

Role AI ve správě dat

Martin Nečaský

martin.necasky@matfyz.cuni.cz

Katedra softwarového inženýrství
Matematicko-fyzikální fakulta Univerzity Karlovy



Co je AI?

Simulace lidské inteligence a chování strojem.

Klasifikace Rozhodování
Odvozování
Tvoření Dokazování

Co je AI?

Symbolická

Diagnostické
systémy v medicíně

Virtuální asistenti
(Siri)

Systémy pro detekci
podvodů

Znalostní grafy

Sub-symbolická

DeepL ChatGPT

Gemini

Midjourney

Copilot

Sora

Lovable

Synthesia

Symbolická AI

Založena na reprezentaci světa (entit, vztahů, pravidel) v exaktních logických symbolech.

Pizza:

- pokrm
- placka z **pšeničného** těsta
- rajčatová omáčka
- alespoň 1 další ingredience

Margherita

Pizza:

- druh **pizzy**
- šunka
- mozzarella
- bazalka

Pizza Picante:

- druh **pizzy**
- plátky salámu
- **chilli papričky**

Pšenice:

- druh **obiloviny**

Obilovina:

- rostlina
- obsahuje **lepek**

Chilli paprička:

- druh papriky
- je **pálivá**

Pokrm vhodný pro děti:

- pokrm
- neobsahuje **pálivou** ingredienci

Pokrm vhodný pro bezlepkovou dietu:

- pokrm
- neobsahuje ingredienci s **lepkem**

Celiakie:

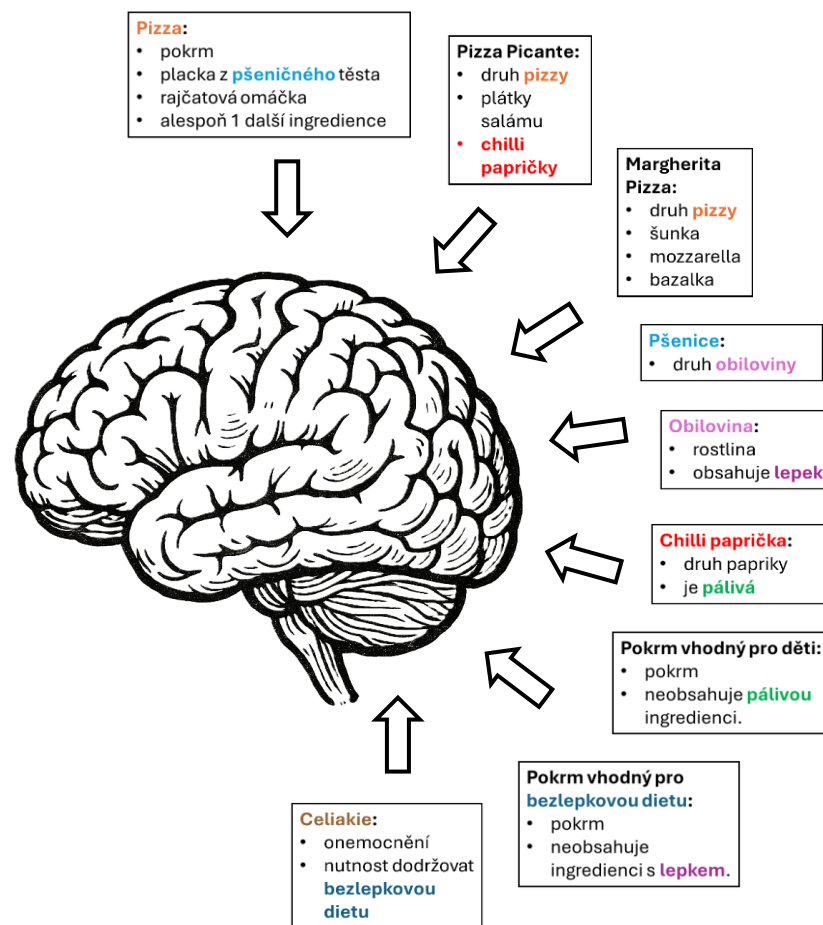
- onemocnění
- nutnost dodržovat **bezlepkovou dietu**

Symbolická AI

Je Pizza Picante vhodná pro děti s celiakií?

