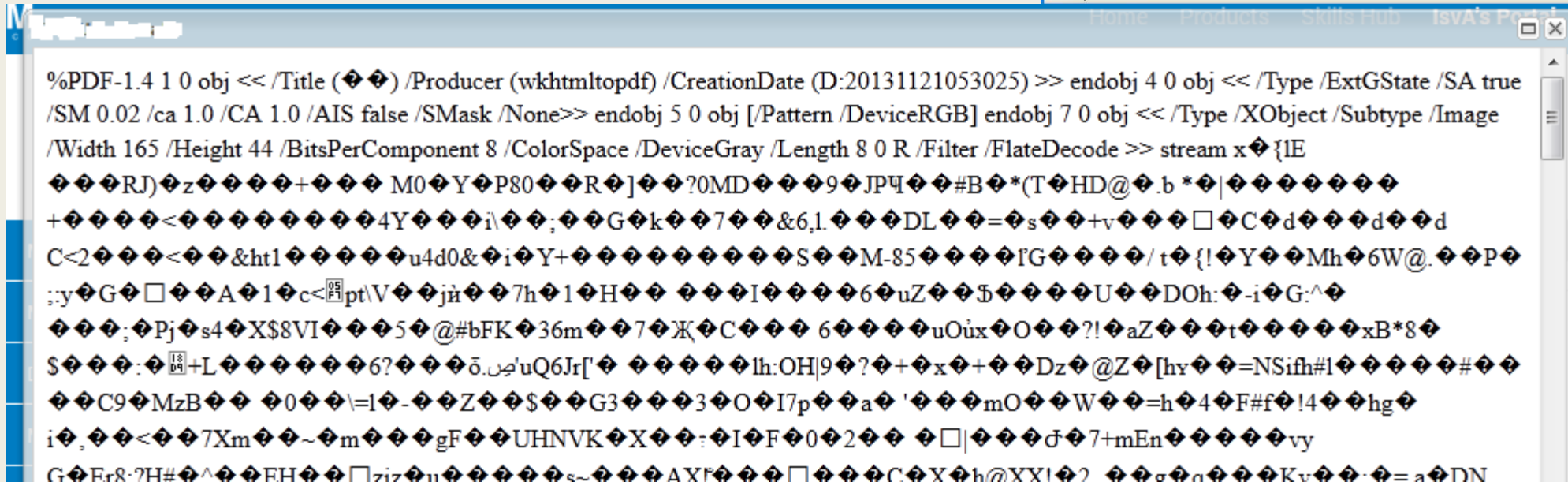
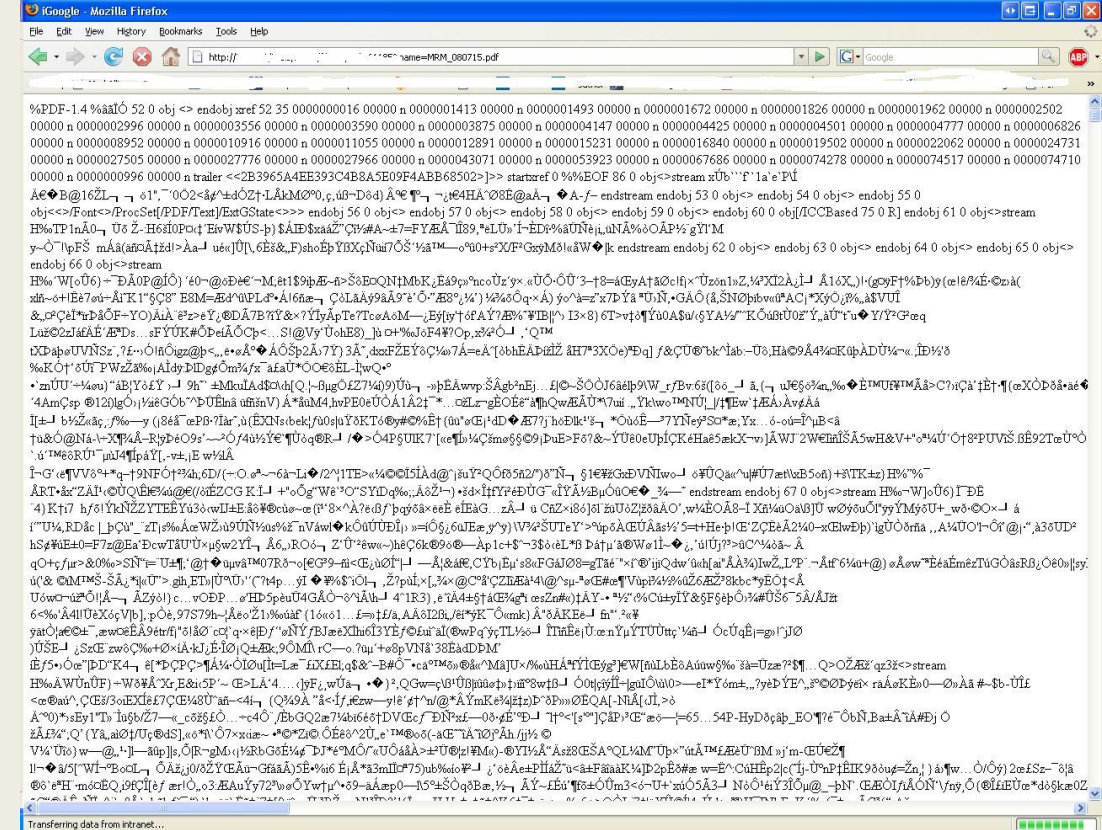
Sometimes...
 intatrace...
 TraceMe...
 catching...
 because...
 program...
 includ...
 atrace...
 trace...
 Aresult...

