

Kompletan scenario k – anonimnost

Cilj ovog priručnika je demonstracija anonimizacije skupa podataka koristeći algoritam koji pruža k – anonimnost.

Učitavanje skupa podataka

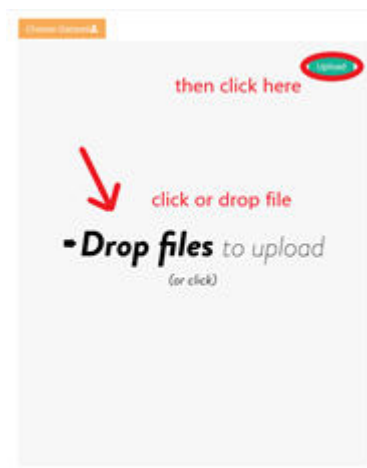
Prvi korak za anonimizaciju skupa podataka D je učitavanje podataka D u aplikaciju Amnesia. Potrebno je imati na umu da aplikacija Amnesia uzima kao ulaz originalni neanonimni skup podataka D . Ako se aplikacija Amnesia ne koristi lokalno na računaru, podaci D se prenose mrežom na server aplikacije Amnesia i podložni su napadima zlonamernih trećih strana.

Skup podataka D može se učitavati na sledeće načine:

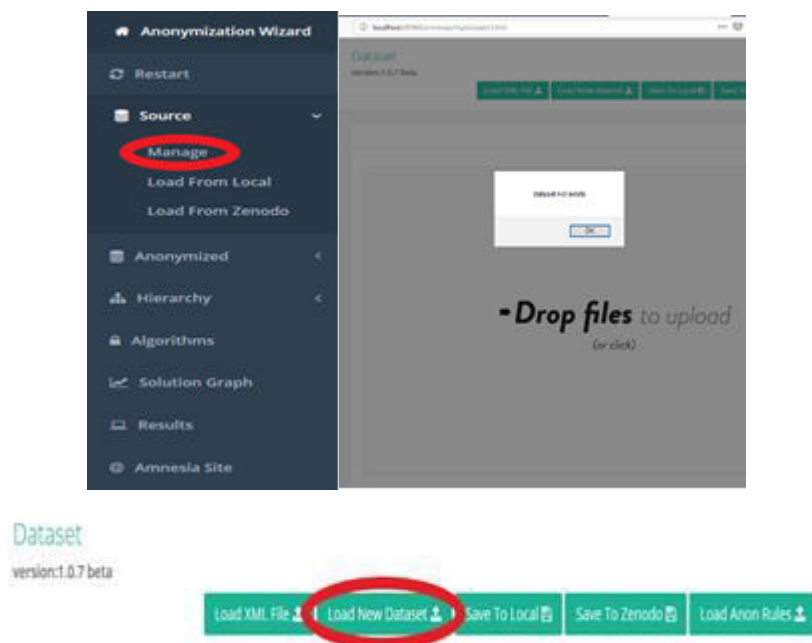
Opcija 1: Klikom narandžasto dugme, a zatim odabirom D za učitavanje



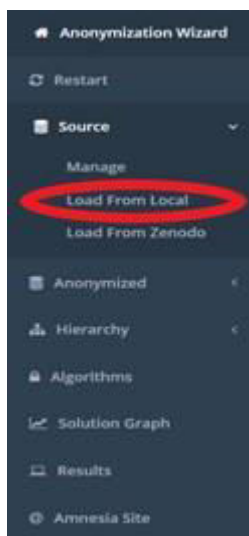
Opcija 2: Klikom da bi se izabralo D ili prevlačenjem D , a zatim klikom na „Upload“ dugme.



Opcija 3: Klikom na opciju „Source“ u levom meniju, a zatim na „Manage“. Potrebno je pročitati poruku upozorenja, a nakon toga klikom na „Load New Dataset“ priložiti *D*.



Opcija 4: Klikom na opciju „Source“ u levom meniju, a zatim na „Load From Local“ i odabirom skupa podataka koji se učitava.



