



LIETUVOS MOKSLO TARYBOS FINANSUOTO IMPRESS-U PROGRAMOS PROJEKTO

**„IRIS-HEP: MOKSLO PROGRAMINĖS ĮRANGOS PLĖTIMAS DALELIŲ FIZIKOS REIKMĖMS
UKRAINAI, LENKIJAI IR LIETUVA“**

(Nr. S-IMPRESSU-24-4, reg. data 2024-06-26 d.)

2025 M. STUDENTŲ STAŽUOČIŲ ATASKAITOS

Projekto vykdytojas: Vilniaus Universitetas

Projekto vadovas: dr. Andrius Juodagalvis

2026 Vilnius

ADAPTING THE COFFEA FRAMEWORK FOR FCC

Stażuotoja E. Lenkaitytė, stažuotės konsultantai D. Lange, P. Yadav

2025-09-23

Stażuotė „Adapting the Coffea Framework for FCC“ (2025-06-30 – 2025-09-19) gavo LMT finansavimą pagal projektą „IRIS-HEP: Mokslo programinės įrangos kūrimo bendradarbiavimo plėtimas į Ukrainą, Lenkiją ir Lietuvą“ (sutarties Nr. S-IMPRESSU-24-4, vadovas A. Juodagalvis). Europos branduolinių mokslinių tyrimų laboratorijoje CERN atliktos stažuotės metu buvo vykdomi prieš stažuotę numatyti darbai, kurių aprašas pateiktas pridėtame priede.

Pagrindinis stažuotės užduoties tikslas buvo sukurti dokumentuotą pavyzdį, kaip naudoti Coffea analizės sistemą su FCC modeliuotais duomenimis. Numatytieji stažuotės tikslai buvo įgyvendinti. Svarbiausios stažuotės veiklos buvo šios:

1. Su stažuotės konsultantais sudaryti stažuotės darbų planą. Sulaukti stažuotės projekto vadovų, P. Elmer ir A. Juodagalvio, pritarimų numatytai veiklai. Užsiregistruoti IRIS-HEP.org puslapyje <https://iris-hep.org/fellows>.
2. Susipažinti su Coffea analizės sistema.
3. Sukurti dokumentaciją duomenų analizės kodui ir sujungti į vieną Jupyter notebook.
4. Pritaikyti kodą naujausiai Coffea versijai (2025.1.1 -> 2025.7.3).
5. Sutvarkyti kodą taip, kad būtų lengva suprasti ir naudoti kaip dokumentuotą pavyzdį rašant analizę, bei vykdyti jį Dask klasteryje vietoje lokalaus paleidimo.
6. Atnaujinti rezultatų saugojimą ir histogramų braižymą – pereiti prie kategorijomis pagrįstos histogramos kelių duomenų rinkinių palyginimui.
7. Atliekamas projektas buvo pristatytas 40-tame „Lithuanians@CERN“ susirinkime, kuris įvyko 2025-08-07. Renginio nuoroda: <https://indico.cern.ch/event/1573485/>

Atliekant stažuotei suformuluotas užduotis, buvo naudota ši programinė įranga: Coffea analizės sistema ir Jupyter Notebook aplinka; bei programavimo kalba python.

Stażuotės metu sukurta programinė įranga yra viešai prieinama adresu <https://github.com/egle-git/Coffea for FCC/blob/main/4leptons/4leptons.ipynb>. Ji bus toliau vystoma bei naudojama kitų tyrėjų visame pasaulyje.

Pagrindinė stažuotės nauda buvo tokia:

1. Galimybė aplankyti ir dirbti CERN vasaros laikotarpiu.
2. Tarptautinė aplinka ir bendravimas su įvairių sričių specialistais, galimybė užmegzti ryšius.
3. Praktinė patirtis dirbant su Coffea Framework.

Panašias stažuotes rekomenduoju tęsti, nes jos suteikia ne tik praktinių techninių įgūdžių, bet ir vertingos tarptautinės darbo patirties, leidžia pažinti naujus žmones bei susipažinti su pasaulinio lygio mokslo tyrimų aplinka.