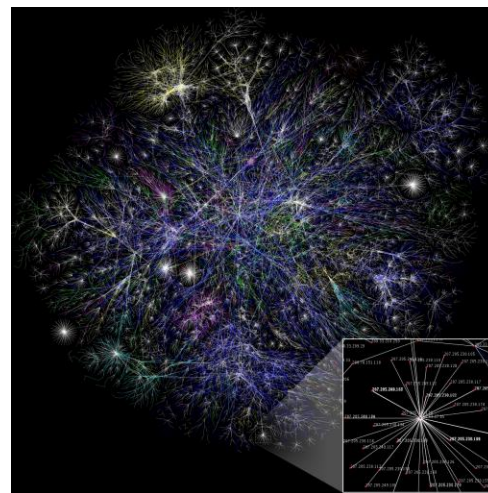
Ne, protože:

- **Pizza** není pokrm vhodný pro osoby s **celiakií**, protože:
 - **Celiakie** vyžaduje dodržování **bezlepkové diety**.
 - Pokrm vhodný pro **bezlepkovou dietu** nesmí obsahovat **lepek**.
 - **Pizza** obsahuje **lepek**, protože:
 - **Pizza** je pokrm, tvořený plackou z **pšeničného těsta**.
 - **Pšenice** je **obilovina**.
 - **Obiloviny** obsahují **lepek**.
- **Pizza Picante** není pokrm vhodný pro děti, protože:
 - **Pizza Picante** obsahuje pálivou ingredienci, protože:
 - **Pizza Picante** je pokrm, který obsahuje **chilly papričky**.
 - **Chilly papričky** jsou **pálivé**.



Sub-symbolická AI

Založena na vzorech opakujících se s určitou pravděpodobností v trénovacích datech.



Sub-symbolická AI

Je Pizza Picante vhodná pro děti s celiakií?

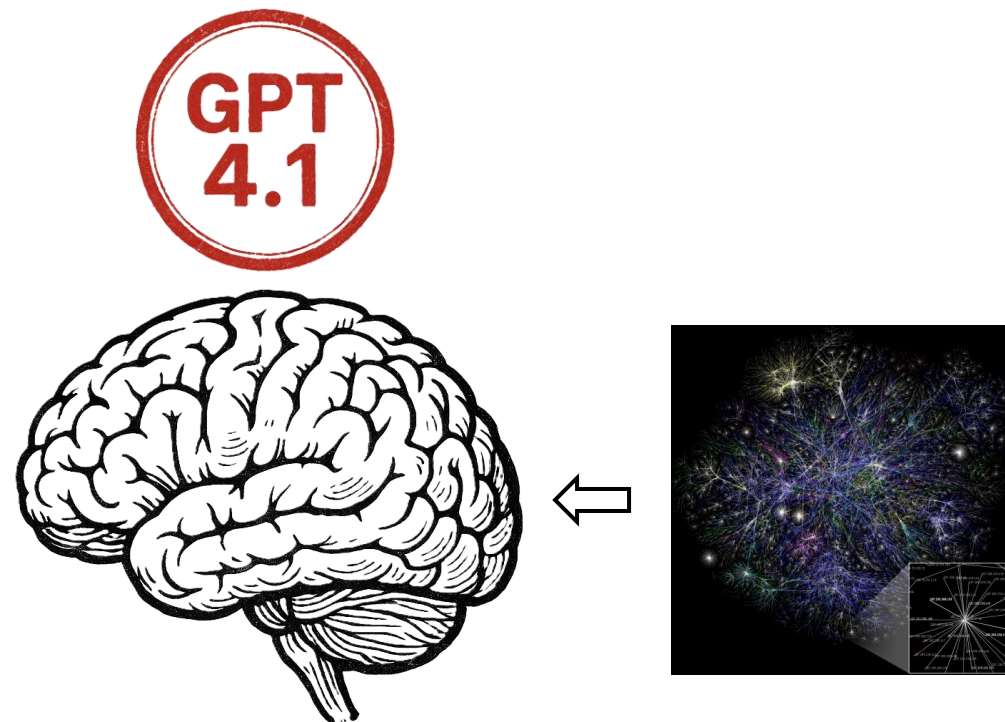
Pizza Picante může být pro děti s celiakií nevhodná, pokud není připravena z bezlepkových surovin a s důrazem na prevenci křížové kontaminace.

