

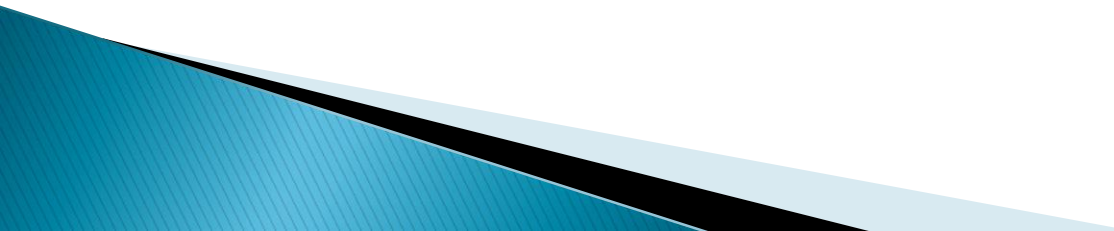
# (UN)ORIGINAL plagiarism checker

Razvoj softvera 2  
–projekat–

dr Vladimir Filipović

Jelena Vidaković  
Dušan Zornić  
Uroš Drakulić

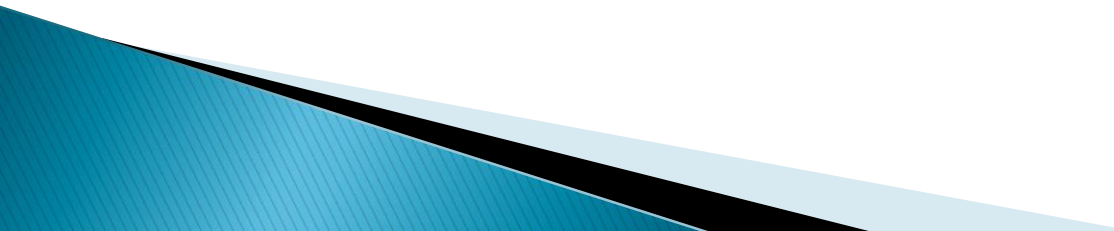
# Pojam plagijarizma

- ▶ Korišćenje tuđe intelektualne svojine u okviru naučnih radova, eseja, programa
  - ▶ Korišćenje celokupne rečenice i pasusa
  - ▶ Korišćenje dela rečenice
  - ▶ Autoplajarizam
  - ▶ Otkrivanje plagijarizma se zasniva na algoritmima za poređenje teksta
- 

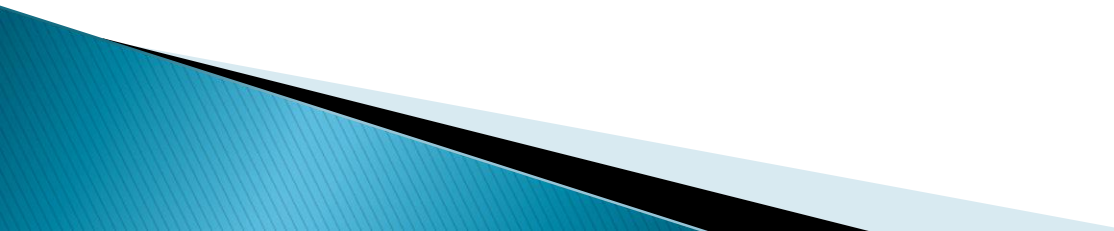
# Postojeći alati

- ▶ Optimizovani uglavnom za englesko govorno područje, ali moguća primena i za naš jezik
- ▶ Neki od najpoznatijih su:
  - [Turnitin](#)
  - [Viper](#)
  - [PlagTracker](#)
  - [Plagiarism Checker](#)
  - [Copyscape](#)
  - [Plagium](#)

