TECHNOLOGICKÉ CENTRUM PRAHA

ČÁST 3: POKROČILÉ TECHNIKY A KONCEPTY

Kristýna Meislová & Adéla Kučerová

Seminář pro RIS3 analytičky a analytiky, 4. 3. 2026

Můžete se ptát i online.

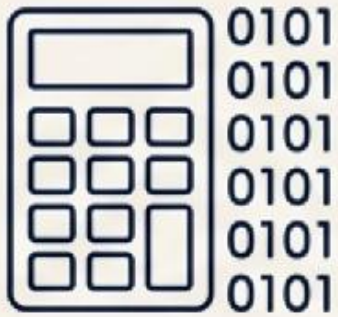
slido.com
kód # 1820 4031

AGENDA

- Sémantické vektory, využití vektorové databáze a RAG
- Strukturování nestrukturovaných dat
- Sklízení dat z webu pomocí agenta
- Komplexní workflow – demonstrační projekt

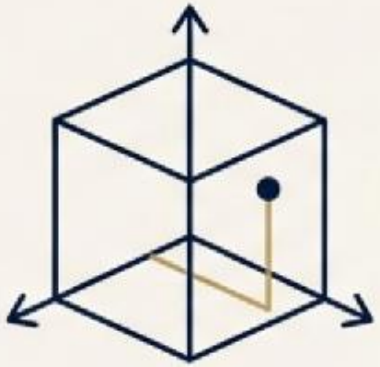
SÉMANTICKÉ VEKTORY

- Způsob, jak zachytit význam slov v číslech



Překlad na čísla

Počítač neumí číst písmena. Každé slovo převede na dlouhou řadu čísel (souřadnic).

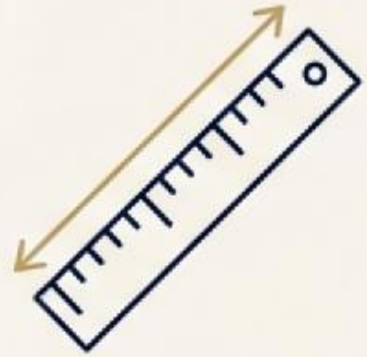


Souřadnice v prostoru

Tato čísla určují přesnou polohu slova ve vícerozměrném virtuálním prostoru.

Kontext

Slova, která se používají v podobných větách, leží v tomto prostoru blízko sebe.



Měření vzdálenosti

Počítač nehledá shodu textu, ale měří matematickou vzdálenost mezi myšlenkami.

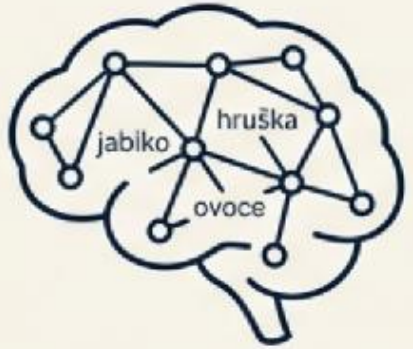
SÉMANTICKÉ VEKTORY

Minulost - Klíčová slova

Počítač vidí pouze řetězec znaků ‘j-a-b-l-k-o’. Pokud hledáte ‘ovoce’, počítač slovo ‘jablko’ nenajde, protože text je jiný.

0 % pochopení významu

Současnost - Vektory



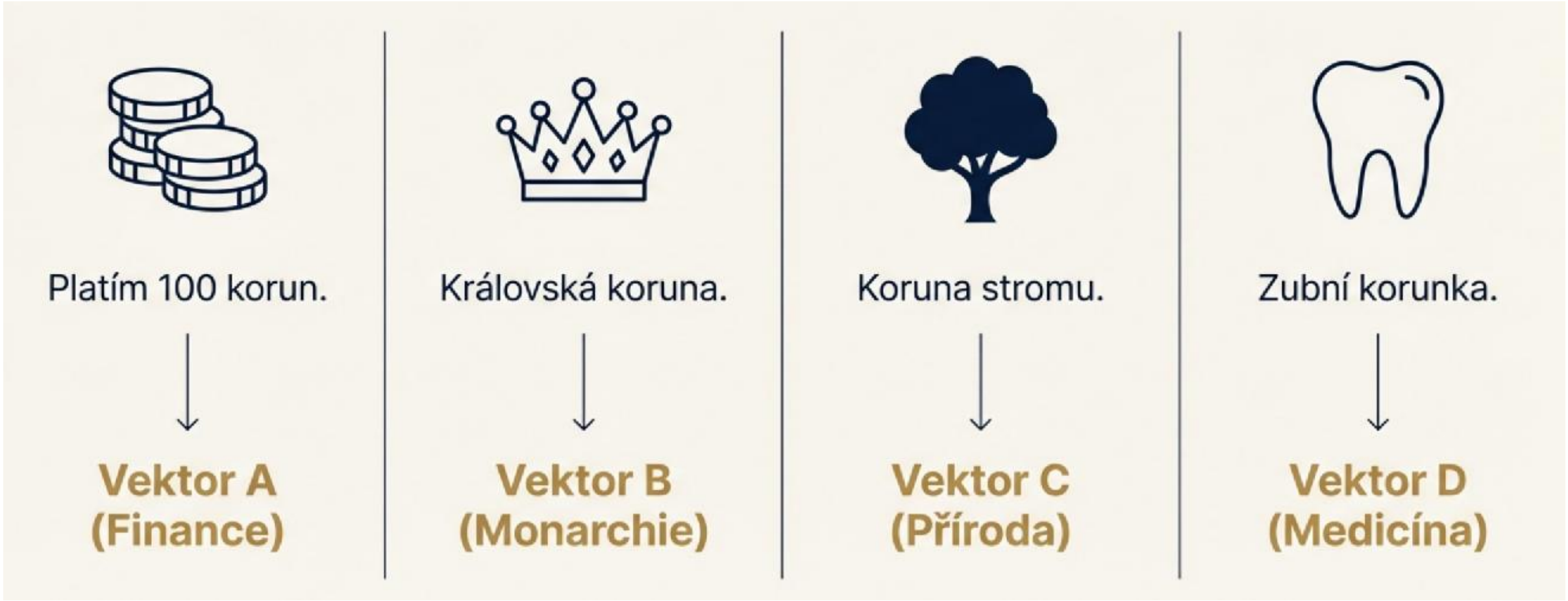
Počítač vidí souřadnice významu. Vektor slova ‘jablko’ leží v prostoru těsně vedle vektoru ‘hruška’ a ‘ovoce’.

100 % kontextuální shoda

SÉMANTICKÉ VEKTORY

SÉMANTICKÉ VEKTORY

- Kontext určuje souřadnice sémantického vektoru



SÉMANTICKÉ VEKTORY

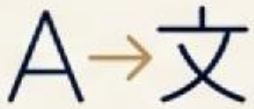
Tato technologie běží na pozadí aplikací, které používáte denně.

Chytré vyhledávání

Najde to, co myslíte, i když to napíšete nepřesně nebo použijete synonyma.

Doporučování obsahu

Líbilo se vám X, bude se vám líbit Y – protože vektory těchto filmů leží blízko sebe.



Strojový překlad

Nepřekládá slovo od slova, ale mapuje význam celé věty do jiného jazyka.

Generativní AI

Předvídá následující slovo na základě vektorové pravděpodobnosti a kontextu konverzace.

SÉMANTICKÉ VEKTORY

- Embedding věty: *„Industry needs to be more involved in innovation.“*

[-0.02227073162794113, 0.026993807405233383, -0.01209342759102583, -0.009084777906537056, 0.005109662190079689, -0.025144917890429497, -0.00309689249843359, 0.041045378893613815, -0.020825235173106194, 0.010816012509167194, 0.03963349759578705, -0.0059374612756073475, 0.03506169468164444, -0.005433218088001013, -0.01751404069364071, 0.02990160882472992, 0.034406181424856186, -0.004584409296512604, 0.05102939158678055, -0.013824662193655968, -0.007160250563174486, -0.03573402017354965, -0.052441272884607315, 0.027111465111374855, -0.013622964732348919, 0.03627187758684158, 0.012185872532427311, 0.018875496461987495, -0.07409010082483292, 0.05943344160914421, -0.010278153233230114, 0.009202434681355953, 0.03886032849550247, -0.06871151179075241, 0.009731889702379704, 0.016345877200365067, -0.001146102324128151, -0.015110481530427933, 0.01369019690901041, 0.0043070754036307335, 0.03818800300359726, -0.02785102091729641, 0.012471609748899937, 0.06571967154741287, 0.031313490122556686, -0.008072090335190296, -0.0137154096737504, -0.029044397175312042, -0.04040667042136192, 0.016160987317562103, 0.004949985537678003, -0.015597916208207607, 0.06891320645809174, 0.017698928713798523, -0.028641002252697945, 0.00043674797052517533, -0.003393135266378522, -0.04870987311005592, -0.004525580909103155, -0.009160414338111877, 0.0075342305935919285, 0.039969660341739655, 0.012043003924190998, 0.02482556365430355, 0.027279546484351158, -0.009958798997104168, -0.02319517731666565, 0.03936456888914108, -0.02786782942712307, 0.011026113294064999, -0.024405360221862793, -0.025212150067090988, 0.06568605452775955, 0.04857540875673294, 0.009563809260725975, -0.007118230219930 41, -0.012698519043624401, -0.015564300119876862, 0.020505880936980247, -0.009647849015891552, 0.03758291155099869, -0.017295533791184425, -0.004996207542717457, -0.018371252343058586, -0.036439958959817886, 0.047768618911504745, 0.053012747317552567, -0.049751974642276764, 0.01316074188798666, -0.009941991418600082, -0.024001967161893845, 0.03667527437210083, -0.010597506538033485, -0.03657442703843117, -0.0330447256565094, 0.04272618889808655, -0.05153363570570946, 0.010698355734348297, -0.0024413764476776123, -0.008324211463332176, 0.028288032859563828, 0.049819208681583405, 0.020253760740160942, 0.012732136063277721, 0.034406181424856186, 0.041247077286243 44, -0.04091091454029083, 0.023027095943689346, -0.010463042184710503, 0.08948632329702377, -0.009000737220048904, -0.014513794332742691, 0.03573402017354965, -0.011631204746663 57, 0.004244045354425907, 0.02744762785732746, -0.02316156215965748, 0.025413846597075462, 0.026102978736162186, 0.04491124302148819, -0.013463287614285946, -0.05472717434167862, -0.0009633142035454512, -0.005815602373331785, -0.033750664442777634, 0.03003607504069805, -0.030943712219595 91, 0.012404377572238445, 0.03519616276025772, -0.02324560284614563, -0.03963349759578705, 0.042658958584070206, 0.04165047034621239, -0.001012688037008047, 0.005618107505142689, -0.011488336138427258, 0.06619029492139816, 0.023598572239279747, -0.03926372155547142, -0.01751404069364071, 0.0639716312289238, -0.0008645666530355811, -0.03768375888466835, -0.02936374954879284, 0.03407001867890358, 0.001658749300986528 4, -0.019379738718271255, 0.05829048901796341, -0.014816340059041977, 0.07893083989620209, -0.03509531170129776, 0.019278891384601593, 0.034893617033958435, 0.001280567143112421, 0.026086170226335526, 0.005172692704945803, -0.000698061368893832, 0.010622719302773476, -0.06074447184801102, 0.0007800008752383292, 0.024052390828728676, -0.020858852192759514, -0.04645758867263794, 0.00761827128008008, -0.0016324867028743029, -0.07274545729160309, -0.0030233568977564573, 0.003367922967299819, -9.32980838115327e-05, 0.048810720443725586, -0.008097302168607712, 0.011269831098616123, 0.010303365066647 53, -0.03647357597947121, -0.008689787238836288, -0.01966547593474388, -0.02070757932960987, -0.04124707728624344, -0.028237607330083847, -0.0643077865242958, -0.0006754755158908665, -0.009975607506930828, -0.004958389326930046, 0.02264050953090191, -0.007521624676883221, 0.043129585683345795, 0.05180256441235542, -0.008437666110694408, 0.013337227515876293, -0.005836612544953823, 0.058593034744262695, -0.023901117965579033, -0.04534825310111046, -0.004571803379803896, -0.013017873279750347, -0.028674617409706116, -0.010261344723403454, 0.02010248787701 13, -0.009412535466253757, -0.005496248602867126 5, 0.003903681179508567, 0.03252367302775383, 0.020758002996444702, -0.0016797594726085663, -0.002598952502012253, 0.003601135453209281, -0.025094492360949516, 0.032893452793359756, 0.011026113294064999, 0.030859671533107 758, -0.031145408749580383, -0.04292788729071617, 0.01085803285241127, 0.007849382236599922, 0.001064162817783653 7, 0.022825399413704872, -0.01601811870932579, 0.031212640926241875, -0.059063661843538284, -0.01761488802731037, -0.021060548722743988, -0.008748616091907024, 0.030826054513454437, -0.018656991422176 36, 0.03423810005187988, -0.018203172832727432, 0.0052147130481898785, -0.018875496461987495, 0.021480752155184746, 0.017883818596601486, -0.015606320463120937, ***... a ještě 1297 souřadnic ...]***

