KÖZPONTI STATISZTIKAI HIVATAL

III. Országos HUN-REN Cloud Találkozó

A HUN-REN Cloud használatának bemutatása – eredmények, tapasztalatok a KSH szempontjából

Jáger Viktor
Módszertani és innovációs főosztály
KSH

2026. november 05.

Az együttműködés

KSH – MILAB – HUN-REN SZTAKI – HUN-REN Cloud

01

KSH – MILAB

2025. május Együttműködési megállapodás

Olyan együttműködés létrehozása a felek között, amely a Nemzeti Laboratórium munkájának megismerésén alapszik, és a Nemzeti Laboratórium eredményeinek hasznosítását, a Nemzeti Laboratóriummal közös kutatás-innovációs kooperációt célozza.

02

KSH – HUN-REN SZTAKI

2025. december Együttműködési megállapodás

A szöveges adatok nómenklatúrába klasszifikálásának mesterséges intelligenciával (MI) és/vagy gépi tanulással történő támogatása céljából jön létre.

03

KSH → HUN-REN Cloud

2026. I. negyedév Beköltözés a felhőbe



A kiinduló helyzet

amire a HUN-REN Cloud megoldási lehetőséget nyújtott

01

Alapprobléma

TUDÁS – ADAT – SZÁMÍTÁSI KAPACITÁS
megfelelő együtt állását összehozni.

02

Peremfeltételek

- Törvényi megfelelés, adatvédelmi és információbiztonsági megfelelés
- Költségek
- IT lehetőségek
- Közérdek, társadalmi hasznosság



Adatosztályozás meghatározása (NKI)

Az adatosztály (F) a bizalmasság (B), a sértetlenség és a rendelkezésre állás (SR) értékeinek összege alapján dől el.

(Kiberbiztonsági tv. - 2024. évi LXIX. törvény Magyarország kiberbiztonságáról - Hatályos Jogszabályok Gyűjteménye)

B	SR	F	Felhős korlátozás	Lokációs korlátozás	Külföldi adattárolás esetén az adat hazai rendelkezésre állásának biztosítása	Külföldi adatkezelés vagy nem privát felhőszolgáltatás esetén titkosítás
1	1	**F1**	Nincs	Nincs	Nem szükséges	Nem szükséges
1	2	**F2**	EGT-n kívül harmadik fél által tanúsított nem privát felhőben, EGT-n belül nincs korlátozás	EGT	Szükséges	Nem szükséges
2	1			Nincs	Nem szükséges	Kötelező, amennyiben technológiailag lehetséges
2	2	**F3**	Magyarországon kívül harmadik fél által tanúsított nem privát felhőben, nem tanúsított nem privát felhőben csak Magyarországon belül	EGT-n belül	Szükséges	Kötelező, amennyiben technológiailag lehetséges
3	1				Nem szükséges	Kötelező
3	2	**F4**	Harmadik fél által tanúsított privát felhőben vagy kormányzati felhőben	Magyarországon belül	Nem releváns	Nem releváns
4	1					
4	2					

Szoftveres környezet kialakítása

Időigény és megvalósítási tapasztalatok

01

Szoftverek telepítése

4–6 hét

Egy fő, részmunkaidőben
Ismert vagy korábban használt eszközök

02

Biztonsági megfelelés

4 hónap

IT-biztonsági és egyéb követelmények ellenőrzése
Egyeztetések és közös feladatok a KSH informatikával

03

Rendszer- és eszközhasználati finomítások

További 1–2 hét a későbbi beállításokra és pontosításokra

04

A telepítés egyszerű volt,
a rendszerek jól használhatók.

Fő tapasztalatok

A fő nehézség: hálózati elérés, VPN és DNS
a szigorú biztonsági környezetben.
Időnként vezetői eszkaláció segítette az előrelépést.

Technikai csatlakozás

Hozzáférés a Coder fejlesztői környezethez

01 Jelenlegi hozzáférés

A KSH hálózatáról VPN-klienssel csatlakozunk az OpenStacken futó VPN-szerverhez. A Coder így jelenleg a KSH irodáiból érhető el.