Obsah lepku

Tradiční Pizza Picante se obvykle připravuje z pšeničného těsta, které obsahuje lepek. Pro osoby s celiakií je nezbytné zvolit bezlepkovou variantu těsta. Některé restaurace a výrobci nabízejí bezlepkové pizzy, ale je důležité ověřit, zda jsou skutečně bezpečné.

Pálivost

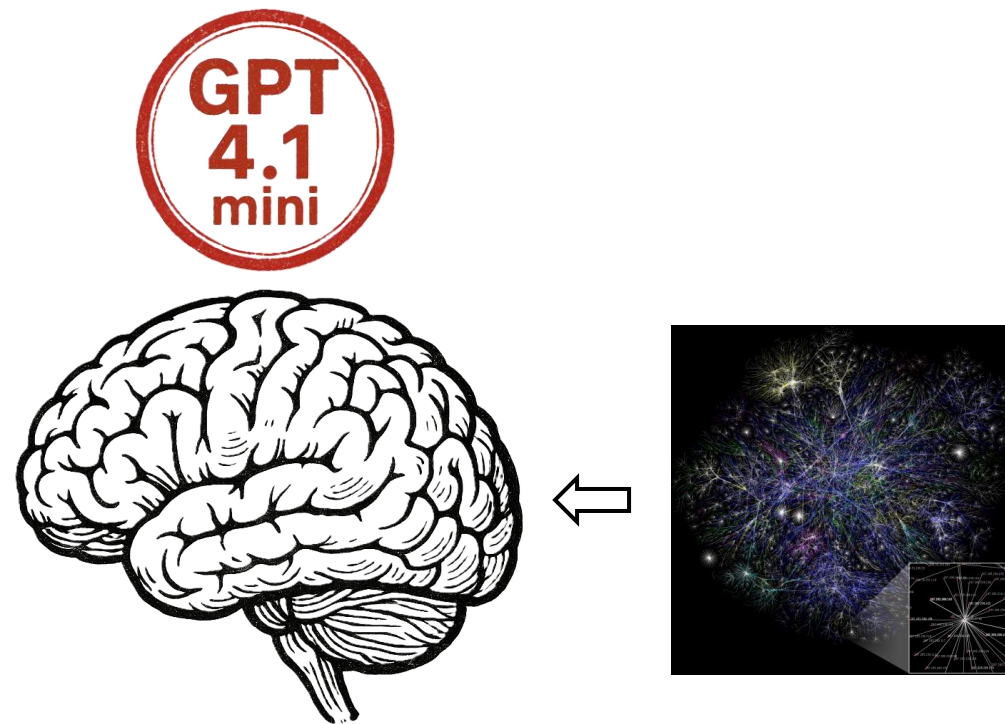
Pizza Picante je charakteristická svou pikantní chutí, často obsahuje ostré ingredience jako chilli papričky nebo pikantní salámy. Pro děti může být taková úroveň pálivosti nevhodná, zejména pokud nejsou zvyklé na ostrá jídla.