UKÁZKA VYUŽITÍ VEKTOROVÉ DATABÁZE

- V TC využíváme
  - model text-embedding-3-large
  - s vektory o 1536 dimenzích

- Místo tradičních vyhledávání textových řetězců

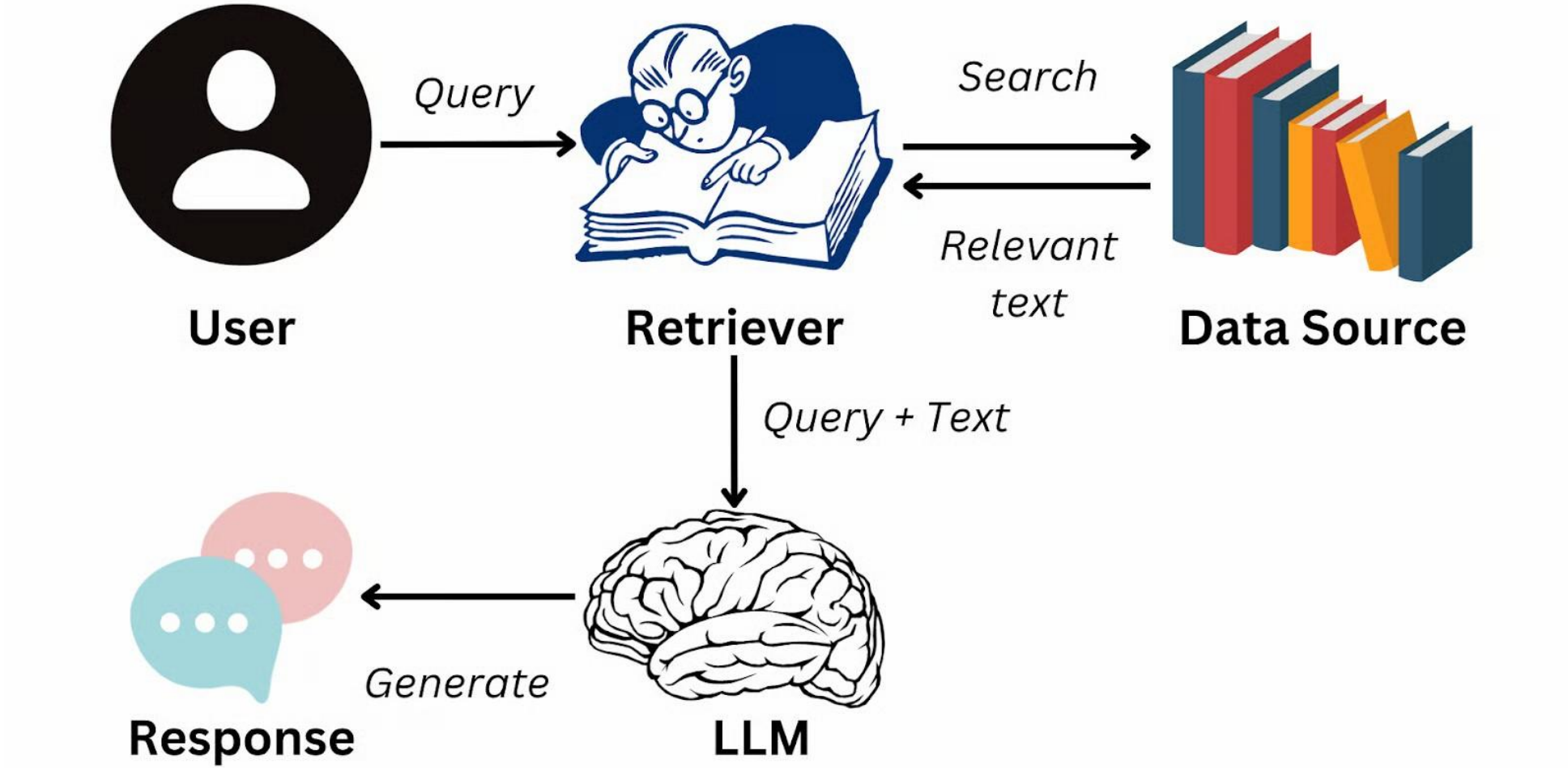
```
select * from projekty_lex pl
where pl.nak ilike '%sustainable%'
;

select * from projekty_lex pl
where pl.lex @@ plainto_tsquery('english', 'sustainable')
;
```

- Hledání blízkých výsledků ve vektorovém prostoru

```
select e.przidk, pl.nak, e.embedding <=> [vektorizované 'sustainable'] as distance
    -- značka <=> je pro cosine distance
from is_vavai_2_projekty_embedding e
left join projekty_lex pl on pl.przidk = e.przidk
order by distance
;
```

RETRIEVAL-AUGMENTED GENERATION (RAG)



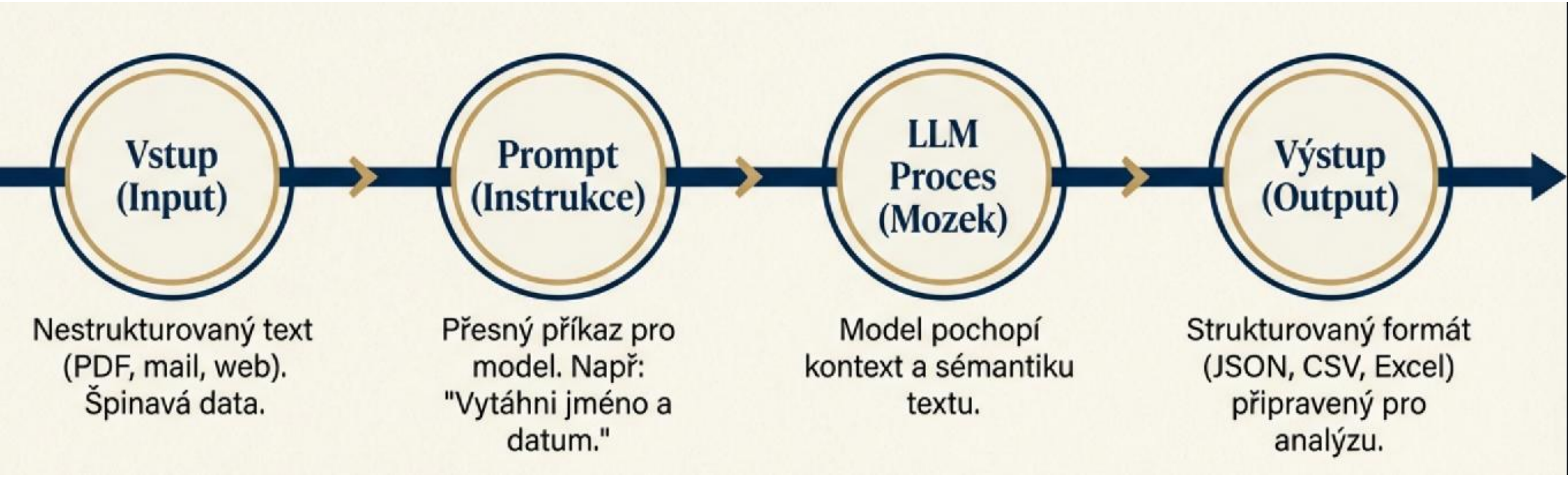
Zdroj: https://www.datacamp.com/blog/what-is-retrieval-augmented-generation-rag

RETRIEVAL-AUGMENTED GENERATION (RAG)

- RAG vlastně používáme sami často, když do AI chatu nahráváme nějaké dokumenty jako podklady ke konverzaci
- V pravém slova smyslu je však RAG technika reagující na omezenou velikost kontextového okna - LLM si vytáhne z externí databáze zdrojů to, co nejlépe odpovídá dotazu – tj. to, co je sémanticky nejbližší k dotazu

Zdroj: https://www.datacamp.com/blog/what-is-retrieval-augmented-generation-rag

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT



STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT

- *Může LLM klasifikátor podat kvalitnější výkon než živý člověk?*

  ❌ Ne, LLM bude řešit obdobná dilemata u hraničních případů jako člověk a navíc dělá chyby

  ✅ Ano, přistoupí k řešení dilemat u hraničních případů více systematicky, chybovost lze snížit kvalitním promptováním a výsledky budou zpracovány rychle

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

Dej mi aktuální a kompletní doporučení, jak vytvořit optimální zadání pro provedení tohoto úkolu:

Kategorizace ekonomických činností v dokumentech podle klasifikace NACE (v příloze). Jedné činnosti lze přiřadit více kódů NACE.