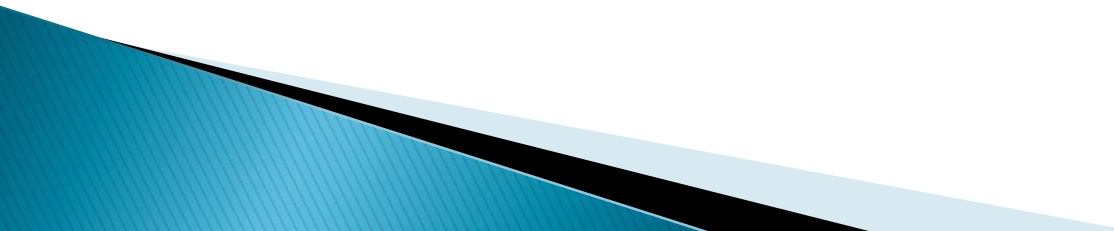
# plagiarism checker – Ideja

- ▶ Otkrivanje identičnih radova
    - Poređenje dva rada
    - Poređenje više radova (iz istih i različitih kategorija)
    - Pronalaženje ključnih reči
  - ▶ Klasifikacija radova
  - ▶ Otkrivanje plagijarizma na osnovu koeficijenta sličnosti
- 

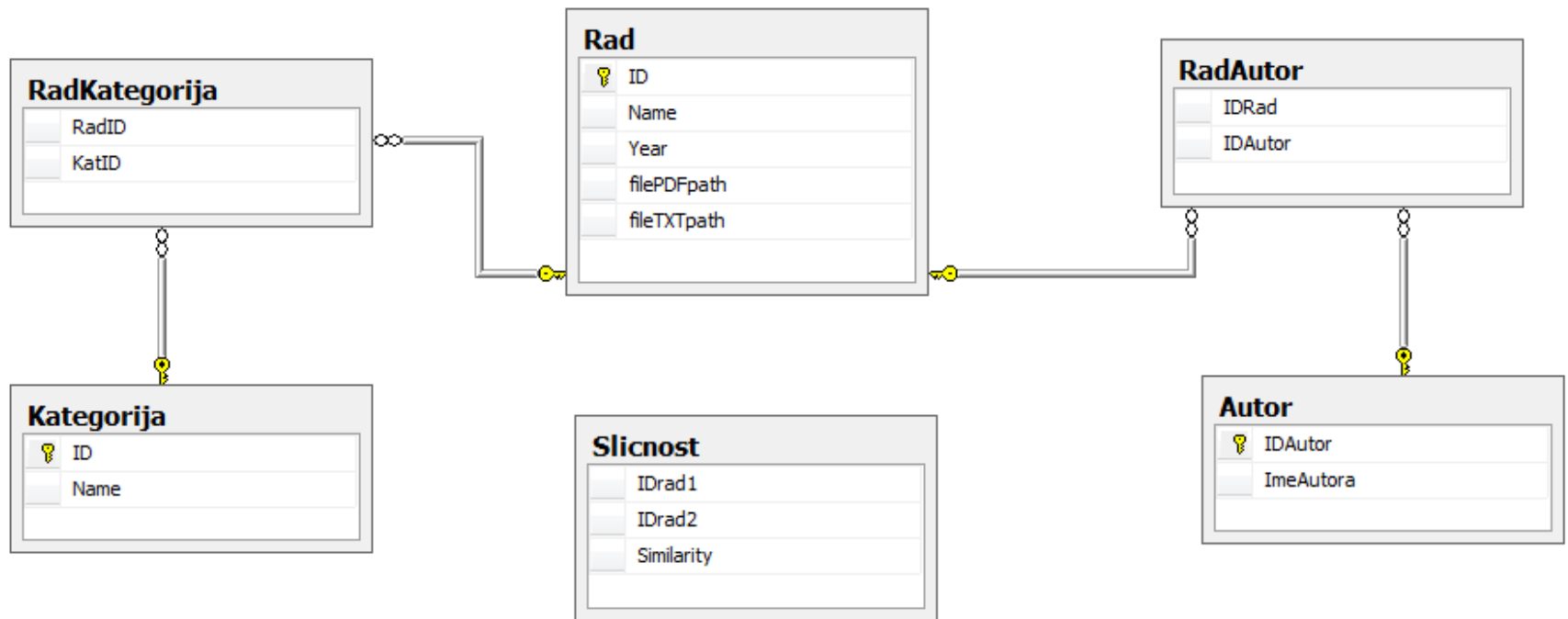
# Korišćene tehnologije

- ▶ MS SQL Server 2008 R2
  - ▶ Visual Studio 2010
  - ▶ ASP .NET 4.0
- 


# Plagiarism checker – razvoj

- ▶ Baza
  - ▶ Interfejs sa pozadinskom logikom za komunikaciju sa bazom
  - ▶ Proces
    - unos i skladištenje rada u bazi
    - brza pretraga
    - algoritam za konverziju rada u oblik pogodan za proveru sličnosti
    - algoritam za proveru sličnosti
- 