Aplikacija Amnesia prihvata ulazne tekstualne datoteke gde je svaki zapis (slog) prenet u jedan red, a pojedine vrednosti su odvojene fiksnim graničnikom (delimiter). K – anonimnost se primenjuje samo na tabelarne podatke, tj. podatke koji imaju fiksni broj odvojenih vrednosti u svakom zapisu. To znači da D mora imati fiksni broj kolona. Jednom kada je D učitano u aplikaciju Amnesia, korisnik koristi čarobnjaka da bi D bio pravilno protumačen. Korisnik mora navesti graničnik i tip skupa podataka (skupovi podataka i skupovi relacionih podataka mogu biti anonimizirani samo primenom k^m – anonimnosti). U ovom slučaju graničnik je zapeta „“.

Dataset Load Wizard

version:1.0.7 beta

Dataset Load

1. Delimiter

2. Variables

Choose delimiter

This is how the dataset looks like :

zipcode,age,creditcard,gender,salary

56335,58,5557783527541459,Male,8700

57255,36,5418686973265201,Female,9700

98559,32,5527060358825468,Female,6800

...

Delimiter *

.

DataSet Type :

Tabular

Tabular

Set

Relational and Set

Amnesia predstavlja sve kolone iz skupa podataka i pogađa (predlaže) njihov tip. Neophodno je da korisnik verifikuje tip svake kolone i označi kolone koje će učestvovati u postupku anonimizacije.

Dataset Load

1. Delimiter

2. Variables

What type is your data?

Choose the columns and their types.

zipcode	age	creditcard	gender	salary
int	int	string	string	int
57255	36	5418686973265201	Female	9700
56335	58	5557783527541459	Male	8700
98559	32	5527060358825468	Female	6800

Finish

Cancel

Kada se izvrše svi izbori, korisnik mora kliknuti dugme „Finish“ da bi učitao skup podataka. Dalje, korisnik može preći na hijerarhije ili može sačuvati skup podataka klikom na „Save To Local“.

3

newData.txt

Show 10 entries

zipcode	age	creditcard
56335	58	5557783527541459
37255	36	5418686973265201
98559	32	5527060358825468
28700	58	5312916958971375
68925	52	5541858987662877
96338	38	5155271703366251
10840	38	5485337334153888
48772	32	5293894792403628
79641	19	5275938856549264
72861	82	5303041772852809

Showing 1 to 10 of 999 entries

Previous 1 2 3 4 5 ... 100 Next

Check Anonymization Proceed to Hierarchies

Dataset
version:1.0.7 beta

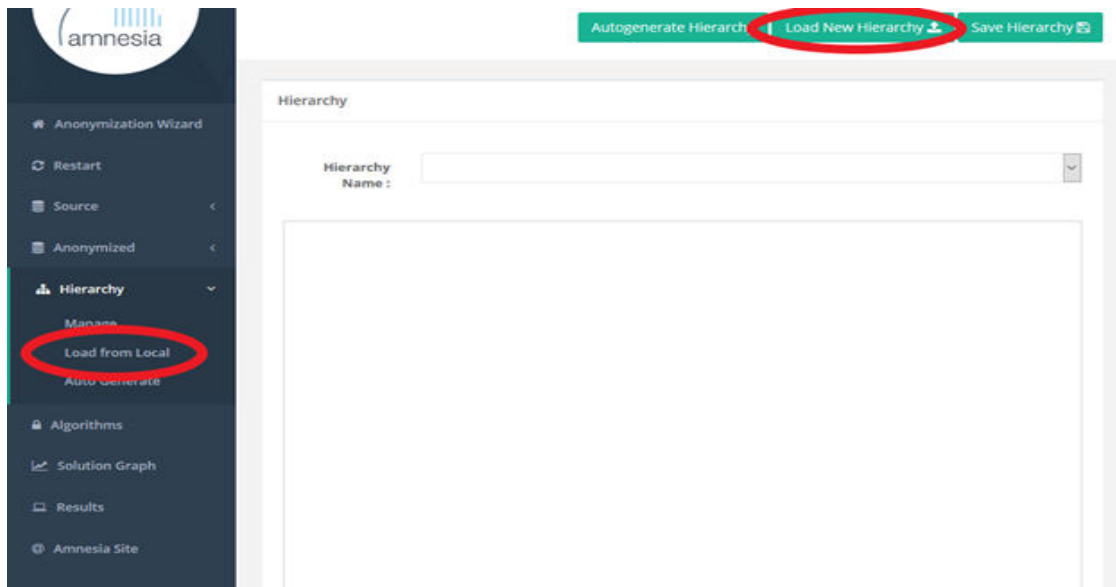
Load XML File Load New Dataset **Save To Local** Save To Zenodo Load Anon Rules

