

# NAČO SÚ NÁM DÁTA

## Úvod

Toto je tretia metodika zo série 3 metodík, ktoré sú venované problematike umelej inteligencie. Tematicky sa venuje využitiu dát pre strojové učenie a dolovaniu v dátach, čím môže poslúžiť aj na motiváciu žiakov pre ich ďalšie štúdium v modernom a perspektívnom odbore dátovej vedy (Data Science) alebo dátovej analýzy<sup>1</sup>. Žiaci prakticky preskúmajú rôzne dáta vytvorené pri práci na internete a vyskúšajú si jednoduché úlohy na prácu s dátami cez stránku **Databasic.io** ([databasic.io](https://databasic.io)). Žiaci majú k dispozícii pracovný list, ktorý obsahuje zadania úloh, miesto na žiacke riešenie a miesto pre poznámky.

### Poznámka:

Pracovné listy obsahujú aj odkazy na online zdroje, preto je pri tejto metodike výhodné, pokiaľ učiteľ poskytne žiakom pracovné listy v elektronickej podobe.

## PRIEBEH VÝUČBY

Osnova vyučovacej hodiny (podľa modelu 5E):

- **Zapojenie (5 minút)** – ukážka nástrojov na internete a rozhovor so žiakmi
- **Skúmanie (10 minút)** – praktické úlohy na internete s použitím pracovného listu (úlohy 1 a 2)
- **Vysvetlenie (10 minút)** – zhrnutie výsledkov úloh z pracovného listu, ukážka ďalších nástrojov
- **Rozpracovanie (10 minút)** – praktická úloha na internete s použitím pracovného listu (úloha 3)
- **Hodnotenie (5 minút)** – reflexia a spoločná diskusia so žiakmi

## ZAPOJENIE (CCA 5 MIN.):

Hodinu začneme krátkym motivačným rozhovorom so žiakmi. V súčasnosti každodenne produkujeme veľké množstvo dát z rôznych zdrojov (napr. sociálnych sietí, online nakupovania, verejnej dopravy alebo GPS senzorov), ktoré nám môžu pomôcť pri rýchlejšom a efektívnom riešení rôznych problémov a úloh. S výsledkami dátovej analýzy sa môžeme stretnúť v rôznych podobách – napríklad na základe správania sa zákazníkov v obchode sú nám odporúčané ďalšie produkty. Netflix používa analýzu dát svojich divákov pri výbere a vytváraní nových seriálov. Dopravné dáta sa používajú pri

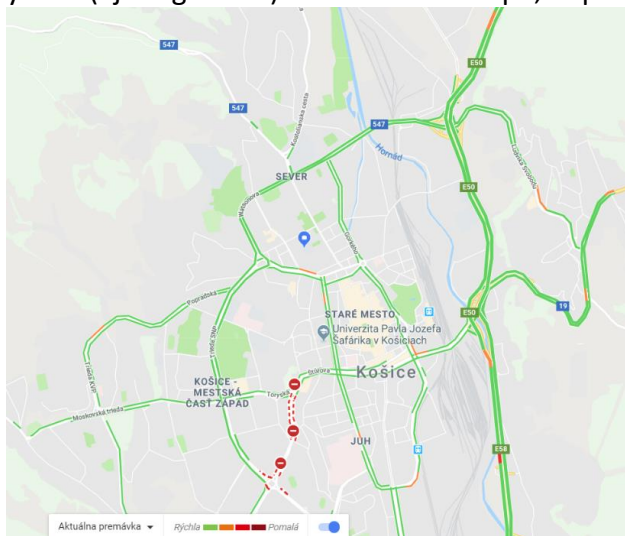
---

<sup>1</sup> V tejto súvislosti je dobré spomenúť aj v súčasnosti mimoriadne aktuálny pojem **Big Data**, resp. veľké dáta, ktorým sa však v tejto metodike nebudeme explicitne venovať, ale niektoré zaujímavé informácie pre učiteľa je možné získať napr. z článku <https://touchit.sk/big-data-ked-je-velky-naozaj-velky/61952>

vývoji autonómneho vozidla a analýzou dát z DNA vzoriek budeme schopní vytvárať efektívnejšie a kvalitnejšie lieky.