# Plagiarism checker – baza

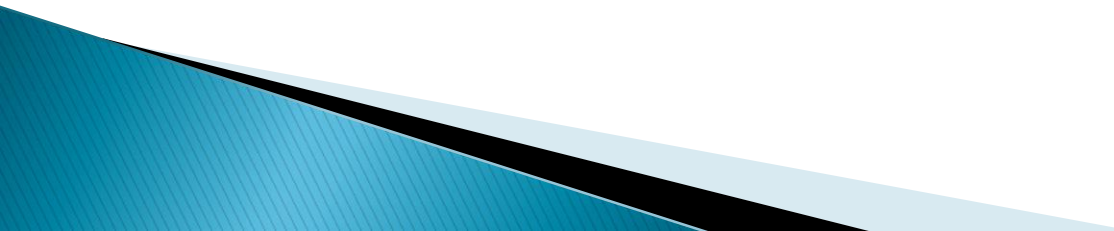


# Plagiarism checker – proces

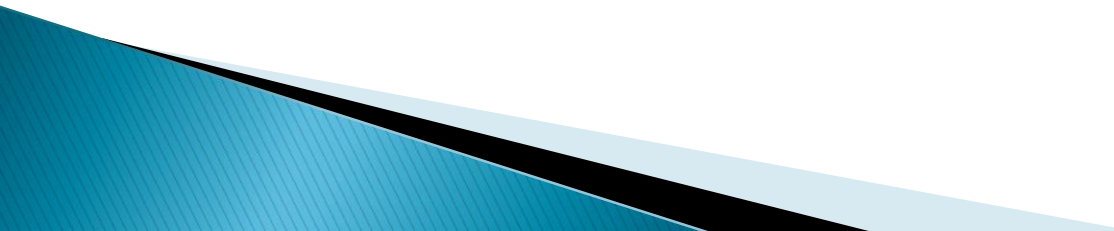
FORMA ZA UNOS RADOVA		
<div>Browse... No file selected.</div>		
Naziv:	<input type="text"/>	
Godina:	<input type="text"/>	
Autor:	<input type="text"/>	
Kategorije:	<div><input type="checkbox"/> Kriptografija</div> <div><input type="checkbox"/> Informacioni sistemi</div> <div><input type="checkbox"/> Analiza algoritama</div> <div><input type="checkbox"/> Istrazivanje podataka</div> <div><input type="checkbox"/> Vestacka inteligencija</div> <div><input type="checkbox"/> Programski jezici</div> <div><input type="checkbox"/> Naucno izracunavanje</div> <div><input type="checkbox"/> Programske paradigme</div> <div><input type="checkbox"/> Bioinformatika</div> <div></div>	
		Label
<div>Unesi rad</div>		



# Plagiarism checker – unos rada

- ▶ Čuvanje originalnog pdf fajla i konvertovanje u txt
  - ▶ `using org.apache.pdfbox.util;`
  - ▶ `using org.apache.pdfbox.pdmodel;`
  - ▶ Dodatni podaci o radu
    - kategorija
    - godina
    - autor
- 

# Plagiarism checker – poređenje

- ▶ Myers–diff algoritm
  - ▶ Snake algorithm
  - ▶ Koeficijent sličnosti se dobija na osnovu broja reči korišćenih u oba rada, u odnosu na celokupan tekst.
  - ▶ Ima vrednost u intervalu  $[0,1]$
- 

# plagiarism checker –poređenje

- ▶ Sličnost između rada i svih ostalih
- ▶ Sličnost između dva rada
- ▶ Sličnost između radova iste kategorije
- ▶ Tabela sličnosti

# Cilj projekta

- ▶ Svest o plagijarizmu
  - ▶ Dalji razvoj i planovi
    - unapređenje dizajna
    - dalje formatiranje za uneti rad
    - dodatni parametri pretrage
    - Web crawler
    - Login i admin panel
    - Integracija sa google-om
    - Integracija sa web servisima za prevod (google translate)
- 