	A	B	C	D
1	CZ-NACE 2025 s opravami zavedenými oproti nařízení Komise (EU) 2023/137 a sdělení ČSÚ č. 400/2024 Sb.			
2	ÚROVEŇ	KÓD	NÁZEV POLOŽKY - ČESKY	NÁZEV POLOŽKY - ANGLICKY
3	1	A	ZEMĚDĚLSTVÍ, LESNICTVÍ A RYBÁŘSTVÍ	AGRICULTURE, FORESTRY AND FISHING
4	2	01	Rostlinná a živočišná výroba, myslivost a související činnosti	Crop and animal production, hunting and related service activit
5	3	01.1	Pěstování plodin jiných než trvalých	Growing of non-perennial crops
6	4	01.11	Pěstování obilovin jiných než rýže, luštěnin a olejnatých semen	Growing of cereals, other than rice, leguminous crops and oil se
7	5	01.11.0	Pěstování obilovin jiných než rýže, luštěnin a olejnatých semen	Growing of cereals, other than rice, leguminous crops and oil se
8	4	01.12	Pěstování rýže	Growing of rice
9	5	01.12.0	Pěstování rýže	Growing of rice
10	4	01.13	Pěstování zeleniny a melounů, kořenů a hlíz	Growing of vegetables and melons, roots and tubers
11	5	01.13.0	Pěstování zeleniny a melounů, kořenů a hlíz	Growing of vegetables and melons, roots and tubers
12	4	01.14	Pěstování cukrové třtiny	Growing of sugar cane
13	5	01.14.0	Pěstování cukrové třtiny	Growing of sugar cane
14	4	01.15	Pěstování tabáku	Growing of tobacco
15	5	01.15.0	Pěstování tabáku	Growing of tobacco
16	4	01.16	Pěstování přadných rostlin	Growing of fibre crops
17	5	01.16.0	Pěstování přadných rostlin	Growing of fibre crops
18	4	01.19	Pěstování ostatních plodin jiných než trvalých	Growing of other non-perennial crops
19	5	01.19.0	Pěstování ostatních plodin jiných než trvalých	Growing of other non-perennial crops
20	3	01.2	Pěstování trvalých plodin	Growing of perennial crops

```
CZ-NACE 2025 s opravami zavedenými oproti nařízení Komise (EU) 2023/137 a sdělení ČSÚ č. 400/2024 Sb.,,,
ÚROVEŇ,KÓD,NÁZEV POLOŽKY - ČESKY,NÁZEV POLOŽKY - ANGLICKY
1,A,"ZEMĚDĚLSTVÍ, LESNICTVÍ A RYBÁŘSTVÍ","AGRICULTURE, FORESTRY AND FISHING"
2,01,"Rostlinná a živočišná výroba, myslivost a související činnosti","Crop and animal production, hunting and related service activities"
3,01.1,Pěstování plodin jiných než trvalých,Growing of non-perennial crops
4,01.11,"Pěstování obilovin jiných než rýže, luštěnin a olejnatých semen","Growing of cereals, other than rice, leguminous crops and oil seeds"
5,01.11.0,"Pěstování obilovin jiných než rýže, luštěnin a olejnatých semen","Growing of cereals, other than rice, leguminous crops and oil seeds"
4,01.12,Pěstování rýže,Growing of rice
5,01.12.0,Pěstování rýže,Growing of rice
4,01.13,"Pěstování zeleniny a melounů, kořenů a hlíz","Growing of vegetables and melons, roots and tubers"
5,01.13.0,"Pěstování zeleniny a melounů, kořenů a hlíz","Growing of vegetables and melons, roots and tubers"
4,01.14,Pěstování cukrové třtiny,Growing of sugar cane
5,01.14.0,Pěstování cukrové třtiny,Growing of sugar cane
4,01.15,Pěstování tabáku,Growing of tobacco
5,01.15.0,Pěstování tabáku,Growing of tobacco
4,01.16,Pěstování přadných rostlin,Growing of fibre crops
5,01.16.0,Pěstování přadných rostlin,Growing of fibre crops
4,01.19,Pěstování ostatních plodin jiných než trvalých,Growing of other non-perennial crops
5,01.19.0,Pěstování ostatních plodin jiných než trvalých,Growing of other non-perennial crops
3,01.2,Pěstování trvalých plodin,Growing of perennial crops
```

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

1. Stáhni mi z internetu informace o ekonomických aktivitách následujících firem:

   - KOVOLIS HEDVIKOV a.s.
   - Teva Czech Industries s.r.o.
   - ŠKODA JS a.s. ,
   - Oracle Czech s.r.o.
   - STARTECH spol. s r.o.

   Výsledek exportuj do CSV se sloupci "nazev_firmy", "aktivita"

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

```
nazev_firmy,aktivita
KOVOLIS HEDVIKOV a.s.,"Tlakové lití hliníkových odlitků, výroba kovových součástí a kovodělných výrobků, povrchové úpravy a obrábění (slévárenství a výroba kovových polotovarů)"
Teva Czech Industries s.r.o.,"Farmaceutická výroba léčivých přípravků, farmaceutických substancí a doplňků - výroba a prodej humánních léčiv a kosmetických přípravků"
ŠKODA JS a.s.,"Inženýrské a výrobní služby pro jaderný průmysl - projekty, dodávky zařízení a servis pro jaderné elektrárny"
Oracle Czech s.r.o.,"Informační technologie - programování, vývoj a prodej softwarových řešení, IT poradenství a služby (IT hardware a software)"
STARTECH spol. s r.o.,"Strojírenská výroba - přesné CNC obrábění, výroba složitých kovových dílců a montáže pro letecký, vědecký a průmyslový sektor"
```

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

2. Dej mi aktuální a kompletní doporučení, jak vytvořit optimální zadání pro provedení tohoto úkolu:

   Kategorizace ekonomických činností v dokumentech podle klasifikace NACE (v příloze). Jedné činnosti lze přiřadit více kódů NACE.

Kritérium	ChatGPT (Prompt 1)	Claude (Prompt 2)	Gemini (Prompt 3)
Role / persona	Odborník na regulační ekonomickou klasifikaci	Odborný ekonomický analytik specializující se na evropskou klasifikaci	Odborný analytik evropských ekonomických dat a specialista na klasifikaci
Struktura promptu	Plochá – cíl, pravidla, formát	Plochá – pravidla, formát, postup	Strukturovaná – XML-like tagy (<cíl>, <pravidla>, <output_format>)
Počet klasifikačních pravidel	8	6	5
Omezení počtu kódů	Max. 5 kódů	Bez omezení	Bez omezení
Primární vs. sekundární kódy	Ano – explicitně vyžaduje 1 primární kód	Ne – žádná hierarchie	Ne – žádná hierarchie
Výstupní formát	JSON s primary_code, secondary_codes, excluded_considered_codes, confidence_score	JSON s activity_description, nace_codes (kód, label, confidence, rationale), notes	JSON pole s document_reference, extracted_activity_description, reasoning, nace_codes
Confidence score	Globální číslo (0.0–1.0) pro celý dokument	Per-kód úroveň jistoty (high/medium/low)	Žádný
Vyloučené kódy	Ano – sekce excluded_considered_codes	Ne	Ne
Postup / chain-of-thought	Ne – není explicitně popsán	Ano – 5-krokový postup na konci	Ano – „krok za krokem" + závěrečná fráze „Zhluboka se nadechněte..."
Vícejazyčné / formátovací pokyny	Ne	Ne	Ano – výslovně zakazuje markdown v outputu
Upozornění na nedostatečné informace	Ano – vrátí INSUFFICIENT_INFORMATION	Ne	Ne
Zaměření na vícenásobné označení	Spíše omezující (max. 5)	Explicitně povzbuzuje více kódů	Explicitně vyžaduje („je VYŽADOVÁNO")

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

Jste odborný ekonomický analytik specializující se na evropskou klasifikaci průmyslových činností.

Byla vám poskytnuta kompletní klasifikace CZ-NACE 2025. Vaším úkolem je analyzovat popisy ekonomických činností a přiřadit jim všechny příslušné kódy NACE.

Pravidla:

- Přiřaďte každé činnosti jeden nebo více kódů NACE – neomezujte se na jeden kód, pokud se na danou činnost vztahuje více kódů.

- Použijte nejkonkrétnější úroveň, která je k dispozici (upřednostněte čtyřmístnou třídu; pouze v případě nedostatečné podrobnosti použijte třímístnou skupinu nebo dvoumístnou divizi).

- Pokud činnost zahrnuje více sekcí (např. výroba A maloobchod), přiřaďte kódy z každé příslušné sekce.

- V nejednoznačných případech uveďte všechny možné kódy a stručně vysvětlete nejednoznačnost.

- Nikdy nevymýšlejte kódy – používejte POUZE kódy uvedené v přiložené klasifikaci.

- Výstupní formát: strukturovaný JSON (viz níže)

- Oznámkujte konfidence zařazení na stupnici: high / medium / low

Výstupní formát pro každou činnost:

```
{
  "activity_description": "...",
  "nace_codes": [
    {"code": "10.13", "label": "Production of meat and poultry meat products", "confidence": "high", "rationale": "..."},
    {"code": "46.32", "label": "Wholesale of meat and meat products", "confidence": "medium", "rationale": "..."}
  ],
  "notes": "Any ambiguity or assumptions made"
}
```

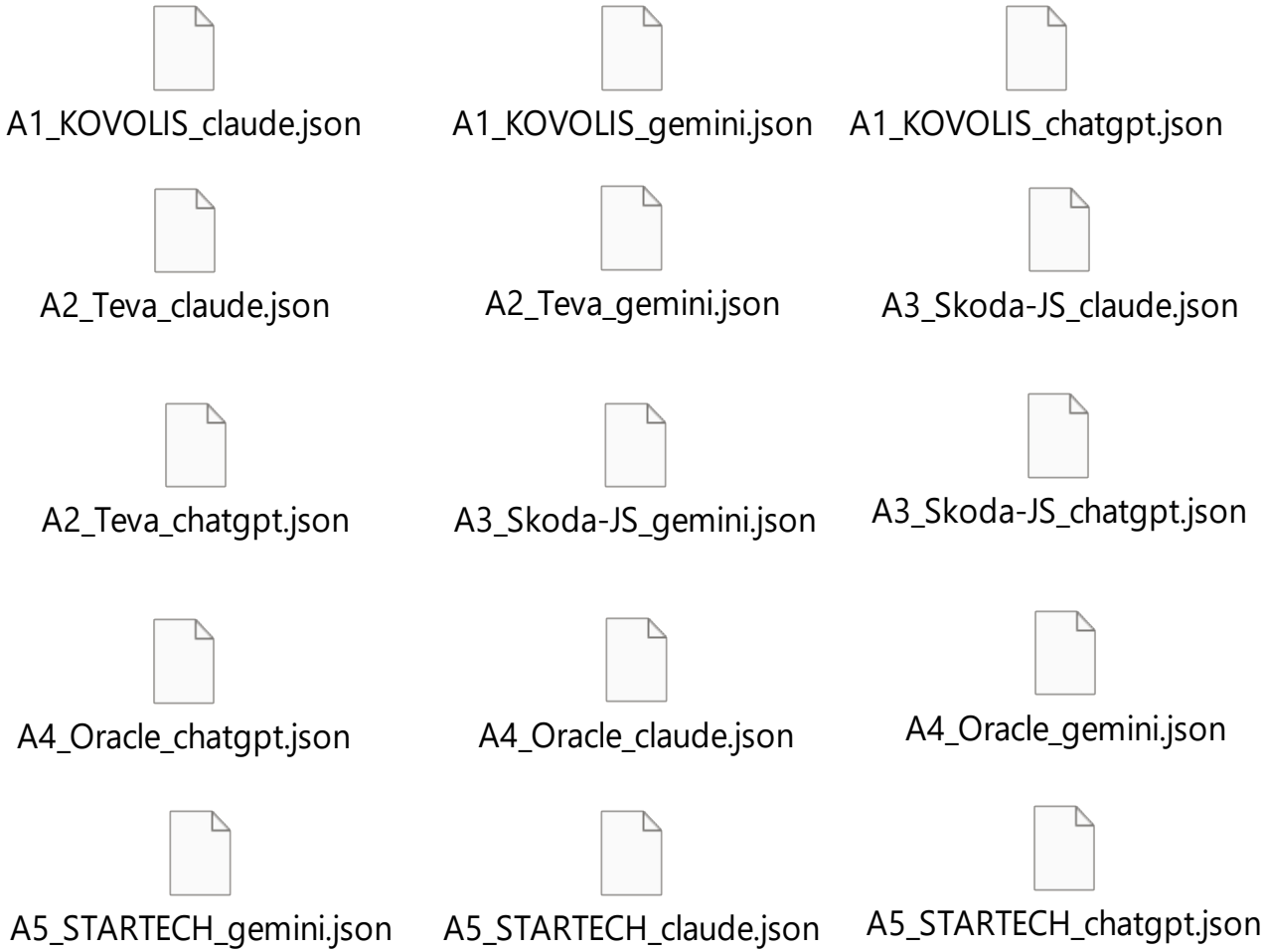
Zde je dokument pro klasifikaci. Identifikujte všechny uvedené ekonomické činnosti a přiřaďte jim kódy CZ-NACE 2025:

<input_dokument>

[VLOŽTE DOKUMENT ZDE]

</input_dokument>

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA



STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

Třem LLM jsem zadala úkol kategorizovat popis činnosti firmy podle klasifikace CZ-NACE 2025. Součástí promptu byla příloha s excelovou tabulkou klasifikace (V případě Gemini to byl txt soubor s CSV tabulkou).