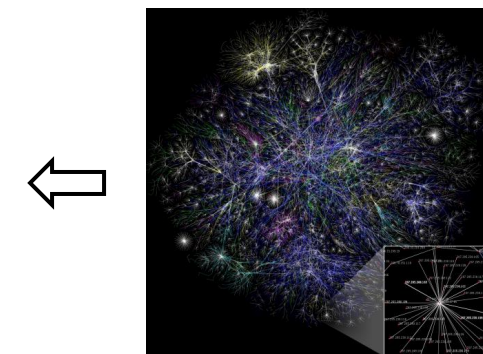
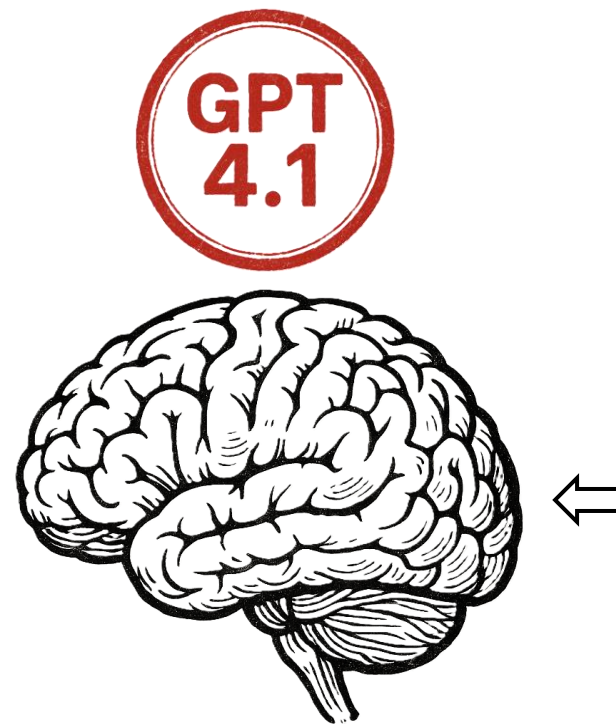
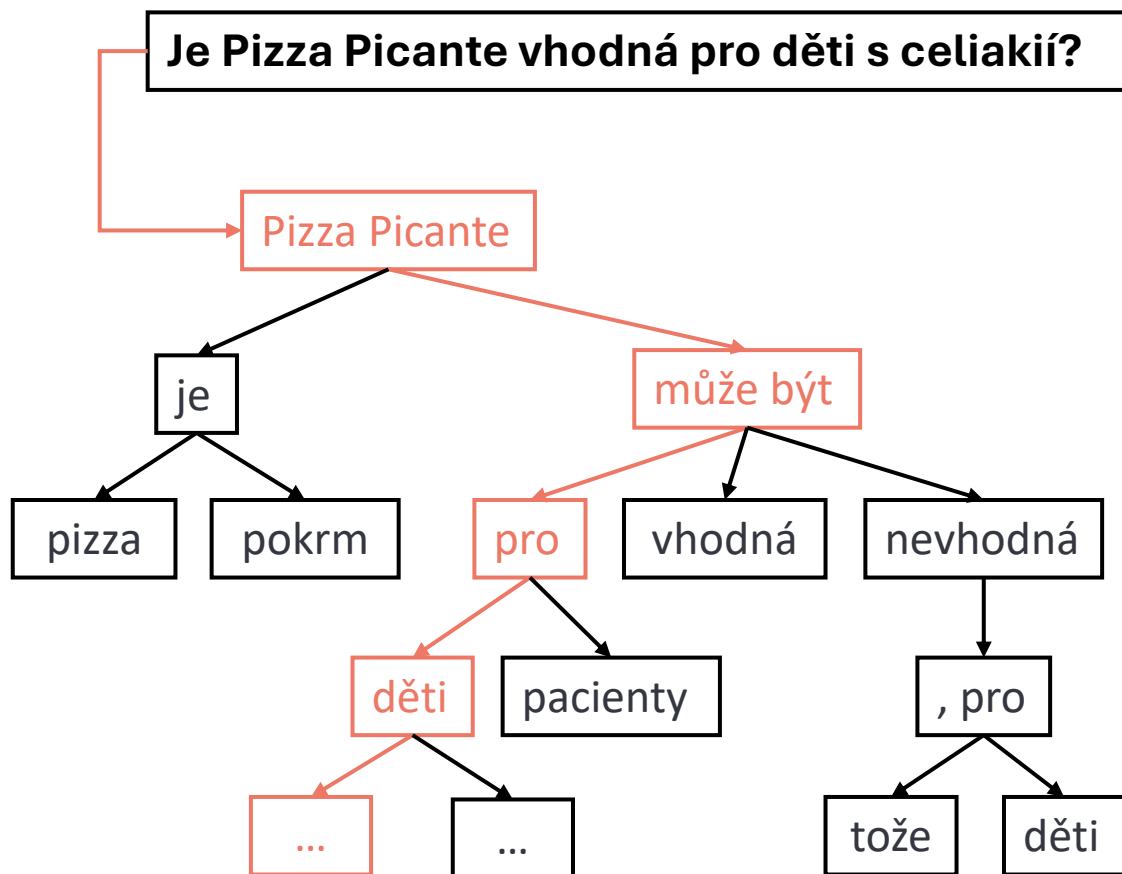
Sub-symbolická AI