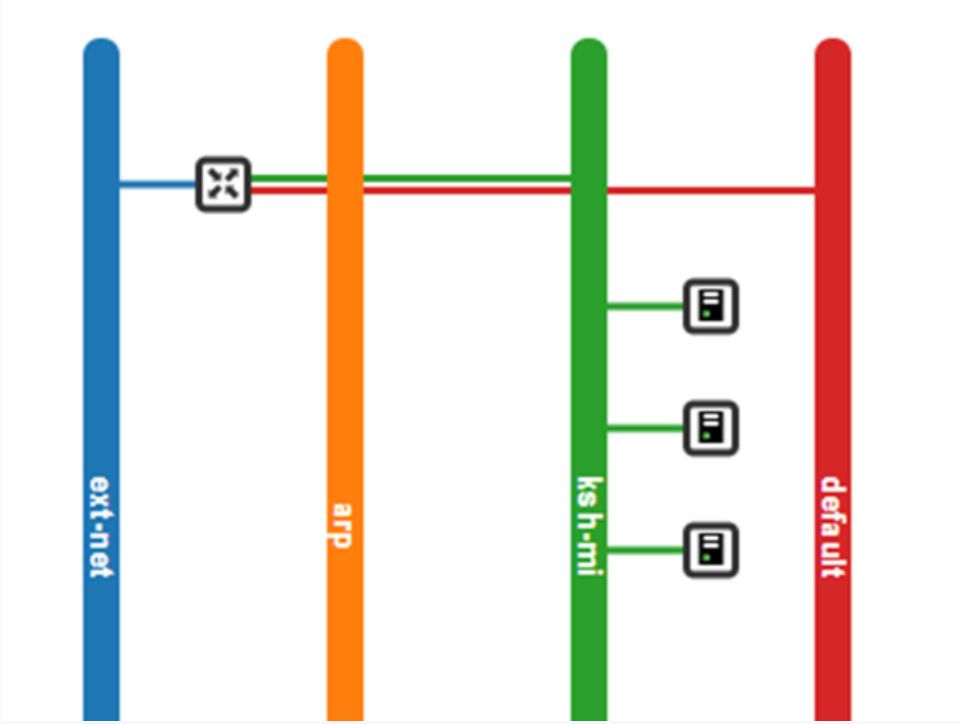
02 Site-to-site VPN tesztelése

Cél: a KSH hálózatáról külön VPN-kliens nélkül is elérhető legyen a Coder. A tesztelést személyi változások és tűzfalproblémák késleltették.

03 Független biztonsági minősítés

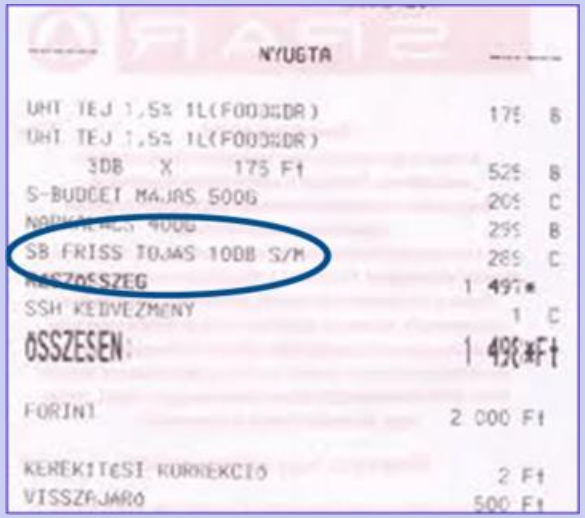
A felhőbe kerülő adatkörök kockázatelemzéséhez külső, független biztonsági minősítésre lenne szükség.

Hálózati felépítés



Automatizálás

Klasszifikációs modellek – COICOP



Feladat és cél

LLM-inferencia a nyugtán szereplő megnevezések alapján.
A termékek besorolása COICOP-kategóriákba.

Besorolási példa: friss tojás

01_1_4_8_1 Tyúk- és más madártojás héjában, frissen

Modell

1D-CNN **BERT** **ReLU** **Softmax**

Tanítóadatok

Több tízezer
korábban manuálisan
kódolt, címkézett adat

Találati pontosság

90% felett
a gyakran vásárolt
termékek esetében

Újrataníthatóság

Az újonnan bekódolt
tételekkel a modell
folyamatosan újratanul.

KÖZPONTI STATISZTIKAI HIVATAL

Információk lekérdezése
e-beszámolók kiegészítő mellékleteiből



Feladat

Teljesítménymutatók összegyűjtése és adatbázisba rendezése, például:

- Árbevétel
- ELÁBÉ
- Bérköltség
- Eredmény

Adatállomány

~400 000

kiegészítő melléklet PDF-formátumban

Tárolás külön meghajtón

Jelenlegi helyzet

Az információk dokumentumonként, egyesével történő kereséssel érhetők el.

Felhasználás

Az adatgyűjtésekből származó adatokkal való összehasonlítás.