As a result...
 couldpot...
 the/Me...
 ablet...
 tosta...
 Oupred...
 haeb...
 inga...
 Vha...
 atha...
 menta...
 pre...
 differ...
 teag...
 mark...
 ag...
 Aft...
 sma...
 amou...
 itha...
 byte...
 head...
 pre...
 simp...
 the...
 whe...
 •The...
 •The...
 •The...
 Inth...
 foa...
 Wth...
 trac...
 bas...
 imple...
 But...
 Afut...
 black...
 byte...



Figure...
 complet...
 mat...
 tos...
 partic...
 cal...
 for...
 The...
 imple...
 But...
 Afut...
 black...
 byte...

Nedostaci (nast.)



PDF/A

- PDF/A je ISO standard za arhiviranje digitalnih dokumenata u PDF formatu. Prvi dio, PDF/A-1, objavljen je 1. listopada 2005. Drugi i treći dio, PDF/A-2 (2011) i PDF/A-3 (2012), poboljšavaju PDF/A uz podršku za dodatne značajke u PDF formatu, uključujući opcije za kombiniranje nekoliko PDF/A datoteka u jednu PDF/A zbirku ili ugrađivanje izvornih datoteka PDF-a ili drugih podataka, podršku za transparentnost i još mnogo toga.
- Stvaranje višedijelnog PDF/A standarda zahtijevalo je dugogodišnju suradnju, od programera softvera, udruga i državnih agencija.
- ISO 19005 definira „format datoteke koji se temelji na PDF-u, poznat kao PDF/A, a koji pruža mehanizam za prikaz digitalnih dokumenata na takav način da se vizualna slika održava tijekom vremena, bez obzira na alate i sustave koji se koriste za njihovu proizvodnju, pohranu i reprodukciju
- PDF/A ne definira strategiju arhiviranja; standard jednostavno navodi tehničke zahtjeve za PDF digitalne dokumente kako bi se osigurala pouzdanost desetljećima ili čak stoljećima nakon stvaranja datoteke.
- Od 2005. PDF/A je postao dobro uspostavljen kao preferirani format za arhiviranje digitalnih dokumenata. PDF format, koji je 2008. godine standardiziran kao ISO 32000-1, koristi se širom svijeta u javnom i privatnom sektoru i nakon HTML-a ostaje najpopularniji mrežni format.
- PDF/A osigurava pouzdan pristup PDF dokumentima, što PDF čini idealnim za dugoročno arhiviranje.

<https://www.pdfa.org/>

PDF U KONTEKSTU ZNANSTVENOG IZDAVAŠTVA

...i znanstvene komunikacije

Kako izgledaju publikacije danas?