Predstavljanje generalizacionih hijerarhija

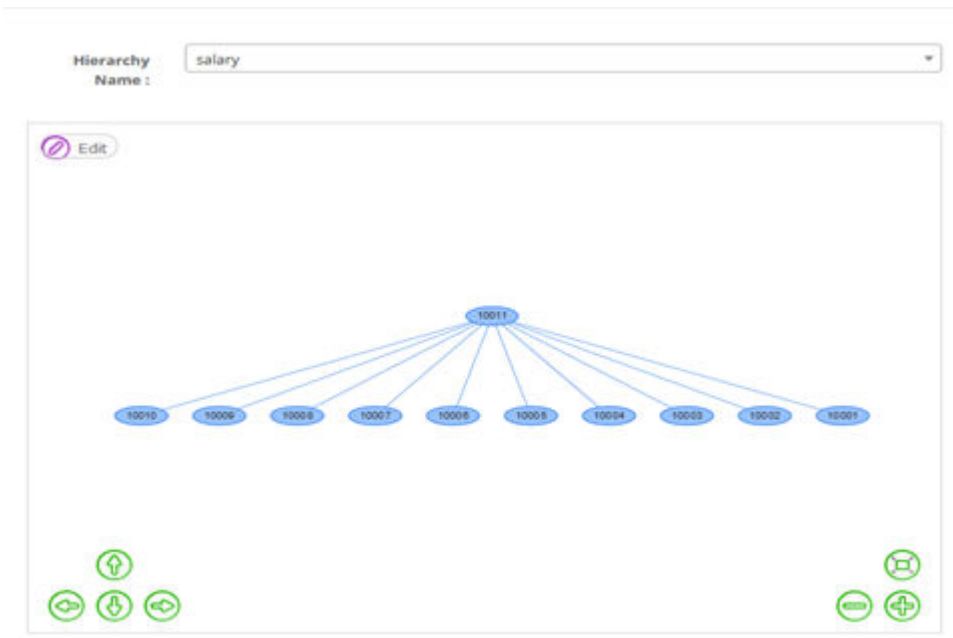
Algoritam aplikacije Amnesia k – anonimnosti koristi generalizaciju za stvaranje grupa identičnih podataka. Pojedinačne ili retke (pojavljuju se manje od k puta u skupu podataka) vrednosti ili kombinacije vrednosti zamenjuju se opštim vrednostima. Na primer, ako u skupu podataka postoji jedna osoba koja živi u Grčkoj, onda Grčka (i ostatak zemalja EU) biće zamenjen sa EU da bi se stvorilo više od k identičnih vrednosti u Zemlji prebivališta. Hijerarhiju generalizacije koja definiše kako se odvijaju ove zamene obezbeđuje korisnik kao ulaz u algoritam.

Amnesia pomaže korisniku da sam kreira hijerarhije koristeći postojeće skupove podataka, više o tome kako da kreira i uredi postojeću hijerarhiju je dato u [Tutorial](#).