Je Pizza Picante vhodná pro děti s celiakií?

Pizza Picante je bezlepková pizza, která je vhodná pro děti s celiakií, pokud je připravena a podávána za přísných hygienických podmínek. Je však důležité mít na paměti, že i malé množství lepku může být pro osoby s celiakií škodlivé. Tento typ pizzy je běžně dostupný v některých pizzeriích a restauracích, například na Slovensku .



Sub-symbolická AI



Využití sub-symbolické AI v analýze dat

	A	B	C	D	E	F	G	H	I	J	K	L
1	DRUH	DRUH_T	FORMA	FORMA_T	OBLAST	OBLAST_T	OBOR	OBOR_T	ROCNIK	STAV	STAV_T	POCET
47	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	1	P	přerušeno	1
48	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	1	R	rozlož.ročník -	4
49	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	1	S	studuje	195
50	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	1	Z	zanechal	64
51	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	2	R	rozlož.ročník -	8
52	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	2	S	studuje	146
53	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	2	V	výjezd na stáž	1
54	B	bakalářské	D	prezenční	I	Informatika	IPP	Informatika	2	Z	zanechal	12

What are you working on?



studium-25-05-29.csv
Spreadsheet



Na základě dat v přiloženém souboru urči, jaký studijní obor má nejvyšší počet studentů.



Tools



What are you working on?



studium-25-05-29.csv
Spreadsheet



Na základě dat v přiloženém souboru urči, jaký studijní obor má nejvyšší počet studentů.



Tools



Obor s nejvyšším počtem studentů

Pořadí	Studijní obor	Počet studentů (aktuálně studuje)
1.	Informatika	341
2.	Fyzika	307
3.	Obecná matematika	267
4.	Finanční matematika	175
5.	Computer Science (angl.)	113

Pokud započítáme všechny evidované stavy (včetně přerušených, zanechaných apod.), zůstává pořadí stejné a Informatika vede s celkovým součtem **431** záznamů.

Takže **Informatika** je studijní obor s nejvyšším počtem studentů na základě údajů v přiloženém souboru.

What are you working on?



studium-25-05-29.csv
Spreadsheet



Na základě dat v přiloženém souboru urči, jaký studijní obor má nejvyšší počet studentů.



Tools



```
1 import sys
2 import pandas as pd
3 from pathlib import Path
4
5 CSV_FILE = "studium-25-05-29.csv"
6 OBOR_COL = "Obor"
7 STATUS_COL = "Status"
8 ACTIVE_STATUSES = {"studuje", "aktuálně studuje", "active", "enrolled"}
9
10 def main(csv_path: str | Path = CSV_FILE) -> None:
11     df = pd.read_csv(csv_path)
12
13     df[OBOR_COL] = df[OBOR_COL].str.strip()
14     df[STATUS_COL] = df[STATUS_COL].str.strip().str.lower()
15
16     current = df[df[STATUS_COL].isin(ACTIVE_STATUSES)]
17     counts_current = current[OBOR_COL].value_counts()
18     top_obor_cur = counts_current.idxmax()
19     top_count_cur = counts_current.iloc[0]
20
21     counts_all = df[OBOR_COL].value_counts()
22     top_obor_all = counts_all.idxmax()
23     top_count_all = counts_all.iloc[0]
24
25     print("=== Aktuálně studuje ===")
26     print(counts_current.to_string())
27     print(f"\nNejvíce studentů má: {top_obor_cur} ({top_count_cur})")
28
29     print("\n=== Všechny záznamy ===")
30     print(counts_all.to_string())
31     print(f"\nNejvíce záznamů celkem má: {top_obor_all} ({top_count_all})")
32
33 if __name__ == "__main__":
34     csv_arg = Path(sys.argv[1]) if len(sys.argv) > 1 else CSV_FILE
35     main(csv_arg)
```