Tady je prompt, který byl využit:

<input_prompt>

</input_prompt>

A nyní ti předám soubory JSON s výsledky. Každý LLM zpracovával 5 úloh (přiřazení kódů klasifikace NACE). Jedné úloze odpovídá jeden soubor JSON. Soubory jsou číslované podle pořadí úloh. V názvech souborů jsou za posledním podtržítkem názvy LLM.

Potřebuji od tebe do tabulky jednoduché shrnutí, jak se LLM s úkolem vypořádaly. Zejména mě zajímá:

- přesnost

- použitelnost výsledků

- analýza shody navzájem mezi LLM

Srovnání výsledků klasifikace CZ-NACE 2025 – Claude vs. Gemini vs. ChatGPT

Hodnotí se: přesnost kódů, konzistence nomenklatury, pokrytí činností, shoda mezi modely a použitelnost výstupu

Úloha	Firma	LLM	Přiřazené kódy (normalizované)	Počet kódů	Chybné / sporné kódy	Chybějící klíčové kódy	Hodnocení přesnosti	Hodnocení použitelnosti	Poznámky / Shoda
A1	KOVOLIS	Claude	24.53 · 25.99 · 25.51 · 25.52 · 25.53	5	—	—	★★★★★ Výborná	★★★★★ Výborná	Nejpřesnější pokrytí; správně odděluje 25.51 a 25.52; detailní rationale
		Gemini	24.53 · 25.51 · 25.53 · 25.99 · 25.40	5	25.40 (kování – neopodstatněné)	25.52 (tepelné zpracování)	★★★★ Dobrá	★★★★ Dobrá	Přidán 25.40 bez přímé opory v textu; jinak shodné s Claude
		ChatGPT	24.53 · 25.11 · 25.62 · 25.61 · 25.99 · 24.10	6	25.62 místo 25.53; 25.61 místo 25.51; 25.11 a 24.10 neopodstatněné	25.53; 25.51	★★ Slabá	★★★ Průměrná	Kódy 25.61 a 25.62 neexistují v CZ-NACE 2025 (existují jen 25.51–25.53); 24.10 zcela mimo téma
A2	Teva	Claude	21.10 · 21.20 · 46.46 · 47.73 · 46.45 · 47.75	6	—	20.42 (výroba kosmetiky)	★★★★ Dobrá	★★★★★ Výborná	Chybí 20.42 pro výrobu kosmetiky; obchodní kódy správně podmíněny
		Gemini	21.10 · 21.20 · 20.42 · 46.46 · 46.45 · 47.73 · 47.75 · 10.86	8	10.86 sporné (doplňky stravy)	—	★★★★ Dobrá	★★★★ Dobrá	Nejúplnější pokrytí; 10.86 je vzdálená možnost, ale lépe odůvodnit; správně zahrnuje 20.42
		ChatGPT	21.10 · 21.20 · 20.42 · 46.46 · 47.73 · 47.75	6	—	46.45 (velkoobchod kosmetika)	★★★★★ Výborná	★★★★★ Výborná	Čistý a přesný výstup; správně zahrnuje 20.42; chybí jen 46.45
A3	Škoda JS	Claude	71.12 · 28.99 · 33.12 · 33.20 · 46.64 · 35.11 · 72.10	7	35.11 irelevantní (pokud firma nevyrábí elektřinu)	—	★★★★★ Výborná	★★★★★ Výborná	Nejlepší odůvodnění; explicitně vysvětluje podmíněnost kódů 35.11 a 72.10
		Gemini	71.12 · 33.12 · 33.20 · 25.21 · 28.11 · 28.99 · 46.64	7	25.21 sporné (parogenerátory – blízké, ale 25.30 přesnější)	—	★★★★ Dobrá	★★★★ Dobrá	25.21 vs. 25.30 – metodická diskuse; jinak solidní pokrytí
		ChatGPT	71.12 · 25.30 · 28.99 · 33.20 · 33.12 · 42.22	6	42.22 okrajový (stavebnictví – jen pro EPC projekty)	46.64 (distribuce zařízení)	★★★★ Dobrá	★★★★ Dobrá	25.30 (parní kotle/jaderné reaktory) – přesnější než 25.21; 42.22 podmíněný
A4	Oracle	Claude	62.10 · 62.20 · 62.90 · 58.29 · 46.50 · 47.40	6	—	—	★★★★★ Výborná	★★★★★ Výborná	Korektní kódy; dobré rozlišení 62.10 vs 58.29; hardware kódy podmíněné
		Gemini	62.10 · 62.20 · 62.90 · 58.29 · 46.50 · 47.40 · 95.10	7	—	—	★★★★★ Výborná	★★★★★ Výborná	Stejné jako Claude + 95.10 (opravy HW) jako nízká konf.; použití podkódu 62.10.9 správné
		ChatGPT	62.01 · 62.02 · 62.09 · 58.29 · 46.51 · 47.41	6	62.01/62.02/62.09 – kódy neexistují v CZ-NACE 2025 (správně 62.10, 62.20, 62.90); 46.51/47.41 – nesprávná čísla	—	★★ Slabá	★★ Slabá	Záměna NACE Rev. 2 (starší) za CZ-NACE 2025; kódy jsou systematicky nesprávné
A5	STARTECH	Claude	25.53 · 25.99 · 30.31 · 33.20	4	—	—	★★★★★ Výborná	★★★★★ Výborná	Nejstručnější a nejpřesnější; výborně vysvětluje kdy použít 30.31 vs. 25.53
		Gemini	25.53 · 25.99 · 30.31 · 28.99 · 26.51	5	26.51 – nízká relevance pro 'vědecký sektor'	—	★★★★ Dobrá	★★★★ Dobrá	Přidány sektorové kódy; 26.51 diskutabilní bez hlubší opory v textu
		ChatGPT	25.62 · 30.30	2	25.62 neexistuje v CZ-NACE 2025 (správně 25.53); 30.30 – správně 30.31	25.99; 28.99; 33.20	★ Velmi slabá	★★ Slabá	Minimum kódů + nesprávná čísla; výrazně podpokrývá popsanou činnost

Celkové srovnání LLM – klasifikace CZ-NACE 2025

Kritérium	Claude	Gemini	ChatGPT	Vítěz	Komentář
Správnost kódů CZ-NACE 2025	☑ Vždy platné kódy	☑ Vždy platné kódy	⚠ 3/5 úloh s neplatnými kódy	Claude / Gemini	ChatGPT opakovaně používá kódy z NACE Rev.2 (62.01, 25.62, 47.41…), které v CZ-NACE 2025 neexistují
Pokrytí činností (úplnost)	Dobré až výborné	Nejúplnější (více alternativ)	Často neúplné	Gemini	Gemini systematicky přidává alternativní kódy s podmíněnou platností; ChatGPT často výrazně podpokrývá
Přesnost (bez falešných kódů)	Nejlepší – min. chybných kódů	Dobré – ojediněle sporné	Nejhorší – časté chyby	Claude	Claude vysvětluje podmíněnost kódů a vyhýbá se spekulativním přiřazením
Konzistence nomenklatury	☑ Konzistentní (4-místné kódy)	☑ Konzistentní (používá .0 suffix)	✗ Nekonzistentní (mísí verze NACE)	Claude / Gemini	ChatGPT mísí kódy různých verzí klasifikace bez rozlišení
Kvalita odůvodnění (rationale)	Výborná – detailní a přesná	Dobrá – stručnější	Průměrná – často zjednodušená	Claude	Claude poskytuje nejhlubší metodické vysvětlení; Gemini je stručnější ale správné
Práce s nejednoznačností	Výborná – explicitně komentuje	Dobrá – uvádí alternativy	Slabá – nejednoznačnost ignoruje	Claude	Claude nejlépe rozlišuje podmíněné kódy a vysvětluje kdy je použít
Použitelnost výstupu v praxi	★★★★★ Výborná	★★★★ Dobrá	★★ Slabá	Claude	Výstup ChatGPT nelze bez korekce použít; Claude je přímo nasaditelný
Shoda mezi modely	—	—	—	—	Všechny 3 modely se shodují na klíčových kódech (24.53, 21.10/21.20, 71.12, 25.53). Neshody nastávají při výběru sekundárních/podmíněných kódů
Celkové pořadí	🥇 1. místo	🥈 2. místo	🥉 3. místo	Claude	Claude nejlépe plní zadání – správné kódy, hloubka analýzy, praktická použitelnost

Analýza shody klíčových NACE kódů mezi Claude, Gemini a ChatGPT

Úloha	Kód NACE	Popis	Claude	Gemini	ChatGPT	Stupeň shody
A1 – KOVOLIS	24.53	Odlévání lehkých kovů	☑	☑	☑	Plná shoda
A1 – KOVOLIS	25.53	Obrábění kovů	☑	☑	✗ (25.62)	Částečná – ChatGPT neplatný kód
A1 – KOVOLIS	25.51	Povlakování kovů	☑	☑	✗ (25.61)	Částečná – ChatGPT neplatný kód
A1 – KOVOLIS	25.99	Ostatní kovové výrobky	☑	☑	☑	Plná shoda
A2 – Teva	21.10	Výroba farm. substancí	☑	☑	☑	Plná shoda
A2 – Teva	21.20	Výroba farm. přípravků	☑	☑	☑	Plná shoda
A2 – Teva	20.42	Výroba kosmetiky	✗	☑	☑	Částečná – Claude vynechal
A2 – Teva	46.46	Velkoobchod farm.	☑	☑	☑	Plná shoda
A3 – Škoda JS	71.12	Inženýrské činnosti	☑	☑	☑	Plná shoda
A3 – Škoda JS	33.12	Opravy strojů	☑	☑	☑	Plná shoda
A3 – Škoda JS	33.20	Instalace strojů	☑	☑	☑	Plná shoda
A3 – Škoda JS	28.99	Stroje pro spec. účely	☑	☑	☑	Plná shoda
A3 – Škoda JS	25.30	Parní kotle/jad. reaktory	✗	✗ (25.21)	☑	Neshoda – různé přístupy
A4 – Oracle	62.10	Programování	☑	☑	✗ (62.01)	Částečná – ChatGPT neplatný kód
A4 – Oracle	62.20	IT poradenství	☑	☑	✗ (62.02)	Částečná – ChatGPT neplatný kód
A4 – Oracle	62.90	Ostatní IT služby	☑	☑	✗ (62.09)	Částečná – ChatGPT neplatný kód
A4 – Oracle	58.29	Vydávání softwaru	☑	☑	☑	Plná shoda
A5 – STARTECH	25.53	Obrábění kovů	☑	☑	✗ (25.62)	Částečná – ChatGPT neplatný kód
A5 – STARTECH	25.99	Ostatní kovové výrobky	☑	☑	✗ (vynecháno	Částečná – ChatGPT vynechal
A5 – STARTECH	30.31	Výroba letadel	☑	☑	✗ (30.30)	Částečná – ChatGPT neplatný kód