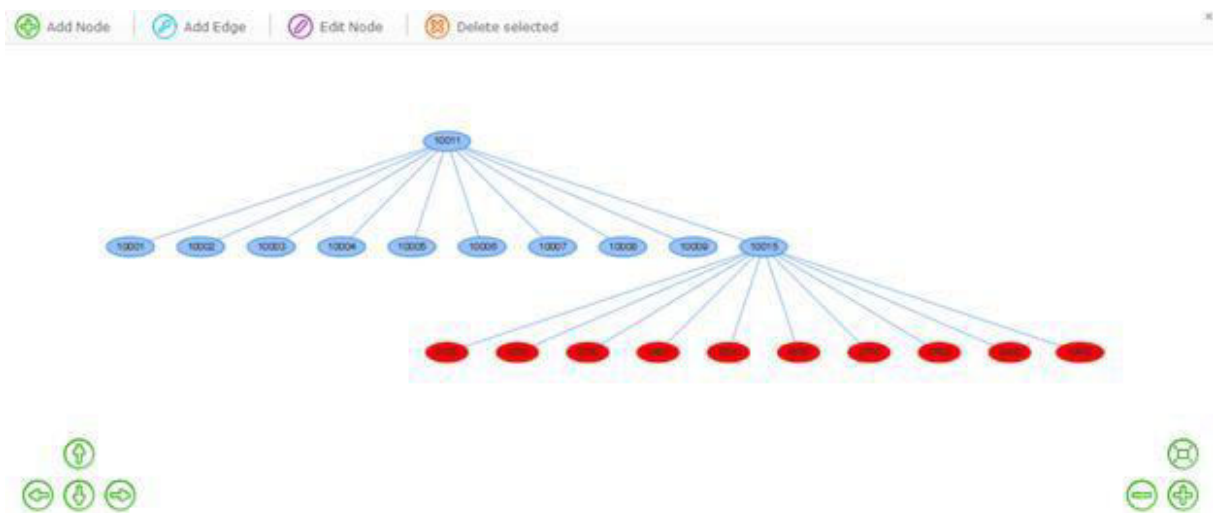
Korisnik može učitati hijerarhiju na dva načina: a) klikom na dugme „Load New Hierarchy“ ili b) klikom na „Load from Local“ u levom bočnom meniju.



Učitane hijerarhije pojavljuju se na ekranu kao grafikoni gde njegovi nadređeni čvor ukazuje na to kako su njihovi podređeni čvorovi generalizovani.



Dvoklikom na čvor mogu se videti njihovi podređeni članovi (deca).



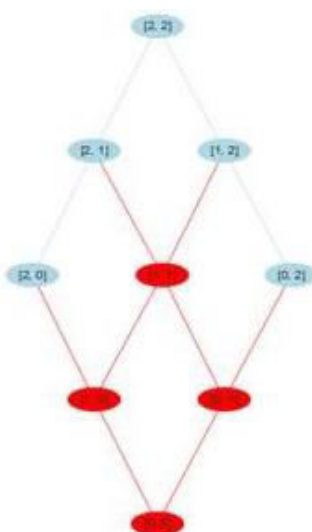
Izvršenje algoritama i rezultati

Kada se skup podataka i hijerarhija učitaju, korisnik može pristupiti izvođenju algoritma anonimizacije. Potrebno je da korisnik: a) poveže svaki atribut sa hijerarhijom koja će biti korišćena na njemu, b) definiše vrednosti k . Dugme za izvršenje „Execute“ pokreće algoritam anonimizacije.