matfyz


Create an app that allows to explore statistical data in the attached file. I need these features: The user can search in "OBLAST" and "OBOR" and filter the records...

★ Show thinking

Okay, I can create a self-contained HTML application for you that will allow you to explore the statistical data from your CSV file. This app will include search functionality for "OBLAST_T" (Oblast description) and "OBOR_T" (Obor description), and filters for "DRUH", "OBLAST" (code), "OBOR" (code), "ROCNIK", and "STAV". After applying these, it will display the total, average, minimum, and maximum number of students ("POCET") from the selected records.

You'll be able to save the code below as an `.html` file and open it in any modern web browser. For the app to work, the CSV file `studium-25-05-29.csv` must be in the same folder as the HTML file when you open it.

Here's the code for your data exploration app:

 **Student Data Explorer App**
1 Jun, 16:01

[Try again without Canvas](#)

How to Use:

1. **Save the Code:** Copy the entire code block above and paste it into a plain text file. Save this file with an `.html` extension (e.g., `student_explorer.html`).
2. **Place CSV File:** Download your `studium-25-05-29.csv` file and place it in the exact same folder where you saved the `student_explorer.html` file.

Ask Gemini

+  Video  Deep Research  Canvas

```
Student Data Explorer App  Code Preview Share X
1 <!DOCTYPE html>
2 <html lang="en">
3 <head>
4   <meta charset="UTF-8">
5   <meta name="viewport" content="width=device-width, initial-scale=1.0">
6   <title>Student Data Explorer</title>
7   <script src="https://cdn.tailwindcss.com"></script>
8   <style>
9     body { font-family: 'Inter', sans-serif; }
10    .filter-select {
11      min-width: 150px;
12    }
13    #dataTable th, #dataTable td {
14      border: 1px solid #e2e8f0; /* Tailwind gray-300 */
15      padding: 8px;
16      text-align: left;
17    }
18    #dataTable th {
19      background-color: #f1f5f9; /* Tailwind slate-100 */
20      position: sticky;
21      top: 0;
22      z-index: 10;
23    }
24    .stats-card {
25      background-color: #f8f9fa; /* Tailwind slate-50 */
26      border-radius: 0.5rem; /* rounded-lg */
27      padding: 1.5rem; /* p-6 */
28      box-shadow: 0 1px 3px 0 rgba(0, 0, 0, 0.1), 0 1px 2px -1px rgba(0, 0, 0, 0.1); /* shadow-md */
29    }
30    .loader {
31      border: 5px solid #f3f3f3; /* Light grey */
32      border-top: 5px solid #3498db; /* Blue */
33      border-radius: 50%;
34      width: 50px;
35      height: 50px;
36      animation: spin 1s linear infinite;
37      margin: 20px auto;
38    }
39    @keyframes spin {
40      0% { transform: rotate(0deg); }
41      100% { transform: rotate(360deg); }
42    }
43    /* Custom scrollbar for table container */
44    .table-container::-webkit-scrollbar {
45      width: 8px;
46      height: 8px;
47    }
48    .table-container::-webkit-scrollbar-track {
49      background: #f1f1f1;
50      border-radius: 10px;
51    }
52    .table-container::-webkit-scrollbar-thumb {
53      background: #888;
54      border-radius: 10px;
55    }
56    .table-container::-webkit-scrollbar-thumb:hover {
57      background: #555;
```

Student Data Explorer

Analyze student records with interactive search and filters.