Skóre souladu s RES a věcné správnosti (hodnocení 0–3 bodů na úlohu)

Úloha	Firma	Stav RES	Claude (shoda/věcnost)	Gemini (shoda/věcnost)	ChatGPT (shoda/věcnost)	Komentář k shodě
A1	KOVOLIS	☑ Správný	☑ Shoda (24.53 = 245)	☑ Shoda (24.53 = 245)	☑ Shoda (24.53), ale +chybné kódy	Plná shoda všech 3 LLM s RES. LLM úplnější o sekundární kódy.
A2	Teva	☑ Správný	☑ Shoda (21.20 = 21200)	☑ Shoda (21.20 = 21200)	☑ Shoda (21.20 = 21200)	Plná shoda. RES pokrývá jen hlavní kód; Claude vynechal 20.42.
A3	Škoda JS	! Chybný	☑ Věcně správnější než RES	☑ Věcně správnější než RES	⚠ Přiřadil 25.30 = kód RES (omylem správně?)	RES chybný (zbraně). LLM správně přiřadily inženýrské a výrobní kódy.
A4	Oracle	⚠ Chybný formát	☑ Správný kód (62.10)	☑ Správný kód (62.10)	✗ Neplatné kódy (62.01 = stará NACE)	RES i ChatGPT sdílejí chybu – staré NACE kódování. Claude/Gemini správně.
A5	STARTECH	! Chybný	☑ Věcně správnější než RES	☑ Věcně správnější než RES	✗ Neplatné kódy (25.62, 30.30)	RES chybný (elektro). Claude/Gemini správně 25.53. ChatGPT min. kódů + chybné.
CELKOVÉ SKÓRE (shoda s RES nebo věcná správnost)			**5/5 správně ☑**	**5/5 správně ☑**	**2/5 správně ✗**	

STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

- Začít rychle, vzniknou nová strukturovaná data spolu s nápady, jak je použít…
- Strategicky zvolit potenciálně zajímavá data a vyhodnotit vhodnou techniku a nástroje k jejich získání
- Nebát se učit…

SKLÍZENÍ DAT Z WEBU POMOCÍ AGENTA

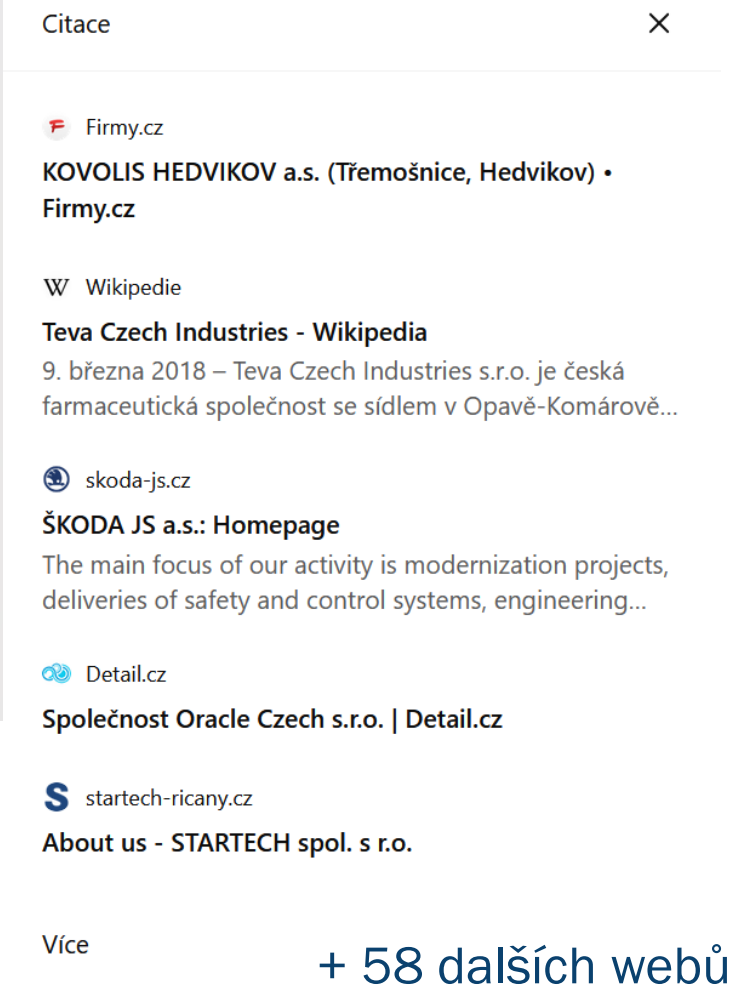
STRUKTUROVÁNÍ NESTRUKTUROVANÝCH DAT - UKÁZKA

1. Stáhni mi z internetu informace o ekonomických aktivitách následujících firem:

   - KOVOLIS HEDVIKOV a.s.
   - Teva Czech Industries s.r.o.
   - ŠKODA JS a.s. ,
   - Oracle Czech s.r.o.
   - STARTECH spol. s r.o.

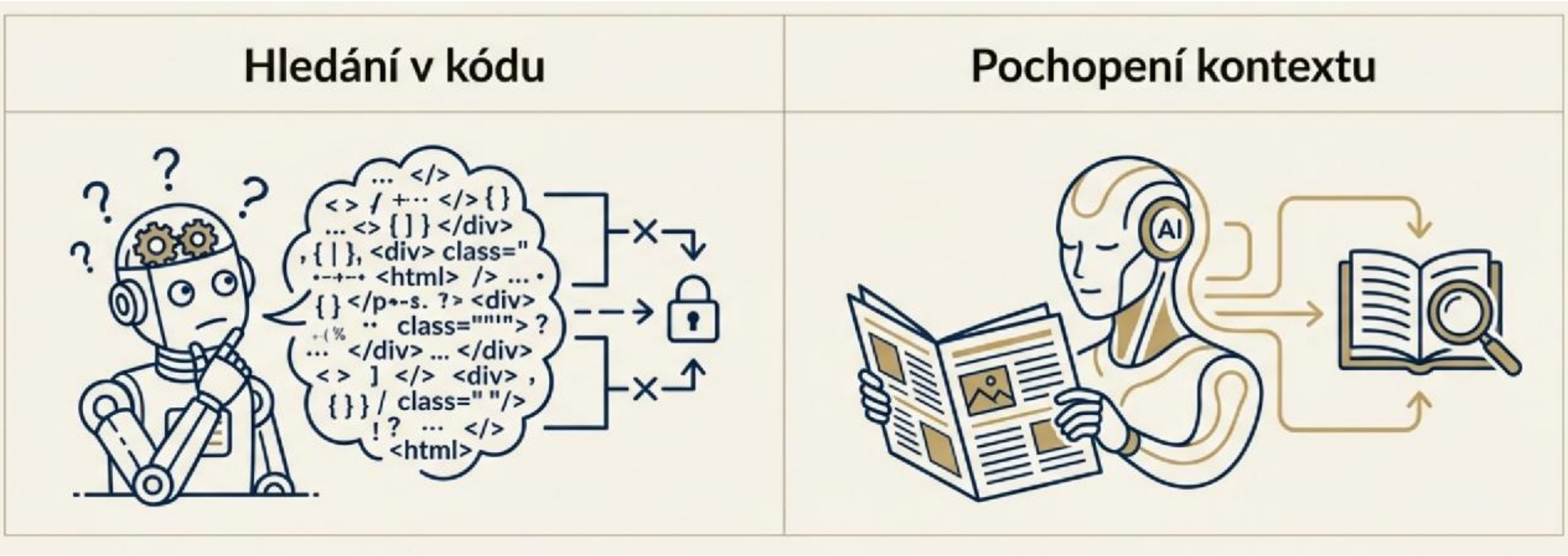
   Výsledek exportuj do CSV se sloupci "nazev_firmy", "aktivita"

17



+ 58 dalších webů

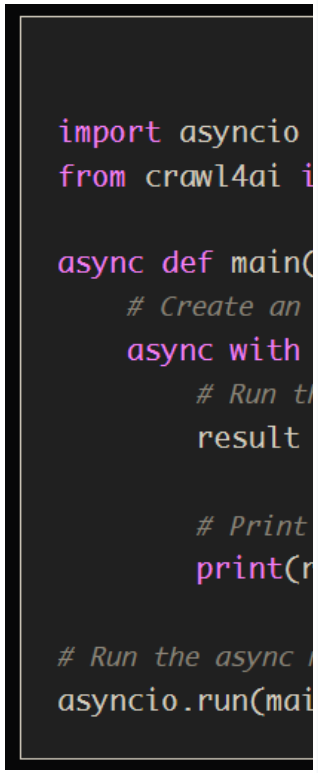
SKLÍZENÍ DAT Z WEBU POMOCÍ AGENTA



- Tradiční scrapping vyžadoval zkoumání HTML kódu a hledání konkrétních tagů (např. div.contact)
- AI čte web jako člověk, nezajímají ho tagy, ale význam

SKLÍZENÍ DAT Z WEBU POMOCÍ AGENTA

- Crawl4AI (Open Source) – stáhne stránku z webu a převede na markdown; běží lokálně
- Firecrawl – mapuje celý web, řeší anti-bot ochranu; cloudová služba
- Stagehand (Open Source) – AI interaktivně kliká na webu, příkazy nativním jazykem
- *Pydantic (Open Source) – způsob, jak dát datům z různých webů jednotný formát*




```python
import os

from playwright.sync_api import sync_playwright

from env import load_example_env
from stagehand import Stagehand


def main() -> None:
    load_example_env()

    with Stagehand(
        server="remote",
        browserbase_api_key=os.environ.get("BROWSERBASE_API_KEY"),
        browserbase_project_id=os.environ.get("BROWSERBASE_PROJECT_ID"),
        model_api_key=os.environ.get("MODEL_API_KEY"),
    ) as client:
        session = client.sessions.start(
            model_name="anthropic/claude-sonnet-4-6",
            browser={"type": "browserbase"},
        )

        cdp_url = session.data.cdp_url
        if not cdp_url:
            raise RuntimeError("No cdp_url returned from the API for this session.")

        with sync_playwright() as p:
            browser = p.chromium.connect_over_cdp(cdp_url)
            context = browser.contexts[0] if browser.contexts else browser.new_context()
            page = context.pages[0] if context.pages else context.new_page()

            client.sessions.navigate(session.id, url="https://news.ycombinator.com")
            page.wait_for_load_state("domcontentloaded")

            observe_stream = client.sessions.observe(
                session.id,
                instruction="find the link to view comments for the top post",
                stream_response=True,
                x_stream_response="true",
            )
```

KOMPLEXNÍ WORKFLOW

Od sběru dat přes analýzu po interaktivní vizualizace online

Ukážeme si demo projekt, který pokrývá témata:

- Využití asistenta kódu (Claude Code) pro netechnické, ale analyticky zdatné uživatele
- Webscraping
- Práce s nestrukturovanými daty
- Pipelining a řetězení úloh
- Validace a iterování
- Vizualizace
- Možnost automatizace
- Sdílení a týmová práce
- ...a další

KOMPLEXNÍ WORKFLOW

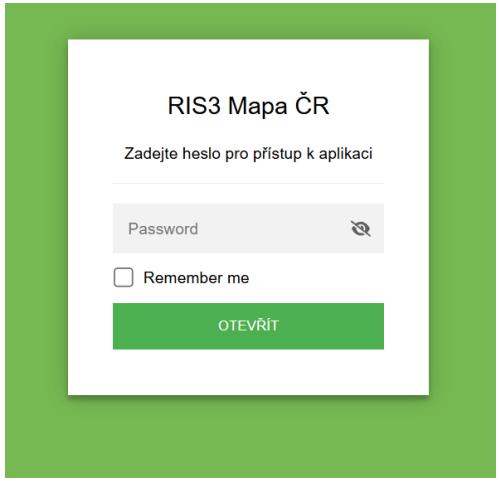
Disclaimer!!!