- časopisi i knjige kroz PDF-ove oponašaju tiskane inačice, čak i kada one ne postoje
 - statičnost i fiksiranost PDF-a kao formata utječu na čitatelje (sigurnost? povjerenje?), održavanje privida znanstvene valjanosti putem složenog *layout*-a?
 - veliki izdavači i („poželjna”) ograničenja PDF-a kao formata, nemogućnost korištenja i samoarhiviranja „izdavačke” verzije rada
-
- Koliko je to u funkciji kvalitetne znanstvene komunikacije?

Zahtjevi Plana S – Obavezni tehnički uvjeti za sva mjesta izdavanja

- Korištenje trajnih identifikatora (PID) za znanstvene publikacije (s verzijama, na primjer, u slučaju revizija), kao što je DOI (poželjno), URN ili Handle.
- Pohrana sadržaja pomoću programa za dugoročnog digitalno čuvanje ili arhiviranje (kao što su CLOCKSS, Portico ili ekvivalent).
- Visokokvalitetni metapodaci na razini članka u standardnom interoperabilnom ne-vlasničkom formatu, pod CC0 namjenom javne domene. Metapodaci moraju sadržavati cjelovite i pouzdane informacije o financiranju koje pružaju financijeri cOAlition S (uključujući najmanje ime davatelja sredstava i broj projekta/identifikator).
- Strojno čitljive informacije o statusu otvorenog pristupa i licenci ugrađenoj u članak, u standardnom ne-vlasničkom formatu.

Zahtjevi Plana S – Preporučeni dodatni kriteriji za sva mjesta izdavanja

- Podrška PID-ovima za autore (npr. ORCID), financijere, programe financiranja i projekte, institucije i druge relevantne subjekte.
- Registriranje politike samoarhiviranja mjesta izdavanja u SHERPA / RoMEO.
- Dostupnost za preuzimanje cjelovitog teksta za sve publikacije (uključujući dopunske tekstove i istraživačke podatke) u strojno čitljivom standardnom formatu zajednice kao što je JATS XML.
- Izdavač izravno pohranjuje publikacije (u strojno čitljivom standardnom formatu zajednice kao što je JATS XML i uključujući cjelovite metapodatke kako je gore opisano) u odabrane od autora ili centralizirane repozitorije otvorenog pristupa koji ispunjavaju kriterije Plana S.
- Sukladnost metapodataka s OpenAIRE-om.
- Povezivanje s podacima, kodom i drugim rezultatima istraživanja koji su temelj publikacije i dostupni su u vanjskim repozitorijima.
- Otvoreno dostupni podaci o citatima prema standardima Inicijative za otvorene citate (I4OC).

ANALIZA MANJEG BROJA RADOVA IZ HRVATSKIH ČASOPISA

Nazivi PDF datoteka

Izbjegavati:

- 1.pdf
- rad1.pdf
- 06Medic.pdf

Mogućnosti:

- [10.1016@j.lisr.2006.06.002.pdf](#) (DOI rada!)
- 10-1108_JD-05-2018-0071.pdf (DOI rada!)
- 0001-3765-aabc-201820170959.pdf (ISSN časopisa, u slučaju da časopis nema integriran DOI)
- Archambault2006_Article_BenchmarkingScientificOutputIn.pdf („deskriptivni” oblik, ako ništa od prethodnog nije primjenjivo)

Nedostaje informacija o časopisu

- ovaj primjer je prisutan kod velikog broja časopisa/PDF-ova na Hrčku
- informacija o časopisu je nužna kako bi se članak, do kojeg korisnik najčešće dolazi putem Google Search-a, mogao povezati s časopisom u kojem je objavljen
- nerijetko nije prisutan broj stranica, a nema niti „*article numbers*” koji zamijenjuju brojeve stranica (npr. eLocator, article ID i sl.)

(pre)Složeni *layout*

- ako časopis smatra da će formatiranje rada u tri stupca i sl. biti korisno za čitatelje i pridonijeti preglednosti i čitljivosti rada, to je u redu
- no, napor i vrijeme koje se ulaže u složeni *layout* možda bi bilo korisnije utrošiti na jasno autorstvo (uz afilijaciju autora), numerirane i ujednačene potpise uz grafike i tablice i druge korisne informacije.

Needitabilni PDF

- najčešće rezultat procesa skeniranja starijih godišta, bez korištenja OCR-a
- nemoguće je računalno selektirati tekst
- za ovakve je PDF-ove posebno važno imati ugrađene kvalitetne metapodatke!