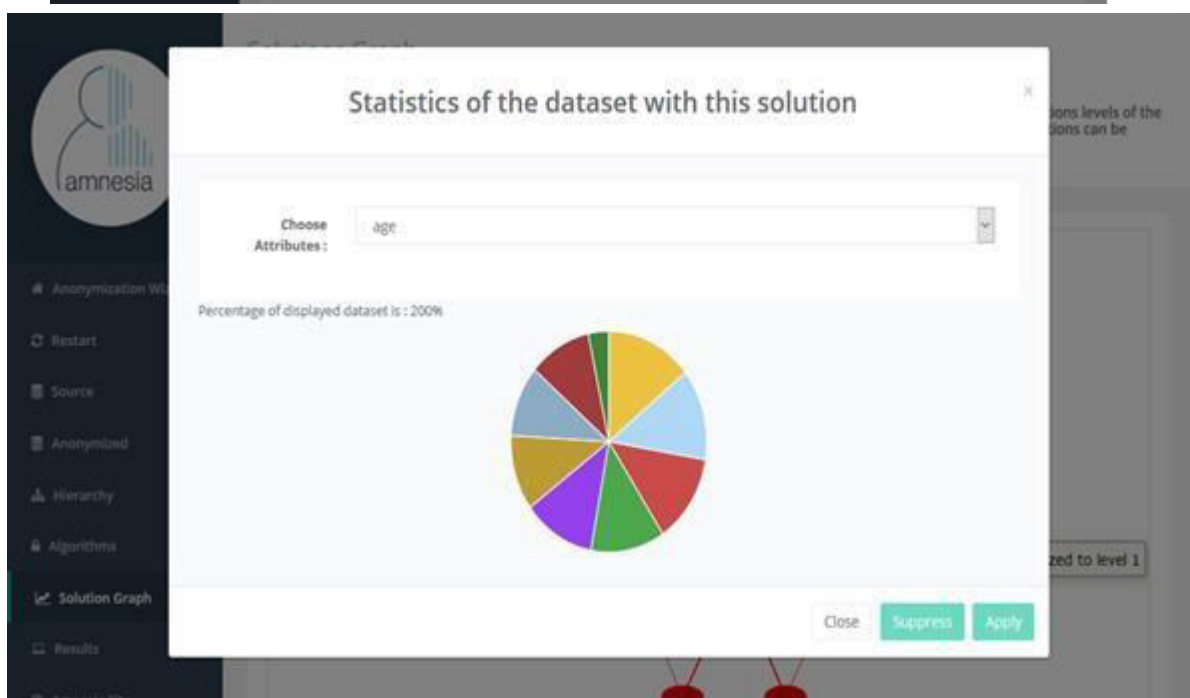
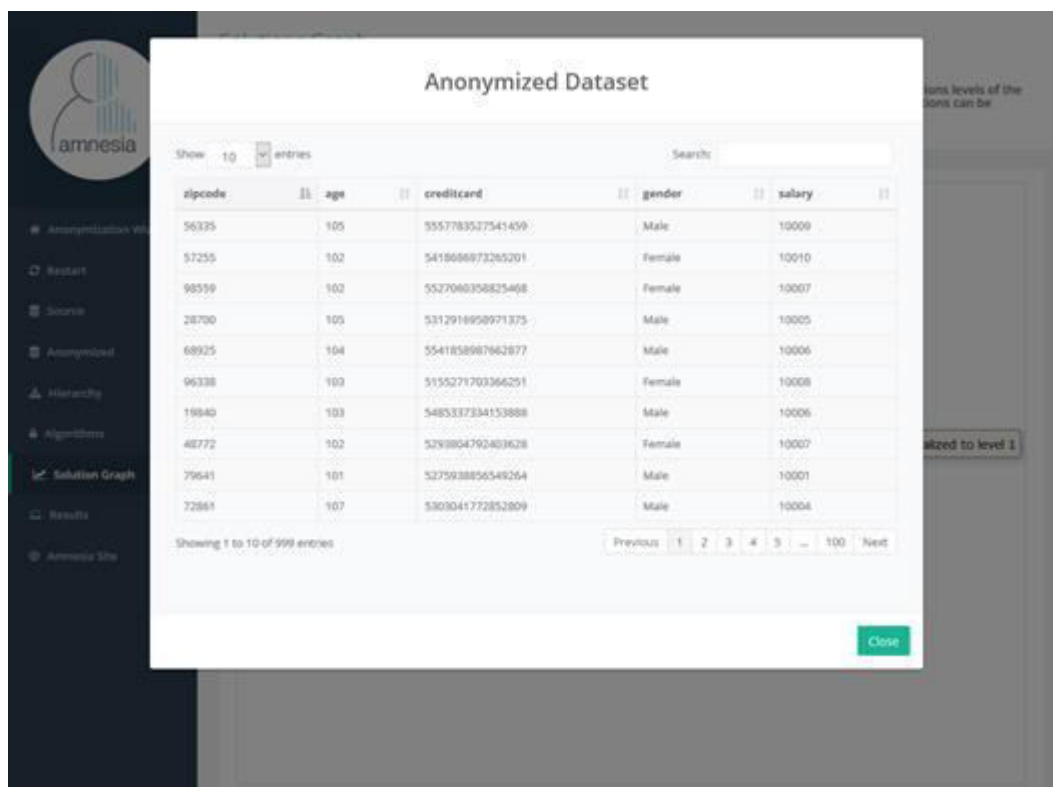
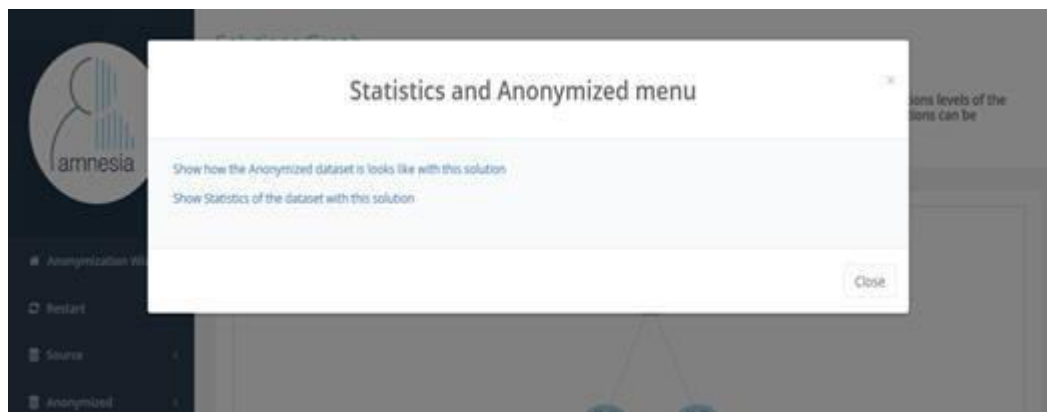
The screenshot displays the user interface for data anonymization. On the left, a table shows sample data with columns: zipcode, age, creditcard, gender, and salary. Below the table is a 'Bind hierarchies' section with a red arrow pointing to a 'Bind Hierarchies with Attributes' area. This area contains dropdown menus for each attribute to select a hierarchy. On the right, a 'Hierarchy' diagram shows a tree structure. Below it, the 'Algorithms' section shows a 'Type' dropdown set to 'Flash' and a red oval highlighting the 'Execute' button.

