

Izveštaj o aplikaciji Amnesia

Cilj aplikacije Amnesia (<https://amnesia.openaire.eu/>) koja je razvijena u okviru OpenAIRE projekta (<https://www.openaire.eu/>) jeste anonimizacija baza podataka u statističke i istraživačke svrhe. Anonimizacija znači da se iz prikupljenih podataka za neko istraživanje izbacе svi podaci o ličnosti koji bi zadirali u privatnosti i ne bi bili u skladu sa zaštitom ličnih podataka koja je definisana zakonom. Anonimizacijom se dobija skup podataka pogodan je za razmenu u naučno istraživačke svrhe bez ugrožavanja privatnosti.

Aplikacija Amnesia se može koristiti instalirana na računaru ili na Zenodo Serveru *online*. Preporučeno je da se *online* varijanta može da koristi samo manje skupove podataka za probu i učenje tako da se preporučuje korišćenje na računaru.

U svrhu korišćenja aplikacije Amnesia na računaru, neophodno je instalirati tekuću Oracle ili OpenJDK Java Runtime Environment verziju 8 ili noviju zavisno od operativnog sistema računara. U okviru Serbia.RDM tima Amnesia je instalirana na Windows 10 operativnom sistemu sa OpenJDK Java Runtime Environment verziju 8 (preuzeta sa sajta: <https://www.openlogic.com/openjdk-download>), a tačna verzija Amnesia softvera je: 1.2.3 beta.

Na početku korišćenja aplikacije, najpre se učitava baza podataka u aplikaciju. Baza podataka je u obliku teksta gde svaki slog (zapis) predstavlja jedan red. Pojedini podaci - atributi su odvojeni graničnicima. Zatim se u sledećem koraku određuju graničnici (obično su to znaci interpunkcije ili specijalni znaci kao ., , : / itd.). Takođe, određuje se i tip baze podataka („Simple table“, „Sets of values“, „Table with a set valued attribute“, „Disc based simple table“) i prelazi se na sledeći korak. Potvrđuje se tip podatka – promenljivih koji je prepoznat u programu (npr. datum celobrojna vrednost integer, string) i potrebno je da korisnik potvrdi tip podatka.

Da bi se izvršila anonimizacija, potrebno je učitati hijerarhiju tj. osnovu po kojoj se vrši anonimizacija. Može se koristiti već zadata hijerarhija koja se učitava iz datoteke, a može se i generisati iz zadatih podataka. U tom slučaju korisnik za određene kolone zadaje parametre za anonimizaciju uz pomoć programa moguće je napraviti odgovarajuću hijerarhiju. Hijerarhija predstavlja način generalizacije - grupisanja podataka kako bi se na taj način onemogućilo prepoznavanje osetljivih podataka. Pri tome se grupisanje može vršiti u više nivoa. Odabirom odgovarajućih nivoa grupisanja raznih parametara (vrednosti varijabli) može se postići optimalan odnos nivoa detalja i anonimnosti.

Nakon toga, se definiše algoritam. U algoritmu je potrebno povezati promenljivu koju je potrebno anonimizirati sa tabelom za anonimizaciju. Potrebno je spojiti hijerarhiju sa odgovarajućim atributom i zadati k parametar. K je nivo anonimnosti. On označava zadati minimalni broj pojavljivanja neke vrednosti u okviru kombinacije varijabli da bi se sačuvala anonimnost.

Uz program se dobijaju primeri za lakše upoznavanje sa načinom funkcionisanja i korišćenja programa.

Rešenje anonimizacije se daje u obliku grafikona na kojima su prikazani čvorovi koji zadovoljavaju i koji ne zadovoljavaju standard za anonimizaciju. Postoji mogućnost odabira odgovarajuće kombinacije atributa prema potrebi istraživanja. Kod čvorova koji ne zadovoljavaju standard za anonimizaciju se može izvršiti „Potiskivanje“ („Suppress“). Potiskivanje označava izbacivanje vrednosti koje kompromituju anonimnost, a koje neće imati veliki uticaj na krajnji rezultat. Time se dobija anonimizirana baza podataka.

Aplikacija Amnesia predstavlja programa za anonimizaciju podataka. Glavne odlike aplikacije su jednostavnost korišćenja i mogućnost izbora – podešavanja pravila anonimizacije prema potrebama istraživača.

Pripremila Jelena Nikolić

u okviru volonterskog rada u Serbia.RDM timu (<https://rdm.open.ac.rs/>)

U Beogradu, 1. april 2021.