Následne žiakom ukážeme využitie dát pre interaktívne zobrazenie plynulosti premávky na cestách pomocou služby **Mapy Google**:

- Otvoríme v okne webového prehliadača službu <https://maps.google.sk>
- V ľavom hornom rohu klikneme na ikonu ponuky ☰.
- Vyberieme položku **Premávka**
- Aktuálny stav (aj s legendou) sa zobrazí na mape, napr. :



Žiakom položíme otázku, či by vedeli vysvetliť, ako sa tieto údaje môžu dostať do mapy. V skutočnosti na to Google využíva dáta zo senzorov z mobilných telefónov s OS Android<sup>2</sup>.

Taktiež môžeme ukázať online mapu verejnej dopravy v Berlíne v reálnom čase (<http://bit.do/fahrinfo>).

Na základe týchto úvodných aktivít predstavíme žiakom špecifický cieľ hodiny:

- uviesť niekoľko typických príkladov úloh (problémov) riešených pomocou umelej inteligencie.

**Poznámka:**

Špecifický cieľ nadväzuje na špecifický cieľ z predošlej hodiny – v rámci tejto hodiny sa zameriame na jednu špecifickú podoblasť úloh, pri ktorých si ukážeme niektoré príklady využitia dát a jednoduchej dátovej analýzy – jej pokročilé nástroje využívajú postupy strojového učenia (tzv. **machine learning**), ktoré zaraďujeme k technikám umelej inteligencie.

---

<sup>2</sup> Bližšie informácie napr. <http://www.businessinsider.com/how-google-maps-knows-about-traffic-2015-11>

## SKÚMANIE (CCA 10 MIN.):

V tejto časti hodiny žiaci pracujú vo dvojiciach s pracovnými listami (úlohu 1 a 2). Učiteľ len monitoruje prácu žiakov, v prípade potreby im poskytne konzultáciu alebo vysvetlí časť, v ktorej im niečo nie je jasné.

**Úloha 1** Pomocou nástroja **Google Trends** (<https://trends.google.com/trends/>) preskúmajte tému pizza od roku 2004 až doteraz na celom svete.

- a) Aký je celkový trend? rastúci – klesajúci – viac-menej stabilný  
b) V ktorých troch krajinách bol najväčší záujem o túto tému?

c) Aký bol najčastejší používaný dopyt? \_\_\_\_\_

Porovnajte výsledky s témou hamburger.

- d) Aký je celkový trend? rastúci – klesajúci – viac-menej stabilný  
e) V ktorých troch krajinách bol najväčší záujem o túto tému?

f) Aký bol najčastejší používaný dopyt? \_\_\_\_\_

Porovnajte výsledky pre obe témy, avšak len na Slovensku.

- g) Aký je celkový trend v porovnaní s celým svetom?  
h) V ktorej časti Slovenska je najvyhľadávanejšia téma pizza? \_\_\_\_\_  
i) V akej súvislosti najčastejšie vyhľadávajú Slováci tému pizza?

j) V ktorej časti Slovenska je najvyhľadávanejšia téma hamburger? \_\_\_\_\_

k) V akej súvislosti najčastejšie vyhľadávajú Slováci tému hamburger?