Prostor za rešenja

Algoritam anonimizacije ne bira automatski rešenja već nudi sva moguća rešenja korisniku. Rešenja su grafički prikazana kao rešetka, gde svaki čvor predstavlja različito rešenje. Plavi čvorovi zadovoljavaju zahteve k – anonimnosti, a crveni ne. Svaki čvor je povezan sa nizom brojeva. Svaki broj pokazuje nivo generalizacije svakog kvazi identifikatora. Prvi broj pokazuje nivo generalizacije prvog kvazi identifikatora, drugi za drugi kvazi identifikator itd. Više detalja se prikazuje ako korisnik zadrži pokazivač (kursor) iznad čvora. Jednim klikom na čvor se aktivira iskačući meni koji nudi dve opcije: a) Prikazati uzorak anonimiziranog skupa podataka za to rešenje ili b) Prikazati statistiku anonimiziranog skupu podataka koji odgovara odabranom rešenju.



Slika 1 – Grafikon rešenja za anonimizaciju dve kolone (plata, starost) sa $k = 4$



Ideja da se prikažu sva moguća rešenja je zbog toga da se omogući korisniku da odabere ono koje nudi najbolji kompromis između privatnosti i potrebne upotrebljivosti. Iako svi plavi čvorovi zadovoljavaju zahteve privatnosti, različiti čvorovi generališu različite attribute do različitih nivoa. Na primer, neko može generalizovati *platu* više od *starosti*, dok drugi može učiniti suprotno, i jedno i drugo rešenje pruža potrebnu garanciju privatnosti. Korisnik može odabrati ono koje bolje odgovara nameni. Štaviše, Amnesia nudi jedan dodatni način za smanjenje gubitka informacija. Rešenja koja ne ispunjavaju zahteve privatnosti tj. pojavljuju se kao crveni čvorovi u grafikonu, možda krše zahtev samo zbog nekoliko zapisa, npr. 0,13 % od ukupnog broja evidencija. Korisnik na ekranu Statistika može videti broj zapisa koji krše garanciju privatnosti, a zatim odabrati da ih suzbije (klikom na dugme „Suppress“), ako se veruje da će više podataka biti sačuvano gubljenjem nekoliko zapisa kako bi se sačuvao nizak nivo generalizacije.