Stażuotoja

E. Lenkaitytė

E-paštas: egle.lenkaityte@ff.stud.vu.lt

Development of an LLM-Based Co-Pilot Tool for CMS Analysts

Stažuotoja Rūta Lomsargytė,
stažuotės konsultantai Peter Elmer, David Lange, Gordon Watts
2025-08-27

Stažuotė „Development of an LLM-Based Co-Pilot Tool for CMS Analysts“ (06-30 – 08-11) gavo LMT finansavimą pagal projektą „IRIS-HEP: Mokslo programinės įrangos kūrimo bendradarbiavimo plėtimas į Ukrainą, Lenkiją ir Lietuvą“ (sutarties Nr. S-IMPRESSU-24-4, vadovas A. Juodagalvis). Europos branduolinių mokslinių tyrimų laboratorijoje CERN atliktos stažuotės metu buvo vykdomi prieš stažuotę numatyti darbai, kurių aprašas pateiktas pridėtame priede.

Pagrindinis stažuotės užduoties tikslas buvo sukurti „Co-Pilot“ įrankį, veikiantį didžiųjų kalbos modelių (LLM) pagrindu, kuris generuotų programinį kodą. Numatytieji stažuotės tikslai buvo įgyvendinti. Svarbiausios stažuotės veiklos buvo šios:

1. Su stažuotės konsultantu sudaryti stažuotės darbų planą. Sulaukti stažuotojų projekto vadovų, P. Elmer ir A. Juodagalvio, pritarimų numatytai veiklai. Užsiregistruoti IRIS-HEP.org puslapyje <https://iris-hep.org/fellows>.
2. Parašyti programinį kodą RAG (retrieval-augmented generation) apdorojimo sekai.
3. Parašyti CMS eksperimento duomenų analizės kodo, kintamųjų ir dažniausiai pasitaikančių klaidų failus, kurie naudojami RAG duomenų apdorojimo sekoje.
4. Atlikti testus pagal IRIS-HEP funkcionalumo vertinimo kriterijus, siekiant įvertinti įrankio sugeneruoto kodo kokybę.
5. Atliekamas projektas buvo pristatytas 40-tame „Lithuanians@CERN“ susirinkime, kuris įvyko 2025-08-07. Renginio nuoroda: <https://indico.cern.ch/event/1573485/>

Atliekant stažuotei suformuluotas užduotis, buvo naudota programavimo kalba python ir bibliotekos awkward, uproot, ollama, chromadb, tiktoken, langchain.

Stažuotės metu sukurta programinė įranga yra viešai prieinama adresu <https://github.com/rrutaa>. Ji bus toliau vystoma bei naudojama kitų tyrėjų visame pasaulyje.

Pagrindinė stažuotės nauda buvo tokia:

1. Programavimo, duomenų analizės įgūdžių tobulinimas: gilintos žinios naudojant naujus įrankius.
2. Darbas tarptautinėje motyvuotų žmonių aplinkoje suteikė naudingų žinių ir įkvėpimo toliau mokytis didelių energijų fizikos.
3. Pažintis su CERN CMS eksperimentu suteikė supratimą, kaip jis vykdomas.
4. Profesinių ryšių, kurie gali būti naudingi tolimesnėje karjeroje, užmezgimas.
5. Karjeros krypties aiškinimasis: suteikta galimybė įvertinti, ar mokslinis darbas atitinka asmeninius lūkesčius.

Panašias stažuotes rekomenduoju tęsti, nes jos suteikia galimybę ugdyti kompetencijas, reikalingas didelių energijų fizikos tyrimams, įgyti darbo ir bendravimo patirties tarptautinėje aplinkoje, užmegzti vertingus kontaktus, detaliau pamatyti, kaip atrodo tyrėjo darbas.