KÖZPONTI STATISZTIKAI HIVATAL

Vállalatok beazonosítása az EGR-IS rendszerben



EGR

EuroGroups Register

Az Eurostat és az EU-tagállamok statisztikai hivatalainak hivatalos adatbázisa a multinacionális vállalatcsoportokról.

EGR-IS

EGR Identification Service

Jogi egységek azonosítása, egyedi azonosítók.
A találatok ellenőrzése és kiválasztása sok manuális utánajárást igényel.

Elkészült az információgyűjtő alkalmazás

North Data-adatokkal és Gemini-összefoglalókkal segíti és gyorsítja a vállalatokhoz kapcsolódó információgyűjtést.

Jelenlegi munka

Vállalati adatok tömeges feldolgozása

Következő lépés

AI-alapú döntési modell fejlesztése a HUN-REN CLoudban: a vállalatok azonosságának ellenőrzése és a megfelelő EGR-IS-találat kiválasztása.

Online álláshirdetések feldolgozása

Online Job Advertisement Eurostat Grant

01

Két adatforrás

Profession.hu: webscraping, havonta kétszer
KÖZSZOLGÁLLÁS: adatátvétel, havonta egyszer

Rendszeresen frissülő álláshirdetési adatok

02

Feldolgozás nyelvi modellekkel

Hirdetések strukturálása feldolgozható táblákká
FEOR-kódolás (foglalkozás)
Az elvárt iskolai végzettség azonosítása

03

Fő elemzési dimenziók

FEOR – Iskolai végzettség – Megye – Foglalkozási osztály

04

Negyedéves kísérleti statisztika

A két forrás külön és integrált elemzése
Egységes, elemzésre kész adatbázis

Összehasonlító vizsgálatok
hazai és nemzetközi eredményekkel

Felhasználói tapasztalatok

01

Modell futtatás és teljesítmény

Feladatonként akár tízezres nagyságrendű modellhívások (Ollamával), illetve feladatspecifikus modellek tanítása

A strukturált adatkinyerés időigényes

02

Hozzáférés

Jelenleg csak KSH belső környezetből érhető el.

A távoli hozzáférést biztosító S2S VPN kialakítása folyamatban van.

03

Három szereplős operatív működés

HUN-REN Cloud – KSH MI – KSH IT

04

Webes VS Code: megszokott és rugalmas környezet

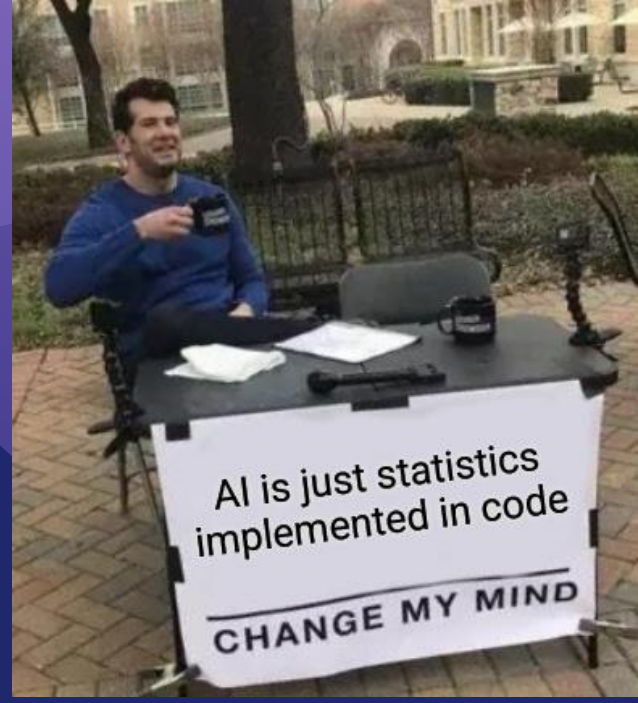
Ismerős felület és eszköztár
GitHub: fejlesztések szinkronizálása több gép között

Könnyen elérhető uv és Python-csomagok
Saját Docker / Coder környezet:
új eszközök rugalmas telepítése és használata

KÖZPONTI STATISZTIKAI HIVATAL

Artificial Intelligence =

linear algebra +
(composite ∘ functions)'
on high-dimensional vectors



Köszönöm

Jáger Viktor
Módszertani és innovációs főosztály
KSH
viktor.jager@ksh.hu
Viktor Jáger | LinkedIn

III. Országos HUN-REN Cloud Találkozó
2026-11-05