**PLAN UPRAVLJANJA ISTRAŽIVAČKIM PODACIMA (PUP)**

|  |  |  |
| --- | --- | --- |
| Opće informacije | | |
|  | Ime i prezime predlagatelja | Vinko Zlatić |
|  | Matična organizacija | Zavod za teorijsku fiziku, Institut Ruđer Bošković |
|  | Naziv projekta | Analiza stohastičkih procesa na mrežama u sustavima s ograničenom informacijom |
|  | Upravitelj podacima | Vinko Zlatić |
| 1. | Prikupljanje podataka i dokumentacija | |
|  | Koje ćete podatke prikupljati, obrađivati, stvarati ili se ponovno njima koristiti? (navedite formate, vrste i opseg svih podataka s kojima ćete raditi, a ne samo krajnji skup podataka koji će biti rezultat istraživanja) | Prikupljat ćemo podatke iz otvorenih repozitorija mrežnih struktura kao što su https://networkrepository.com/, https://networkdata.ics.uci.edu/, https://networks.skewed.de/. Koristit ćemo standardne mrežne formate podataka poput popisa čvorova i veza. Generirati ćemo umjetne podatke koji će biti spremljeni u istim formatima. |
|  | Kako će se podaci prikupljati, obrađivati ili stvarati? (ukratko navedite metodologiju i procese osiguranja kvalitete, načine organiziranja podataka te alate i instrumente kojima ćete se koristiti za prikupljanje i obradu) | Podaci koji su već odabrani bit će uzeti kao realna vrijednost. Kvaliteta podataka koje ćemo generirati bit će osigurana putem javno dostupnog koda. Obrada podataka vršit će se u programima awk, python i matlab. Veliki dio projekta zapravo je generiranje podataka za koje ćemo koristiti python i c++ s bibliotekom graphtool.  Podaci će biti označeni prema – sustavu i procesu koji simuliraju, te nadalje, parametrima za mrežu i procese. |
|  | Koju ćete dokumentaciju i metapodatke izraditi osim podataka? (dokumentacija mora sadržavati informacije i standarde potrebne korisnicima kako bi mogli samostalno čitati i interpretirati podatke u budućnosti, primjerice, kodne knjige, *ReadMe* datoteke i sl.) | Osigurat ćemo svu dokumentaciju podataka potrebnu za ponovno stvaranje podataka, osim niza rand brojeva potrebnih za točnu rekonstrukciju instance modela. Koristit ćemo GitHub za dijeljenje projekata sa zajednicom. |
| 2. | Pravna i sigurnosna pitanja | |
|  | Jeste li ograničeni sporazumom o povjerljivosti? Imate li potrebna dopuštenja za prikupljanje, obradu, čuvanje i dijeljenje podataka? Jesu li osobe čiji se podaci obrađuju informirani o tome i jesu li dali privolu? Kojim ćete se metodama koristiti u svrhu zaštite osjetljivih podataka (GDPR - posebne  kategorije osobnih podataka, navesti metode anonimizacije podataka)? | Ne, nismo ograničeni ugovorom o povjerljivosti. Da, podaci su javni i već se dijele, a umjetni podaci uopće nisu etički upitni. Ne postoje osobni podaci i GDPR se ne odnosi na ovo istraživanje. |
|  | Kako će se regulirati pristup podacima i njihova sigurnost? Koji su potencijalni rizici koje treba uzeti u obzir? Kako ćete osigurati sigurnost pohrane osjetljivih podataka? | Podaci će biti otvoreni kada ih iskoristimo. Osigurat ćemo sigurnu pohranu podataka na našim računalima i vanjskim diskovima. |
|  | Kako ćete upravljati zaštitom autorskih prava i drugog intelektualnog vlasništva? Tko će biti vlasnik podataka? Koje će se licencije primjenjivati na podatke? Koja će se ograničenja primjenjivati na ponovnu  uporabu osobnih podataka? | Vlasnik podataka bit će suradnici na projektu. Koristit ćemo samo dostupne podatke trećih strana koji ne podlježu ograničenjima. |
| 3. | Pohrana i čuvanje podataka | |
|  | Kako će radne verzije podataka biti pohranjene tijekom projekta?  Kako će se napraviti sigurnosne kopije tih podataka (*backup*)?  Koja je očekivana količina podataka koja će se prikupiti i čuvati tijekom projekta (izraženo u MB/GB/TB)? | Podatke ćemo pohraniti na lokalne strojeve i na vanjske diskove. Tijekom procesa generiranja generirat ćemo najmanje 2 kopije podataka. Očekujemo da će se generirati 100GB-1TB podataka, što će ovisiti o količini računalog vremena koje ćemo imati na raspolaganju. |
|  | Kako će se završne verzije podataka dugotrajno pohraniti i čuvati (i nakon završetka projekta)?  U kojim će se formatima čuvati podaci? Koja je očekivana količina podataka koja će se trajno pohraniti (izraženo u MB/GB/TB)? | Podaci će biti sigurno pohranjeni na našim računalima i na vanjskim tvrdim diskovima koje ćemo čuvati. Svako oštećenje podataka lako će se riješiti jer možemo samo odbaciti oštećene podatke i ponovno stvoriti skup podataka. Očekujemo da će se generirati 100GB-1TB podataka, što će ovisiti o količini vremena na računalu koje ćemo imati na raspolaganju. Koristit ćemo standardne mrežne formate kao što su GEXF, GDF, GML, GraphML, Pajek NET, GraphViz DOT, CSV, ali općenito ćemo koristiti csv kao najlakši format podataka. |
| 4. | Dijeljenje i ponovna uporaba podataka | |
|  | Kako i gdje će se podaci dijeliti? Koji repozitorij će se koristit za dijeljenje podataka? Kako će potencijalni korisnici doznati za podatke? | Podaci će se dijeliti na zahtjev i/ili iz repozitorija u koje ćemo staviti instance podataka. Većinu repozitorija zanimaju podaci iz stvarnog svijeta, a ne umjetni, tako da ne očekujemo interes repozitorija da čuvaju naše podatke, za razliku od naših kodova koji će biti zanimljivi i dijeljeni putem githuba. |
|  | Ako postoje podaci koji se ne smiju dijeliti (prijavitelji vezani zakonskim, etičkim, autorskim pravila, povjerljivošću i sl.), pojasnite razloge ograničenja. | Nema ih |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji je u skladu s načelima *FAIR-a*. | Potvrđujemo da će digitalni repozitorij koji ćemo odabrati biti u skladu s FAIR načelima. |
|  | Potvrdite da ćete se koristiti digitalnim repozitorijem koji održava neprofitna organizacija (ako ne, objasnite zašto ne možete dijeliti podatke na digitalnom repozitoriju koji nije  komercijalan). | Potvrđujemo da će digitalni repozitorij koji ćemo odabrati biti održavan od strane neprofitne organizacije. |