Učitavanje radova u softvere za upravljanje referencama

Mendeley – probni import malog uzorka radova

- radovi iz nekih časopisa se učitavaju sa svim potrebnim metapodacima: autor, naslov, časopis, godina, volumen, sveščić, broj stranica, sažetak, ključne riječi, ISSN...
- za neke radove učitava se krivi naziv časopisa (?) ili se naziv časopisa uopće ne učitava
- za neke radove učitava se krivi autor (npr. „Nizozemske, Austrije” ili „Informatike, I” ili su hrvatski znakovi pogrešno kodirani)
- za neke radove nije se učitao niti jedan metapodatak

Google Search

- provjerite kako izgledaju rezultati pretrage vašeg časopisa na Google Search
- prikazuju li se ispravno osnovni podaci o naslovu, autoru i godini objavljivanja?
- provjerite i rezultate pretraživanja na Google Scholar-u



site:hrcak.srce.hr filetype:pdf



[All](#) [Images](#) [News](#) [Maps](#) [More](#) [Tools](#)

About 324,000 results (0.40 seconds)

<https://hrcak.srce.hr> > file [PDF](#)

[Popisi i oprema udavača \(dotali\) iz 18. i početka 19. st.\(Građa ...](#)

by R Matejčić · 1968 — U toku istraživanja građe za povijest (nošnje Rijeke i njene okolice uspjelo nam je pronaći u Historijskom arhivu u Rijeci veći broj do- talnih isprava u fondu: ...

<https://hrcak.srce.hr> > file [PDF](#)

[RANKINGS SCIENTISTS, JOURNALS AND COUNTRIES ...](#)

by G Mester · 2016 · Cited by 65 — Indexes in scientometrics are based on citations. However, in contrast to the journal impact factor, which gives only the ranking of the scientific ...

<https://hrcak.srce.hr> > file [PDF](#)

[Seizmološke i seizmotektonske značajke šireg zagrebačkog ...](#)

by V Kuk · 2000 · Cited by 15 — Seizmološke i seizmotektonske značajke šireg zagrebačkog područja. Prikazani su učestalost i prostorna distribucija pojavljivanja potresa na širem ...

<https://hrcak.srce.hr> > file [PDF](#)

[Izvod iz strogo povjerljive kombinatorike Velimira ... - CORE](#)

by LJB BOBAN · Cited by 1 — Na sredini listopada 1983. urednicima Časopisa za suvremenu povijest na javio sam da ću za naredni broj toga časopisa pripremiti osvrt na drugo.

<https://hrcak.srce.hr> > file [PDF](#)

[hrvatski, pdf \(372 KB\)](#)

by V Valjan Vukić · 2013 — pristupa u Danskoj, na koju su utjecali djelujući na njega ili operirajući njime. brojnih svjetski poznati teoretičari Četvrto načelo teorijsku podlogu nalazi i teorijski ...



Google search



site:hrcak.srce.hr filetype:pdf



https://hrcak.srce.hr > file PDF

→ Dr. BOŽIDAR ŠIROLA: STARA HRVATSKA MUZIČKA ...

by B Širola · 1925 — Doista, davna naša muzička prošlost izaziva spoznaju goleme praznine i nedostajanja pravog umjetničkog pregnuća. SvijeHom u:wru du- brovačkih ...

https://hrcak.srce.hr > file

→ Untitled

by V Sepčić · 1967 · Cited by 1 — work, though I'm sure it has points, and I don't think it retrograde from. The White Peacock. It surprises me by its steady progressiveness – I hate it for its fluid, ...

https://hrcak.srce.hr > file PDF

PTSP I GRUPNA PSIHOTERAPIJA – ISKUSTVA ...

by A CIGROVSKI · 2014 · Cited by 3 — Ljudska bića su orijentirana na život u grupi, koji i počinje u maloj grupi zvanj obitelj. Svaka grupa mala ili velika, obiteljska ili državna, sastoji se od mnoštva ...

https://hrcak.srce.hr > file PDF

→ 100 KB

by T Žigmanov · 2009 · Cited by 3 — Sažetak. Cilj je radnje prezentirati monografske publikacije o Hrvatima u Vojvodini u razdoblju od 1990. do kraja 2008. godine. Radnja donosi popis od 137 ...

https://hrcak.srce.hr > file PDF

O toleranciji kroz povijest1 →

by S Bertoša · 2014 · Cited by 5 — UDK 316.647.5(091). -----
----- . Analizirajući najprije značenje pojma „tolerancija“, a potom i ...