Stažuotoja
E-paštas: ruta.lomsargyte@ff.stud.vu.lt

Rūta Lomsargytė

Backend update of the AlCaVal framework for validation of calibration and alignment conditions data

Stažuotojas Patrikas Styra,
stažuotės konsultantas Matheus Pereira Macedo De Sousa

2025-09-09

Stažuotė „Backend update of the AlCaVal framework for validation of calibration and alignment conditions data“ (2025-06-30 – 2025-09-05) gavo LMT finansavimą pagal projektą „IRIS-HEP: Mokslo programinės įrangos kūrimo bendradarbiavimo plėtimas į Ukrainą, Lenkiją ir Lietuvą“ (sutarties Nr. S-IMPRESSU-24-4, vadovas A. Juodagalvis). Europos branduolinių mokslinių tyrimų laboratorijoje CERN atliktos stažuotės metu buvo vykdomi prieš stažuotę numatyti darbai, kurių aprašas pateiktas pridėtame priede bei atlikti nenumatyti patobulinimai, atsiradę pasikeitus reikalavimams.

Pagrindinis stažuotės tikslas buvo atnaujinti programinę įrangą pagal reikalavimus. Numatytieji stažuotės tikslai buvo įgyvendinti. Svarbiausios stažuotės veiklos buvo šios:

1. Su stažuotės konsultantu sudaryti stažuotės darbų planą. Sulaukti stažuotės projekto vadovų, P. Elmer ir A. Juodagalvio, pritarimų numatytai veiklai. Užsiregistruoti IRIS-HEP.org puslapyje <https://iris-hep.org/fellows>.
2. Atsisiųsti programinės įrangos „AlCaVal“ <https://github.com/cms-AlCaDB/AlCaVal> pirminį kodą. Jį sukompiliuoti ir įvykdyti pirminius patikrinimo darbus.
3. Pagal numatytą planą modifikuoti programinę įrangą: pridėti histogramas detalesniam darbo eigų stebėjimui ir pridėti automatinius darbo eigų atnaujinimus su el. pašto pranešimais apie baigtas eigas.
4. Pristatyti ir pateikti kolegoms įvertinti programinę įrangą ir gauti jų atsiliepimus. Koreguoti programinį kodą.
5. Projekto eiga ir rezultatai buvo pristatyti 40-tame „Lithuanians@CERN“ susirinkime, kuris įvyko 2025-08-07. Renginio nuoroda: <https://indico.cern.ch/event/1573485/>

Atliekant stažuotei suformuluotas užduotis, buvo naudota ši programinė įranga: VSCode, SSH, LLM modeliai, Docker, OpenStack VM, MongoDB; bei programavimo kalbos Python, JavaScript.

Stažuotės metu patobulintos programinės įrangos paskutinė versija yra viešai prieinama adresu <https://github.com/patrikas-git/AlCaVal/tree/feature/workflow-refresh-job>. Tačiau pagal tolimesnius planus, ji nebus toliau vystoma.

Stažuotės metu buvo nuspręsta tobulintą programinę įrangą integruoti su kita. Kita programinė įranga pasiekama CERN vartotojams <https://gitlab.cern.ch/cms-ppd/technical-support/web-services/RelVal>. Paskutinis atliktas atnaujinimas pasiekiamas https://gitlab.cern.ch/cms-ppd/technical-support/web-services/RelVal/-/merge_requests/146. Ji bus toliau vystoma.

Pagrindinė stažuotės nauda buvo tokia:

1. Įgytas supratimas, kaip veikia CERN ir mokslinės bendruomenės
2. Ryšių užmezgimas
3. Įgyta patirties dirbant su technologijomis tokiomis kaip „Docker compose“ ir „Helm charts“
4. Galimybė aplankyti CERN objektus ir susipažinti su Šveicarijos aplinka

Panašias stažuotes rekomenduoju tęsti, nes mokslu besidomintiems žmonėms suteikiama galimybė geriau susipažinti su tokių bendruomenių bei institucijų veikla ir prisidėti prie mokslo inovacijų.

Stazuotojas

Patrikas Styra

El. paštas: patrikass32@gmail.com