Dvostrukim klikom na čvor primenjuje se odabrana anonimizacija i korisniku se omogućava da vidi anonimni skup podataka uporedo sa originalnim skupom podataka.

zipcode	age	creditcard	gender	salary
56335	58	5557783527541459	Male	8700
57255	38	5418686973265201	Female	8700
98559	32	5527060358825488	Female	6800
28700	58	5312916958971375	Male	4700
68925	32	5541858987662877	Male	5700
96338	38	5155271703366251	Female	2100
19840	38	5485337334153888	Male	6000
48772	32	5293804792403628	Female	7000
79641	19	5275938856549264	Male	100
72861	82	5303041772852809	Male	4000

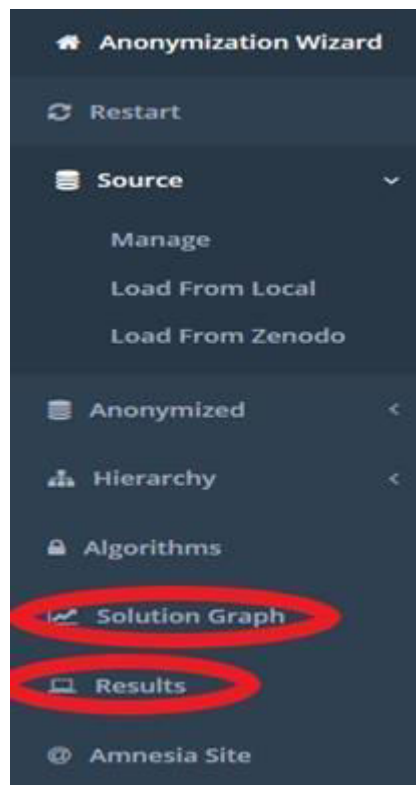
zipcode	age	creditcard	gender	salary
56335	105	5557783527541459	Male	10009
57255	102	5418686973265201	Female	10010
98559	102	5527060358825488	Female	10007
28700	105	5312916958971375	Male	10005
68925	104	5541858987662877	Male	10006
96338	103	5155271703366251	Female	10008
19840	103	5485337334153888	Male	10006
48772	102	5293804792403628	Female	10007
79641	101	5275938856549264	Male	10001
72861	107	5303041772852809	Male	10004

U ovom slučaju može se sačuvati anonimizirani skup podataka odabirom „Save To Local“ ili dodatno sačuvati pravila anonimizacije – „Save Rules“. Pravila anonimizacije predstavljaju prikaz izabranih rešenja.

Results
version:1.0.7 beta



Do ovog ekrana se može doći direktno klikom na opciju rezultata u levom bočnom meniju.



Pravila anonimizacije

Sačuvana pravila anonimizacije mogu se učitati i primeniti na druge skupove podataka koji imaju slične podatke, a skup podataka biće direktno anonimiziran. Nakon učitavanja skupa podataka korisnik ima mogućnost da učitava pravila anonimizacije kao što se može videti dole.

Dataset
version:1.0.7 beta



U ovom trenutku su predstavljena dva scenarija koja obrađuju skupove podataka sa skupom vrednosti i objektno – relacionim podacima.

Ukratko, postupak je jednostavan, korisnik treba da učitava odgovarajući skup podataka, da učitava hijerarhije povezane sa odgovarajućim atributima, da pokrene algoritam i na kraju će dobiti anonimizirani skup podataka na svom ekranu pored početnog skupa podataka.

NAPOMENA: Svi termini koji su korišćeni u ovom dokumentu i koji su izraženi u gramatičkom muškom rodu, porazumevaju prirodni ženski i muški rod.

Prevela dokument „A complete k-anonymity scenario“, pod CC licencom, (<https://amnesia.openaire.eu/Scenarios/AmnesiaTutorialKAnon.pdf>) sa engleskog na srpski Jelena Nikolić u sklopu volontiranja za Serbia.RDA grupu u Beogradu aprila 2021.