Search Oblast Description (OBLAST_T):

Search Obor Description (OBOR_T):

DRUH (Type of Study):

OBLAST (Code):

OBOR (Code):

ROCNIK (Year):

STAV (Status):

Reset Filters

Total Students

3,267

Average Students

5.90

Min Students

1

Max Students

195

Dává nám AI správné výsledky?

Obor s nejvyšším počtem studentů

Pořadí	Studijní obor	Počet studentů (aktuálně studuje)
1.	Informatika	341 ✖
2.	Fyzika	307 ✖
3.	Obecná matematika	267 ✖
4.	Finanční matematika	175 ✖
5.	Computer Science (angl.)	113 ✖

Pokud započítáme všechny evidované stavy (včetně přerušených, zanechaných apod.), zůstává pořadí stejné a Informatika vede s celkovým součtem **431** záznamů.

Takže **Informatika** je studijní obor s nejvyšším počtem studentů na základě údajů v přiloženém souboru.

Studijní obory mají své jazykové varianty.

Studijní obory mají své specializace.

Student je fyzická osoba zapsaná ve studijním oboru.

Rozložení ročníku studentem znamená, že daný ročník studia absolvuje během dvou akademických roků místo jednoho akademického roku.

Zanecháním studia student přestává být zapsán ve studijním oboru.

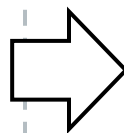
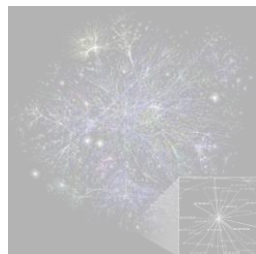
...

Naivní vstup

studium.csv

Jaký studijní obor
má nejvyšší počet
studentů?

Sub-symbolická AI



Studijní obor

Počet studentů (aktuálně studuje)

Informatika

341

Vstup obohacený o **kontext**

studium.csv

obory.csv

specializace.csv

Jaký studijní obor
má nejvyšší počet
studentů?

Symbolická reprezentace světa

studium_schema.json

obory_schema.json

specializace_schema.json

Studijní obory mají své jazykové varianty.

Studijní obory mají své specializace.

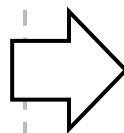
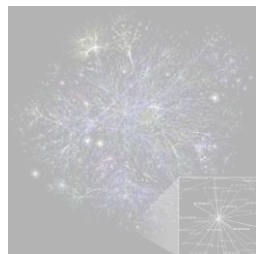
Student je fyzická osoba zapsaná ve studijním oboru.

Rozložení ročníku studentem znamená, že daný ročník studia absolvuje během dvou akademických roků místo jednoho akademického roku.

Zanecháním studia student přestává být zapsán ve studijním oboru.

...

Sub-symbolická AI



Studijní obor s nejvyšším počtem studentů je:

Informatika (kód: IPP) s celkovým počtem **796** studentů.

Počet zahrnuje všechny jeho jazykové varianty a všechny jeho specializace podle pravidel.

Postup výpočtu

1. Za studenta považují pouze osoby ve stavech:

- R (rozložený ročník)
- S (studuje)
- V (výjezd na stáž)

2. Studijní obor zahrnuje:

- všechny jeho jazykové varianty (české i anglické)
- všechny jeho specializace (a jejich jazykové varianty)

3. Součet za obor =

součet počtu studentů pro všechny záznamy v datech, které odpovídají jakékoli variantě daného oboru nebo jeho specializace (bez ohledu na jazykovou variantu), ve stavech R, S, V.



Kde vzít kontext?