- Uvidíte interaktivní vizualizaci pohledu do krajských domén specializace. Celý projekt od vyhledání a zpracování dat, přes analýzu po web vznikl v dialogu s AI asistentem, a to záměrně bez jakékoliv přípravy předem (a bez hands-on programování)!
- Smyslem ukázky je demonstrovat, jak může vypadat cesta od analytického záměru k interaktivnímu webu pomocí AI asistenta. Záměrem je demonstrovat i chyby.
  - Data a vizualizace slouží k demonstraci workflow, ne jako seriózní analytický podklad
    - Analytické metody jsou ilustrativní, jiný analytik by legitimně zvolil jiné přístupy a možná dospěl k jiným výsledkům.
    - Vstupní data byla ověřena pouze zběžně, např. parsování PDF a propojování datasetů může obsahovat chyby, které se promítají do výsledků, žádný z výstupů neprošel metodologickou oponenturou.

KOMPLEXNÍ WORKFLOW

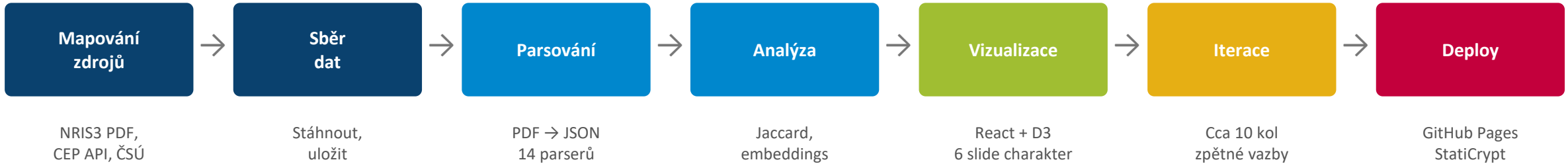
Funkční prototyp výsledku naleznete online na odkaze nebo použijte QR kód:

- https://kristyne.github.io/ris3-mapa-cr/
- *Heslo: RIS3seminarBrezen2026*



KOMPLEXNÍ WORKFLOW

Konkrétní pipeline projektu



Celý pipeline proběhl v dialogu člověka s AI , včetně řešení problémů a slepých uliček.

↑ Validace člověkem v každém kroku ↑

Jak si představit práci s asistentem kódu:

Chat s AI nástrojem, který vidí vaše soubory na disku.

Čte, píše, upravuje soubory. Spouští skripty.

Vy říkáte CO chcete (klidně v češtině), AI řeší JAK

Nemusíte umět programovat (ale hodí se rozumět principům kódování), stačí dobře formulovat, co chcete a vědomě řídit proces.

KOMPLEXNÍ WORKFLOW

Tento projekt vznikl kompletně v Claude Code

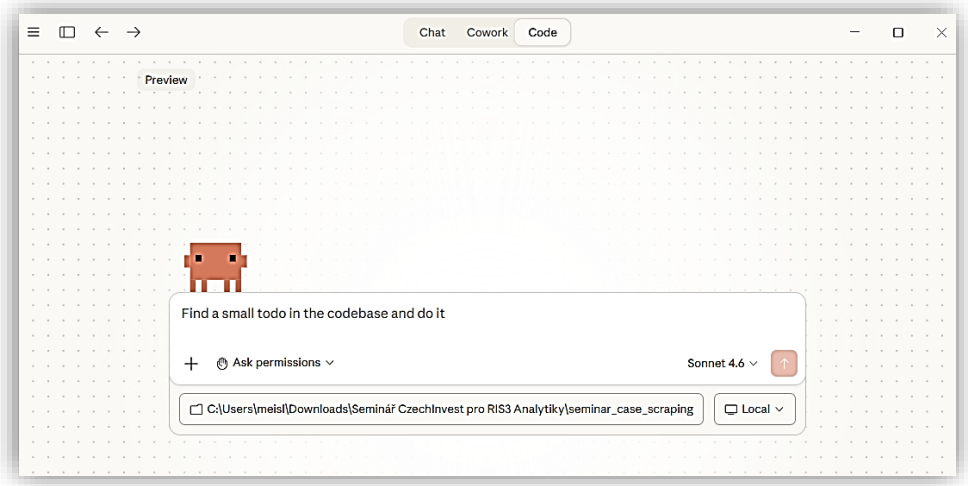
Claude Code

Příkazový řádek / VS Code / JetBrains / Desktopová aplikace

Plná kontrola — vidí soubory, píše kód, spouští skripty, deploy

Integrace s Git, GitHub Actions, MCP servery

Pro technicky orientované uživatele — maximální flexibilita



Claude Cowork

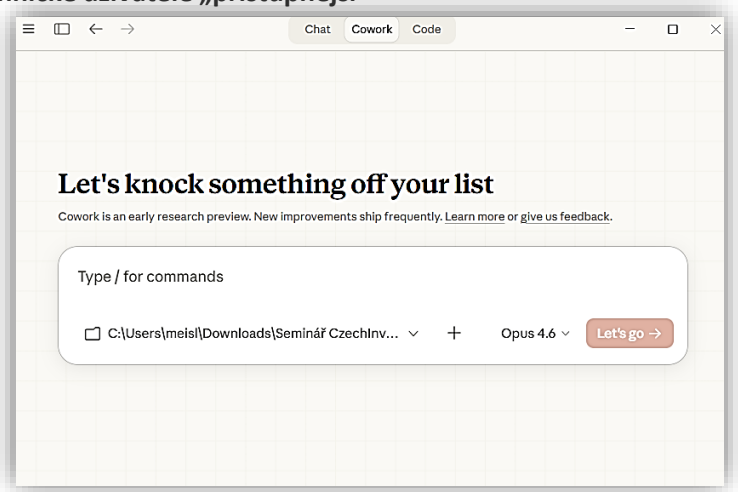
Desktopová aplikace s chat rozhraním (Win / Mac)

Stejné možnosti (soubory, skripty, analýza) bez terminálu

Generuje Excel, PowerPoint, pracuje s dokumenty

Pluginy: Google Drive, Gmail, DocuSign, FactSet…

Pro netechnické uživatele „přístupnější“



Oba nástroje běží lokálně, vidí vaše soubory, ke **kterým jim dáte přístup**. Pracujete v dané složce, kde AI vytváří různé typy souborů. Code pro plnou kontrolu, Cowork pro chatovací komfort a organizaci práce. Většinu kroků této ukázky lze provést i v Cowork.

KOMPLEXNÍ WORKFLOW

Tento projekt vznikl kompletně v Claude Code

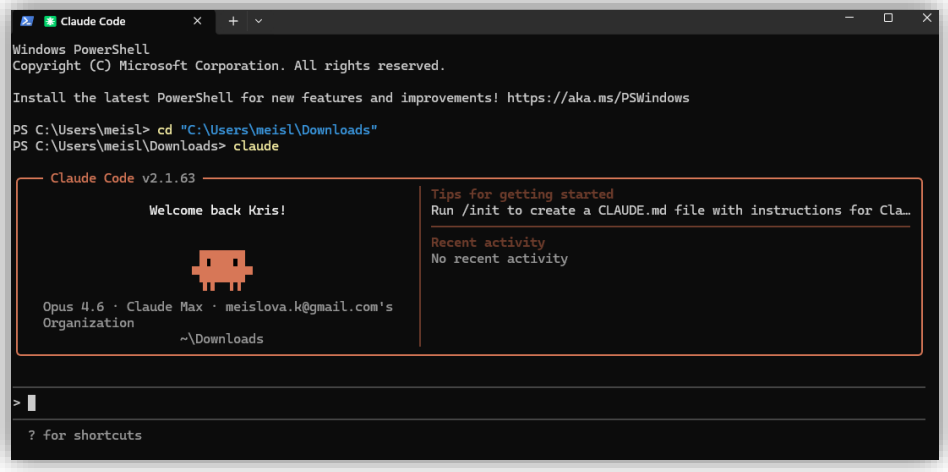
Claude Code

Příkazový řádek / VS Code / JetBrains / Desktopová aplikace

Plná kontrola — vidí soubory, píše kód, spouští skripty, deploy

Integrace s Git, GitHub Actions, MCP servery

Pro technicky orientované uživatele — maximální flexibilita



Claude Cowork

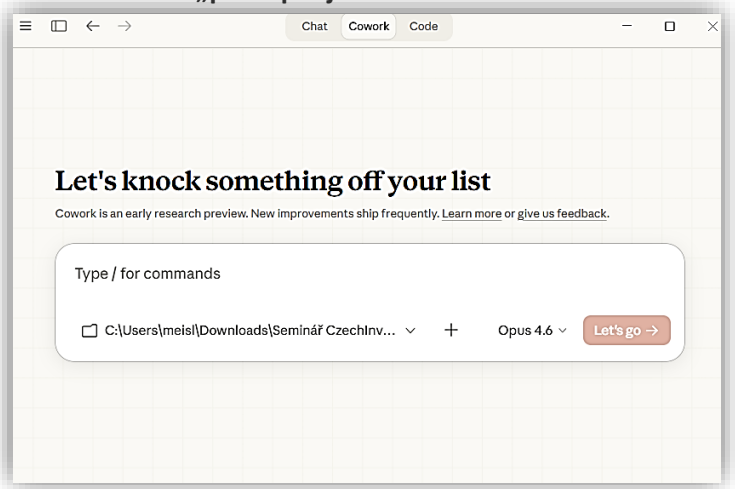
Desktopová aplikace s chat rozhraním (Win / Mac)

Stejné možnosti (soubory, skripty, analýza) bez terminálu

Generuje Excel, PowerPoint, pracuje s dokumenty

Pluginy: Google Drive, Gmail, DocuSign, FactSet…

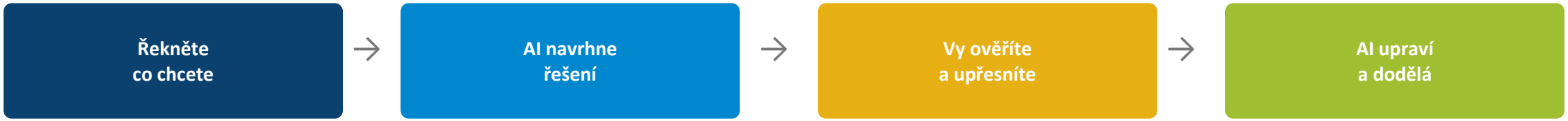
Pro netechnické uživatele „přístupnější“



Oba nástroje běží lokálně, vidí vaše soubory, ke **kterým jim dáte přístup**. Pracujete v dané složce, kde AI vytváří různé typy souborů. Code pro plnou kontrolu, Cowork pro chatovací komfort a organizaci práce. Většinu kroků této ukázky lze provést i v Cowork.

