

Pokročilé vyhledávání a prohledávání dat

Charta Task Force

Charta

Poslední aktualizace: 22. 04. 2026

Verze: 1.0

Přehled Task Force

Vedoucí:

RNDr. Matej Antol, Ph.D. – Ústav výpočetní techniky, Masarykova univerzita; antol@muni.cz

Zástupkyně vedoucího:

RNDr. Terézia Slanínáková - Ústav výpočetní techniky, Masarykova univerzita;
slaninakova@ics.muni.cz

Cíle:

Návrh a vývoj jednotného frameworku pro pokročilé objevování (discovery), prohledávání (browsing) a vyhledávání (searching) dat v rámci jednoho nebo více repozitářů, založeného na:

1. sémantickém porozumění obsahu (anotace i samotná data) - schopnosti rozpoznat relevantní výsledky i při odlišné terminologii,
2. vhodné kombinaci sémantického přístupu s přístupem postaveném na metadatech neboli přesné shody termínů (např. autor, instituce),
3. identifikaci a integraci grafových vztahů (např. vztah autor-instituce-spoluautor), včetně pokročilé vizualizace pro exploraci těchto propojení,
4. integraci vysvětlitelnosti sémantických shod.

Umožnění integrace vyhledávacích a objevovacích mechanismů s AI nástroji (např. Claude Code) pro agentní přístup k službám nad repozitářovými systémy.

Výsledky a výstupy

Plánované výstupy:

- Framework a specifikace pro pokročilé služby objevování, prohledávání a vyhledávání dat v prostředí výzkumných datových repozitářů.
- Rozvoj prohledávacího portálu dostupného na discovery.eosc.cz.
- Referenční návrh integrace těchto služeb s AI agenty a analytickými workflow.
- V závislosti na časových a finančních možnostech členů TF rozvoj dalších specifických nástrojů a služeb.

Časový harmonogram a plán

Plánovaná délka trvání:

18 měsíců; od 05 / 2026

Plánované kroky:

1. Identifikace use-cases napříč PS EOSC CZ.
2. Identifikace aktuálně nasazených vyhledávacích přístupů v platformách datových repozitářů, včetně identifikace limitů vyhledávání na základě shody textů a metadatových záznamů a zhodnocení pokročilých metod práce s daty.

3. Sdílení know-how a případná společná implementace vybraných use-cases.
4. Návrh frameworku pro pokročilé objevování a vyhledávání dat s důrazem na interoperabilitu napříč repozitáři a integraci s AI systémy.
5. Návrh systému uživatelského testování pro sběr feedbacku a benchmarkování pro práci s kvalitou vyhledaných výsledků; exekuce testování a benchmarků na discovery.eosc.cz a případných dalších službách.
6. Návrh mechanismu financování, vývoje a dlouhodobého provozu produktu včetně plánu jeho implementace v zapojených infrastrukturách a spolupracujících infrastrukturních službách.

Účastníci

Členové:

Mgr. Adrián Rošinec - Ústav výpočetní techniky, Masarykova univerzita; adrian@muni.cz

Bc. Jakub Čillík - Ústav výpočetní techniky, Masarykova univerzita; cillik@muni.cz

Mgr. et Mgr. Šárka Portešová - Ústav výpočetní techniky, Masarykova univerzita;
485401@muni.cz

Odborníci na správu výzkumných dat, vyhledávací technologie a integraci AI se zkušeností s repozitářovými systémy, metadaty, zpracováním semantických dat a interoperabilitou v rámci EOSC CZ.