Google Scholar

Google Scholar

hrcak



Articles

About 99,900 results (0.04 sec)

Any time

Since 2021

Since 2020

Since 2017

Custom range...

Sort by relevance

Sort by date

include patents

include citations

Create alert

Figuren der Relationierung von Lehrerwissen und Lehrerkönnen

[GH Neuweg](#) - Pedagogijska istraživanja, 2004 - [hrcak.srce.hr](#)

... FIGUREN DER RELATIONIERUNG VON LEHRERWISSEN UND LEHRERKÖNNEN.

Pedagogijska istraživanja, 1 (1), 136-136. Preuzeto s <https://hrcak.srce.hr/139366> MLA 8th Edition Neuweg, Georg Hans ... 136-136. <https://hrcak.srce.hr/139366>. Citirano 26.08.2020 ...

☆ Cited by 121 [Related articles](#) [All 2 versions](#)

Selektivni turizam, hir ili znanstveno-istraživačka potreba

[T Luković](#) - Acta turistica nova, 2008 - [hrcak.srce.hr](#)

... APA 6th Edition Luković, T. (2008). Selektivni turizam, hir ili znanstveno-istraživačka potreba. Acta turistica nova, 2 (1), 51-74. Preuzeto s <https://hrcak.srce.hr/37744> MLA 8th Edition Luković, Tihomir ... 2, br. 1, 2008, str. 51-74. <https://hrcak.srce.hr/37744>. Citirano 17.06.2020 ...

☆ Cited by 50 [Related articles](#) [All 2 versions](#)

Flavonoid analysis and antimicrobial activity of commercially available propolis products

[I Kosalec](#), [S Pepeljnjak](#), [M Bakmaz](#)... - Acta ..., 2005 - [hrcak.srce.hr](#)

... Flavonoid analysis and antimicrobial activity of commercially available propolis products. Acta Pharmaceutica, 55 (4), 423-430. Preuzeto s <https://hrcak.srce.hr/19048> MLA 8th Edition KOSALEC, IVAN, et al ... 4, 2005, str. 423-430. <https://hrcak.srce.hr/19048>. Citirano 13.11.2020 ...

☆ Cited by 207 [Related articles](#) [All 12 versions](#)

[\[PDF\] srce.hr](#)

Kubelka-Munk theory in describing optical properties of paper (I)

[V Džimbeg-Malčić](#), [Ž Barbarić-Mikočević](#), [K Itrić](#) - Tehnički vjesnik, 2011 - [hrcak.srce.hr](#)

... i Itrić, K. (2011). Kubelka-Munk theory in describing optical properties of paper (I). Tehnički vjesnik, 18 (1), 117-124. Preuzeto s <https://hrcak.srce.hr/65936> MLA 8th Edition Džimbeg-Malčić, Vesna, et al ... 1, 2011, str. 117-124. <https://hrcak.srce.hr/65936>. Citirano 30.07.2020 ...

☆ Cited by 165 [Related articles](#) [All 3 versions](#)

Zaključno

- od iznimnog je značaja da PDF dokument sadrži sve potrebne metapodatke koji opisuju glavne značajke rada
- većina upita stići će od Google tražilice
- prosječan će korisnik pohraniti zanimljiv rad (PDF) i učitati ga u program za upravljanje referencama (ZOTERO, Mendeley i dr.)
- manjkavi prikaz ukazat će (neopravdano) na manjkavosti uređivačkih procesa i narušiti (nepotrebno) ugled časopisa
- manjkavi prikaz smanjit će čitanost i posebno citiranost časopisa/radova
- PDF-ovi se nerijetko odvajaju od metapodataka o radu sadržanih na stranicama časopisa/sveščića ili repozitorija (npr. Hrčka) i stoga je važno da sadrže sve potrebne informacije
- funkcionalni PDF unaprijeđuje pronalaženje, citiranje, interoperabilnost s drugim sustavima i značajno unaprijeđuje ugled časopisa