KOMPLEXNÍ WORKFLOW

Jednoduché vzorce, které fungují a platí pro všechny AI code asistenty



↩ opakovat dokud není hotovo

Nejdůležitější princip: dejte AI možnost ověřit výstup. Bez validace může chybovat. Dělejte testy, screenshoty, dávejte asistentovi konkrétní zpětnou vazbu.

Zopakujme si tři osvědčené vzorce (uvedené instrukce jsou zjednodušené!):

Řetězení kroků

Rozdělíte úkol na kroky.
Výstup kroku 1 → vstup 2.

```
"1. Stáhni data z API."
"2. Naparsuj do JSON."
"3. Udělej graf."
```

→ Data pipeline

Iterace se zpětnou vazbou

AI udělá první verzi.
Vy řeknete co změnit. Klidně 5x…20×.

```
"Udělej mapu."
"Barvy změň na brand."
"Legendu posuň."
```

→ Vizualizace, design

Dialog o metodách

AI navrhne přístup.
Vy se ptáte proč. Zvolte.

```
"Jak porovnat podobnost?"
"Proč embeddings?"
"Implementuj obojí."
```

→ Analytické úkoly

Zdroj: Anthropic "Building Effective Agents" — jednoduchost funguje lépe než složité frameworky

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 1: Zadání a hledání dat

PočátečnízZadání (záměrně obecné bez zdrojových dat):

```
"Chci analyzovat domény specializace krajských inovačních
strategií (RIS3) a z výsledků analýzy vytvořit interaktivní
online vizualizaci. Nejdříve najdi relevantní a aktuální datové
zdroje pro krajské RIS3 v ČR a prozkoumej je. Svá zjištění
popiš."
```

Co asistent našel a měl následně za úkol stáhnout a připravit pro analýzu:

✓ **NRIS3 v08 — Krajské karty**
PDF → parsování textu, 14 krajů, desítky domén

✓ **IS VaVaI / Starfos (TA ČR)**
Web scraping → JSON, 7 963 VaV projektů + 8 757 subjektů

✓ **ČSÚ — výdaje na VaV**
Excel (4 sektory, 2005–2024) → JSON konverze

✓ **ArcČR — geodata**
Stáhl rovnou GeoJSON hranice krajů a okresů, ale zjednodušil si ho na kraje

Ideální postup: začít v Plan Mode (Shift+Tab)

AI asistent analyzuje zadání, navrhne zdroje dat a strukturu projektu, ale nic ještě nestahuje, neskriptuje. Vy zkontrolujete plán, upřesníte co chcete, a až pak přepnete do běžného režimu. Je to prevence zbytečné práce, chyb, nežádoucích operací, nepochopení instrukce a pálení tokenů.

↑ Kde analytik validuje:

Existují zdroje a jsou aktuální? Pokud dále souhlasíme se stažením, obsahují soubory očekávaná data?

Pokrývají všech 14 krajů? Není v datech něco navíc (třeba národní strategie)?

Porovnat s vlastní odborností a doménovou expertízou!!!

Příklad problému: U 2 zdrojů získal asistent špatný URL, po upozornění našel správný. Problém s doménami je definiční (ponecháno na modelu), který ale nezná kontext problému v celé šíři (tzn. reálně by analytik*čka musel*a dodat platnou metodiku).

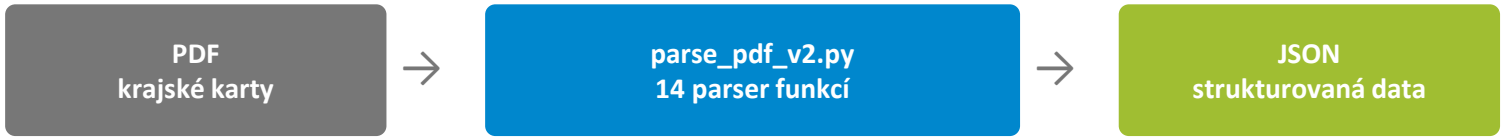
⚠ 4 zdroje, 4 různé formáty

PDF (nestrukturovaný text) · Web scraping (HTML/JSON) · Excel · GeoJSON
AI musel každý zdroj najít, stáhnout a převést do jednotného JSON formátu.

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 2: Parsování a iterativní opravy

1 PDF → 14 krajských karet → identifikovaných 91 domén specializace (záleží jak striktní v tom, co je doména specializace budeme) → 180+ NACE kódů, mnoho různých popisů




```
"Naparsuj PDF krajských karet NRIS3.
Z každé karty extrahuj: název domény,
NACE kódy, popis, klíčové firmy.
Ulož jako JSON. Každý kraj má jiný
formát – přizpůsob parser."
```

Reálné problémy při parsování:

- **Každý kraj má jiný formát karty!**
- **EN DASH (U+2013) vs. pomlčka v NACE regexích**
- **MSK: specifické end markery pro správné ukončení**
- **ZLK: horizontální layout potřeboval stop markery**
- **Windows kódování (cp1250) vs. UTF-8 výstupy**
- **Některé kraje vůbec nemají NACE kódy**

AI asistent zvládne i špatně formátované PDF, iterativně opravuje parser na základě výsledků.

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 2: Parsování a iterativní opravy

Reálná hrubá (sklizená) data jsou obvykle ošklivá.

Evoluce kódu:

Verze 1: generický parser (238 řádků)

Jeden parser pro všech 14 krajů → fungovalo jen pro ~7 krajů.

↓ iterace ↓

Verze 2: 14 specializovaných parserů (1 077 řádků)

Každý kraj svůj parser → identifikováno 91 domén, 180+ NACE.

Reálné chyby a jak je AI opravil:

EN DASH vs. pomlčka

MSK používá typografický EN DASH (U+2013).
Regex NACE kódů nefungoval.

```
nace_re = r'[\-
\u2013:]\s*(\d{2})'
```

Windows kódování

Česká diakritika se rozbíjela.

```
sys.stdout.reconfigure(encod
ing='utf-8')
```

ZLK: dva typy domén

Horizontální + vertikální specializace
v 3stupňové hierarchii.

Typický (zjednodušený) dialog při debugování:

MSK má jen 8 z 10 NACE kódů. Podívej se proč.

Našel jsem problém: PDF používá EN DASH místo pomlčky. Opravím regex.

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 3a: Volba analytických metod

AI navrhuje metody, vy rozhodujete a směřujete další postup. Je to dialog, který (zjednodušeně!) vypadá následovně:

Chci porovnat jak jsou si kraje podobné v doménách specializace. Navrhni přístup.

Navrhuji dva nezávislé přístupy:
1) Jaccard index přes NACE kódy — formální shoda oborů
2) Sémantická podobnost přes embeddings — obsahová shoda textů

Doporučuji prozkoumat oba přístupy, protože ukazují různé dimenze podobnosti. Jaccard nefunguje pro kraje bez NACE (4 ze 14). Pro sémantickou podobnost navrhni vhodné multilinguální modely pro embedding, které dobře zvládnou český jazyk.

Sémantický přístup funguje pro všech 14. Jaccard bude doplňková metrika pro 10 krajů.
Vizualizace označí kraje bez NACE (šedivě).

Souhlasím. Přidej propojení na VaV projekty z CEP a aplikuj dual matching. Tzn. zohledni dvě úrovně: FORD matching + sémantický embedding.

Výsledné skripty:

```
gen_embeddings.py     117 řádků
model: MiniLM-L12-v2
91 × 91 matice podobnosti
```

```
compute_vav_semantic.py  259 řádků
FORD + semantic matching
7 863 projektů → 4 kategorie shody
```

↑ Kde (by) obvykle proběhla další kola iterace a dialogu:

Výběr modelu pro embedding (významně ovlivní výsledky)

Kontrola hodnot podobnosti u skutečně příbuzných domén (kokrola lidskou expertízou)

Kontrola výsledků metod (kontrola lidskou expertízou, další analytická validace)

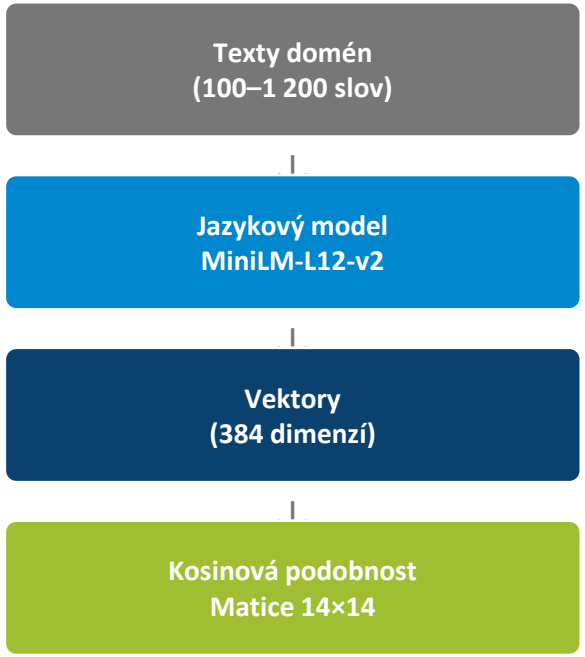
VaV matching: Přiřazené projekty odpovídají doménám?

Zkoušení thresholds (0.35 / 0.38) — lze snadno upravit a otestovat přímo v kódu

….a další…

KOMPLEXNÍ WORKFLOW

Ukázka: Krok 3b — Sémantická analýza (technický detail)




```
"Vygeneruj embeddings plných textů domén
modelem paraphrase-multilingual-MiniLM-L12-v2.
Spočítej kosinovou podobnost.
Threshold 0.35 pro sémantický match."
```

Co proběhlo na pozadí:

Model běží lokálně — žádné API ani login, žádné náklady, žádné odesílání dat

91 domén → 4 095 párů k porovnání

Výstup: semanticka_podobnost.json — matice + průměry per kraj

Aplikovaný vícejazyčný model je rychlý a zvládá češtinu bez problémů (pro experimentování ideální volba)

Kde validovat a přizpůsobit:

Threshold 0.35 — můžete změnit v gen_embeddings.py

Volba modelu — jiný model = jiné výsledky (např. větší model pro vyšší přesnost)

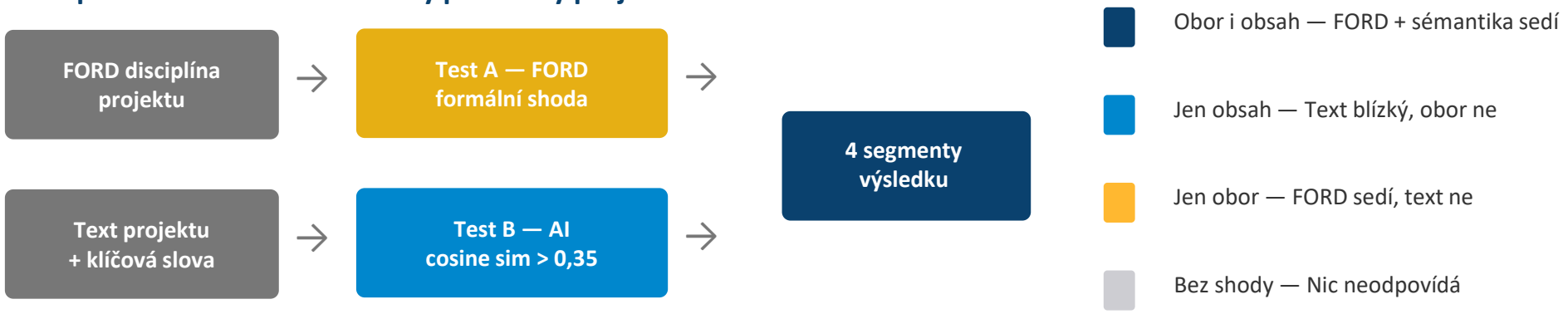
Spot-check: Nejpodobnější páry je vhodné validovat expertně. Dávají výsledky smysl doménově?