• Datové slovníky

- Důležité entity a vztahy mezi nimi jako pojmy
- Definice pojmů, vazba do legislativy

• Datové katalogy

- Metadata, popisky, schémata
- Číselníky
- NKOD
- NKD
- LKOD
- LKD

<https://data.gov.cz>

Portál o datech

SPRÁVA DAT VZDĚLÁVÁNÍ ČLÁNKY DATOVÁ KVALITA NÁRODNÍ KATALOG OTEVŘENÝCH DAT DALŠÍ

Poskytovatelé (1) Co hledáte Hledat

Digitální a informační agentura (358) Rozšířené vyhledávání

Témata (34) Nalezeno 358 datových sad. Název vzestupně

Vláda a veřejný sektor (347)

Populace a společnost (176)

Hospodářství a finance (171)

výkonná moc (45)

veřejná správa (30)

výkonný orgán (30)

místní orgány státní správy (29)

orgány veřejné správy (29)

pravomoc výkonné moci (29)

samospráva (29)

správa věcí veřejných (29)

veřejná služba (29)

Zobrazit další

Agendy

Agendy evidované v Registru práv a povinností ve smyslu § 51 zákona č. 111/2009 Sb. o základních registrech.

JSON JSON-LD

Aplikace využívající otevřená data

Aplikace využívající otevřená data registrovaná na portálu otevřených dat.

RDF Turtle

Asistenční služby dle nařízení EU o Jednotné digitální bráně

Popis asistenčních služeb, o kterých jsou členské státy EU povinny informovat na základě nařízení EU o Jednotné digitální bráně.

JSON

Číselník fází

<https://slovník.gov.cz>

ShowIt 🔍

Silniční vozidlo

Silniční vozidlo je motorové nebo nemotorové vozidlo, které je vyrobené za účelem provozu na pozemních komunikacích pro přepravu osob, zvířat nebo věcí.

<https://esbirtka.opendata.cz/zdroj/předpis/56/2001/sekce/2/1>

Související pojmy

Slovník zákona č. 361/2000 Sb. o provozu na pozemních komunikacích a o změnách některých zákonů – slovník Vozidlo

Silniční vozidlo

Motorové vozidlo

Nemotorové vozidlo

Technicky nezpůsobilé silniční vozidlo

Technicky způsobilé silniční vozidlo

Vztahy

Silniční vozidlo

Datový slovník registru silničních vozidel datum první registrace silničního vozidla v ČR

Datový slovník registru silničních vozidel

Jak kontext pro AI vytvořit a udržovat?

Kvalitní správa dat!!!

<https://data.gov.cz>

Portál o datech

SPRÁVA DAT VZDĚLÁVÁNÍ ČLÁNKY DATOVÁ KVALITA NÁRODNÍ KATALOG OTEVŘENÝCH DAT DALŠÍ

Správa dat

Kvalitní správa dat je nezbytným základem pro to, aby si organizace veřejné správy mohly vyměňovat potřebná data a poskytovaly občanům kvalitní služby. Sdílení dat uvnitř veřejné správy ale nemůže fungovat bez toho, aby úřady svá data dobře znaly a garantovaly jejich využitelnost. Data hrají také zcela zásadní roli pro schopnost státu činit efektivní a účinná rozhodnutí. Systematická péče o data a jejich dobrá znalost má navíc ekonomické přínosy i pro samotné úřady. Na této stránce naleznete důležité informace z oblasti správy dat ve veřejné správě, metodické a další podpůrné materiály, které mohou úřady využít k tomu, aby zlepšily správu svých dat.

Správa dat „od A do Z“

[Co je správa dat](#) > [Legislativa související se správou dat](#) > [Strategie správy dat](#) > [Specifika státní správy](#) >

Správu dat chápeme jako soubor principů, které se vztahují k... Předpokladem pro sdílení dat ve veřejné správě musí být správa a... Efektivní a kvalitní správa dat nemá význam jen pro komerční...

- Může AI pomoci?
 - Tvorba metadat pro datové katalogy.
 - Tvorba datových slovníků.

Děkuji za pozornost

Martin Nečaský

martin.necasky@matfyz.cuni.cz

Katedra softwarového inženýrství
Matematicko-fyzikální fakulta Univerzity Karlovy

