

TÍM č. 12 IS-SI

Paison



Názov projektu:

OpenScience

Členovia tímu (študenti):

Richard Belan, Rania Daabousová, Tomáš Chovaňák, Milan Gajdoš,
Michal Palatinus, Helmut Posch, Martin Žalondek

Ved. tímu (pedagóg):

Ing. Peter Lacko, PhD.

Motto tímu:

Can't make bricks without a clay, can't make research without data.

O ČOM JE NÁŠ PROJEKT?

OpenScience je webový portál umožňujúci verejné zdieľanie datasetov. Vedci často navrhujú nové metódy a vykonávajú experimenty, na ktoré potrebujú dáta. Datasetsy sú kľúčové hodnotiace

faktory metód. Problém korektného porovnávania výsledkov experimentov s iným experimentom podobného typu si mnohokrát vyžaduje použitie rovnakých experimentálnych dát. Často to však nie je jednoduché, pretože datasety nie sú jednoducho dostupné v rámci vedeckej obce. Ich verejné zdieľanie je tak dôležitou súčasťou vedeckých výskumov, umožňujúcou lepšiu verifikáciu a rýchlejšie napredovanie výskumov.

Naším cieľom je vytvorenie nástroja s potenciálom pre globálne použitie, ktorý výskumníkom z celého sveta a rôznych vedeckých oblastí ponúka možnosť pre jednoduché zdieľanie ich dát. Naš webový portál poskytuje rôzne druhy datasetov zoskupených na jednom mieste a dostupných pre každého. Každý návštevník si môže jednoducho vyhľadať dáta vhodné pre svoju prácu a stiahnuť ich bez potreby registrácie. Registrovaní používatelia môžu prispieť svojimi vlastnými datasetmi, čím nielen pomôžu ostatným výskumníkom, ale taktiež môžu dostať spätnú odozvu na svoje datasety.

Výhodou webového portálu je jednoduchá dostupnosť na rôznych zariadeniach kdekoľvek na svete a bez nutnosti akejkoľvek inštalácie. Portál je vyvinutý na technológiách podporovaných Googlom, konkrétne Google App Engine (Python SDK) a Django s podporou NoSQL databáz.

Momentálne dokážeme nahrávať datasety veľkosti niekoľkých gigabajtov, ale pracujeme na zvyšovaní kapacity a rýchlosti nahrávania. Používatelia môžu jednoducho nájsť dataset prostredníctvom vyhľadávania na základe názvu, tagov alebo popisu. Nápomocnou im môže byť funkcia automatického dopĺňania hľadaného textu po stlačení klávesy a zoznam najčastejšie sa vyskytujúcich tagov. Okrem toho je možné aj vyhľadávanie konkrétnych používateľov. Do budúca ešte plánujeme umožniť používateľom spresniť hľadanie podľa nastavených kritérií ako je dátum pridania, veľkosť a pod. K datasetom je možné pridávať referencie na práce, v ktorých bol použitý, a taktiež ich môžu používatelia hodnotiť. Prispievatelia môžu spravovať svoje datasety vo vlastnom profile. Portál podporuje aj registráciu cez rôzne sociálne siete a takisto je možné datasety na týchto sociálnych sieťach zdieľať.

ČO NÁM DÁVA PRÁCA NA TOMTO PROJEKTE?

Práca na tomto projekte nás obohatila hneď vo viacerých smeroch. Za najvýznamnejší prínos považujeme skúsenosti s prácou vo väčšom tíme, nakoľko pre väčšinu z nás to bola prvá skúsenosť tohto typu. Museli sme spoločne riešiť množstvo problémov či už technických, manažmentových alebo komunikačných. Získali sme skúsenosti v konštruktívnom a efektívnom riešení konfliktov, ktoré sa vždy v rámci tímov vyskytujú a nie sú negatívnym znakom, ale naopak ukazujú, že členovia tímu uvažujú nad daným problémom a vyjadrujú vlastné názory a nápady. Každý z členov si vyskúšal nejakú manažérsku pozíciu v rámci tímu a s tým spojenú zodpovednosť. Prakticky sme sa oboznámili s tvorbou a dodržiavaním metodík vhodných pre náš projekt a tím. Získali sme praktické skúsenosti s agilným vývojom s použitím metodológie SCRUM, ktoré môžu byť užitočné v rámci iných projektov či už na škole, alebo mimo nej. S tým súvisí aj používanie manažmentových nástrojov ako JIRA a Trello. V rámci samotného vývoja sme sa zoznámili s novými technológiami, platformami a jazykmi, pričom niektoré boli úplne nové pre nás všetkých. Navzájom sme si pomáhali a vymieňali skúsenosti, aby sme projekt čo najrýchlejšie, a hlavne v čo najlepšej kvalite, dotiahli až do konca. Objavili sme krásu verziovania zdrojového kódu a vyskúšali sme si, aké je to robiť *code review*, čím sme si zlepšili vlastnú štábnu kultúru programovania. Na rozdiel od doterajších, väčšinou vlastných menších projektov, sme teraz boli nútení písať efektívne a prehľadné zdrojové kódy, aby im jednoducho porozumeli všetci členovia tímu. Celkovo sme teda nadobudli nové technické a komunikačné skúsenosti a určite sme sa naučili aj väčšej zodpovednosti nie len voči sebe, ale aj iným.

PREČO JE NÁŠ PROJEKT ZAUJÍMAVÝ?

Projekt OpenScience je dostupný na akomkoľvek zariadení a používa cloud technológie od spoločnosti Google, ktoré sú zárukou stability a rozšíriteľnosti. Má potenciál zjednodušiť prístup k datasetom a podporiť kolaboráciu v rámci vedeckej obce, či už na fakulte alebo globálne. Výskumníci môžu navzájom porovnávať a zdieľať získané

dáta, čo urýchli výskum a potenciálne zvýši jeho kvalitu. S tým súvisí aj zníženie vytvárania redundantných dát, čo ušetrí nielen čas, ale aj finančné prostriedky. Prehľad vedeckých prác s referenciami na datasety a možnosť ohodnocovania datasetov poskytuje nielen spätnú väzbu autorovi datasetu, ale poukazuje aj na jeho relevantnosť.

Použité technológie:

Google App Engine SDK,
Google Cloud Storage,
Datastore NoSQL databáza,
Django non-rel, Python,
JavaScript,
Google Search API.

O ČOM TO VLASTNE JE

OpenScience je webový portál umožňujúci verejné zdieľanie datasetov. Naším cieľom je vytvorenie nástroja s potenciálom pre globálne použitie, ktorý výskumníkom z celého sveta a rôznych vedeckých oblastí ponúka možnosť pre jednoduché zdieľanie ich datasetov. Portál napomôže presnejšej evaluácii výskumov, čo zabezpečí zvyšovanie ich kvality.