Stejně tak nejméně podobné páry. Opravdu jsou vzdálené? A tak dále…

KOMPLEXNÍ WORKFLOW

Ukázka: Krok 3c — Propojení VaV projektů s doménami (analytický detail)

Dva experimentální nezávislé testy pro každý projekt:



7 863 projektů CEP × 91 domén = ~715 000 porovnání

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 4: Vizualizace a iterace

Máme hotová analyzovaná data: GeoJSON mapy krajů a okresů (ArcČR), JSON s doménami specializace a jejich NACE kódy, matici sémantické podobnosti textů domén (z embeddings), VaV výdaje per kraj a sektor, a výsledky párového přiřazení VaV projektů k doménám. Teď potřebujeme výsledky převést do interaktivní prezentace.

```
"Výsledky analýz vizualizuj jako interaktivní webovou prezentaci. Formát: fullscreen slideshow, 6 slidů, přepínání šipkami/swipem/tečkami. Každý slide má:
hlavní vizualizaci přes většinu plochy, nadpis s podnadpisem vlevo nahoře, legendu, komentář k metodologii v malém panelu, rozklikávací tlačítko ⓘ s
podrobným vysvětlením, tooltipy na interaktivních prvcích a zdroj dat v patičce. Každý slide má jiný barevný odstín vizualizace, aby se vzájemně odlišily.
Tón neutrální – žádné rankingy ani doporučení, jen čísla a vztahy. Slidy: (1) titulka s disclaimerem a obrysem mapy ČR na pozadí, (2) choroplethová mapa
VaV intenzity s koláčovými grafy sektorů v centroidech krajů, (3) Jaccard heatmapa krajů podle sdílených NACE kódů, (4) sémantická podobnost textů domén –
choropleth + force-directed  síťový graf + scatter plot Jaccard vs. sémantika, (5) stacked bar chart párování VaV projektů s doménami (FORD + sémantika),
(6) závěr se shrnutím pipeline a poznatky.
```

Výsledek: 6 interaktivních slidů

1	**Titulní slide**	D3 mapa pozadí
2	**Výdaje na VaV mapa s koláči**	GeoJSON + pie charts
3	**Jaccard heatmapa**	Heatmapa + kartogram
4	**Sémantická podobnost**	Force-directed + choropleth
5	**VaV projekty vs. domény**	Stacked bar chart
6	**Závěr**	Pipeline shrnutí

Typická (zjednodušená!) iterace (screenshot nebo popis → úprava):

👤 Na slide XY do kartogramu přidej koláčové grafy per kraj. Grafy by měly vyjadřovat… zobrazovat…

🤖 Hotovo — slide s GeoJSON mapou + pie charts.

👤 [screenshot] Barvy změň na TC Praha paletu (#0A416E, #0087CD, ...) z dokumentu Logo_manual. Popis Středočeského kraje se překrývá s popisem Prahy. Posuň popis...

🤖 Barvy upravené. Centroid popisku posunut o +30,+35px.

📷 Tip: screenshot vizualizace přetažný do Claude = AI asistent uvidí co vy.
@ reference: "@src/slides/SlideMapVav.jsx uprav legendu"

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 4b: Responsivita

Problém:

Na mobilu se nic nenačítalo správně

Position:absolute boxy se na malých obrazovkách vrstvily přes sebe

```
"Na mobilu není aplikace responsivní. Např. legendy se
překrývají. Potřebuji tři breakpointy: desktop overlay,
tablet/ 13pc pod sebou, mobil collapsible panely. Vytvoř
reusable CollapsiblePanel komponentu."
```

Řešení — tři breakpointy (stále není ideální, ale pro rychlý fix ok):

Desktop >1024px	Panely vedle vizualizací (absolute overlay)
Tablet 768–1024px	Vše pod sebou v jednom sloupci (flex column)
Mobil <768px	Zkrácené legendy + rozbalovací panely

Reálný průběh vyžaduje opět několik iterací:

1. AI navrhl systematický přístup: tři layouty per slide + zahrnul CollapsiblePanel komponentu
2. Na 13" monitoru → stále překryvy → přidán isCompact flag pro height < 900px
3. Kontrola→ legendy stále v nehezkých pozicích → pozice upraveny o px (ale různé velikosti zobrazení opakují problémy) → rozhodnutí dále se tím (ne)zabývat

*Design je iterativní, několik kol zpětné vazby je obvykle třeba i v případě práce s AI asistentem, než je výsledek publikovatelný. Jako analytici*čky nemáme obvykle dostatek zkušeností s aplikacemi. Podle velikosti a významnosti projektu je vhodné zvážit, zda nepřenechat tuto část zkušenějším odborníkům na grafický design, frontend dev, UI apod. Tímto způsobem můžete odborníkům a vývojářlm obecně dodat konkrétnější představu o tom, co potřebujete.*

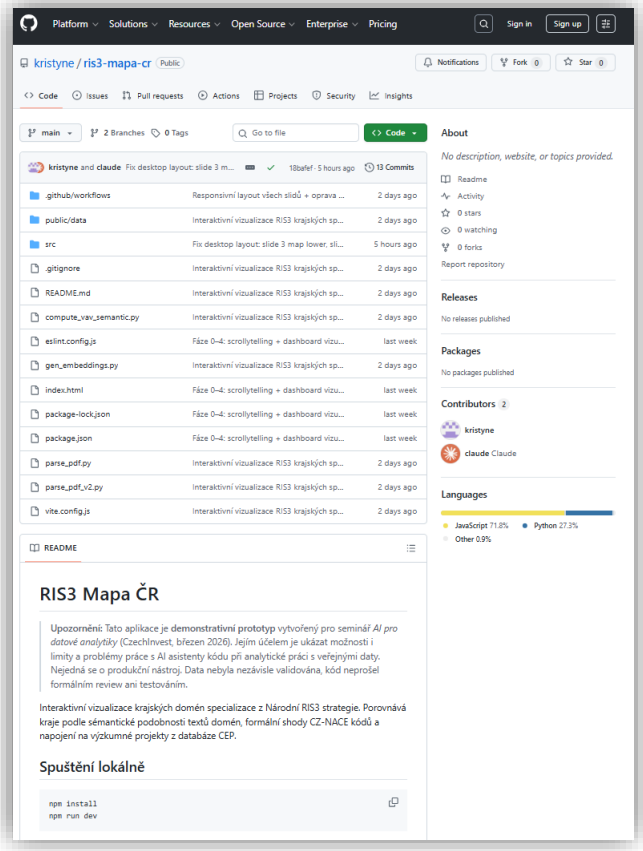
Validace na různých velikostech monitoru vám pomůže. AI nemůže vidět výsledek bez zpětné vazby.

KOMPLEXNÍ WORKFLOW

Ukázka — Krok 5: Základní deploy a sdílení výstupů


```
# .github/workflows/deploy.yml
steps:
  - npm ci && npm run build
  - npx staticrypt dist/index.html
      -p ${{ secrets.SITE_PASSWORD }}
  - cp dist/index.html dist/404.html  # SPA routing
  - actions/deploy-pages@v4
```

"Nasaď aplikaci na GitHub Pages.
Zašifruj heslem přes StatiCrypt.
Nastav deploy při push do main."



Kompletní dokumentaci ukázky naleznete v GitHub repositáři
https://github.com/kristyne/ris3-mapa-cr

Projekt je publikován veřejně a je dostupný všem.

- Propojení na Git server (GitHub není jediný), kde je možné hostovat a spravovat kód. Musíte mít zřízen vlastní účet.
- Vhodné pro sdílení, týmovou práci a synchronizování projektů, zálohování, podporuje vytváření nezávislých větví pro vývoj nových funkcí apod. Repositáře mohou být veřejné i soukromé.
- GitHub také pomůže s jednoduchým statickým webem (viz ukázka).
- **Seriózní produkční nasazení důležité aplikace by vyžadovalo další intenzivní práci, a to nejen analytickou, ale zejména vývojářskou (zaměřenou na robustní back-end, responsivitu, bezpečnost (!!!) a další aspekty podle zaměření aplikace.**

**sestavení (bundlování) webové aplikace pro produkční nasazení*
***StatiCrypt je nástroj (Node.js balíček), který slouží k zaheslování statických HTML stránek bez nutnosti backendu*

ZÁKULISÍ A DOKUMENTACE VÝVOJE

Jak si AI asistent kódu "pamatuje" a jak nastavit pravidla? Jde o institucionální paměť projektu, která je dostupná v dokumentaci (ve složce projektu).

MEMORY.md — co se AI při práci na projektu naučil

Claude si automaticky ukládá poznatky z každé session.
Příště už ví, co fungovalo a co ne. Každý řádek zaznamenává nějaký vyřešený problém.

```
## Parser (parse_pdf_v2.py)
- 14 per-kraj parserů, handles all variants
- Key gotchas:
  EN DASH (U+2013) v NACE regex
  MSK end marker: "Realizace krajské"
  ZLK horizontal parser stop markers
  Windows cp1250 → UTF-8

## NACE pokrytí
Full NACE: JHM, JHČ, KVK, HKK, PAK, PLK
Partial: LBK (4/5), MSK (8/10)
No NACE: PHA, OLK, ZLK, ULK

## Středočeský centroid offset
Slide 2: +30,+35  |  Slide 3: +25,+30
Slide 4: +20,+25  (aby se nepřekrýval
                   s Prahou)
```

/memory — zobrazí a upraví paměť · /init — vygeneruje CLAUDE.md z projektu
A mnohé další podobné příkazy. Sledujte dokumentaci kódovacího asistenta nebo zavolejte /help.

CLAUDE.md — vaše pravidla pro AI

Soubor ve složce projektu. AI ho čte při každém spuštění.
Vaše instrukce, které platí vždy a nemusíte je opakovat.

```
# Naš CLAUDE.md
Žádné soubory nesmí být smazány.
Změny pouze na vyzvání.
# Postupně doplněno:
Brand barvy: #0A416E, #0087CD, ...
"domény specializace" NE "strategie"
Žádné textové popisky na mapách.
```

Plánovací dokumenty — před implementací

V Plan Mode (Shift+Tab) nejdřív navrhneme schéma:
JSON výstup, thresholdy, barevnou paletu.
Teprve pak AI píše kód podle odsouhlaseného plánu.

```
# PLAN_SLIDE5_SEMANTIC.md
## Výstupní JSON:
{ "metadata": {
    "semantic_threshold": 0.35,
    "ford_threshold": 0.38 },
  "kraje": { "PHA": {
    "v_obou": 456,
    "jen_semantic": 123 }}}
```

Děkujeme za pozornost.

A stále se můžete se ptát a sdílet online...

slido.com
kód # 1820 4031

Děkujeme za pozornost

Kristýna Meislová meislova@tc.cz
Adéla Kučerová kucerova@tc.cz

Technologické centrum Praha
www.tc.cz | www.horizonary.cz