\_\_\_\_\_

### Úloha 2

- a) Pomocou nástroja **SocialMention** (<http://www.socialmention.com/>) preskúmajte nasledovné obchodné značky vo videách a doplňte tabuľku:

Značka	Počet spomenutí (mentions)	Sila (strength)	Sentiment
Pepsi			
McDonalds			
Milka			

- b) Zvoľte si jednu z týchto značiek a preskúmajte ju aj pomocou nástroja **Social Searcher** (<https://www.social-searcher.com>) – po ukončení vyhľadávania kliknite na tlačidlo



(nájdete ho vpravo na úvodnom banneri so základným prehľadom výsledkov) a pomocou záložky **SENTIMENT** porovnajte, aké sú reakcie na rôznych sociálnych sieťach – sú rovnaké alebo niekde sa výsledky výraznejšie líšia?

## VYSVETLENIE (CCA 10 MIN.):

Na základe riešení úloh z pracovných listov vyzveme žiakov, aby vyskúšali porozmýšľať, kde by bolo možné využiť dáta, ktoré Google získava od užívateľov pri využívaní jeho služieb (reklama, odporúčacie systémy, vylepšenie prekladu/strojové učenie, reCaptcha – rozpoznávanie obrázkov).

Taktiež sa ich spýtame, kde by vedeli využiť dáta získané napr. službou **SocialMention** alebo **Social Searcher** (alebo obdobnou, ktorá využíva napr. dáta zo sociálnych sietí).

Súčasťou práce s dátami je aj ich vizualizácia, ktorá nám ponúkne možnosť ďalšej analýzy a napovie viac o súvislostiach, ktoré možno zo samotných dát (napr. v tabuľkovej podobe) nemusia byť zrejmé. Výsledkom analýzy big data bola napríklad mapa internetu **The Internet Map** (<https://internet-map.net>), ktorú môžeme spolu so žiakmi v krátkosti preskúmať – do vyhľadávacieho políčka zadáme názov krajiny (po anglicky) alebo URL adresu stránky. Výsledná mapa nám ukazuje vzájomné prepojenie webstránok v danej krajine – čím väčší kruh, tých navštevovanejšia je stránka. Keďže návštevníci prechádzajú zo stránky na stránku, čím vznikajú prepojenia medzi stránkami a čím silnejšie prepojenie medzi stránkami je, tým sú stránky bližšie k sebe. Ako je možné z mapy zistiť, stránky sú farebne zoskupené podľa príslušnej krajiny, kam patria.

### **Poznámka:**

Treba však podotknúť, že táto mapa internetu pochádza z roku 2011, v tejto ukážke je však použitá na ilustráciu ako zaujímavý príklad na interpretáciu dát, ktoré vznikli (a každodenne ďalej vznikajú) používaním internetu.

Iný príklad vizualizácie dát – tentokrát z úplne inej oblasti – ponúka video od NASA **Perpetual Ocean** (<https://www.youtube.com/watch?v=vQk5N0I9Crs>), z ktorého ukážku žiakom premietneme pomocou dataprojektora. Zobrazuje 3D model zemského oceána a neustále pohyby jeho podmorských prúdov na základe dát získaných monitorovaním od júla 2005 do októbra 2007.

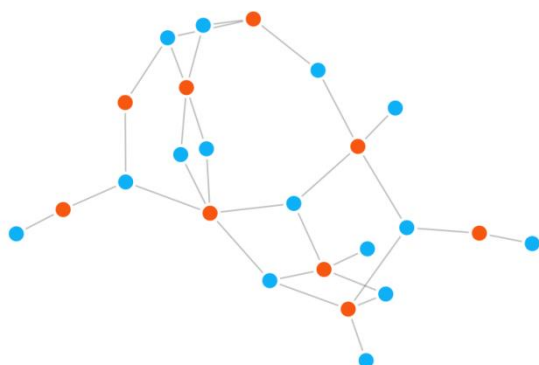
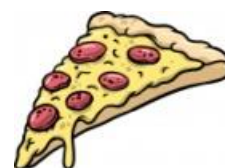
Na záver ukážeme ešte jednu interaktívnu vizualizáciu – **Health Infoscapes** (<http://senseable.mit.edu/healthinfoscapes/>) – unikátny projekt, ktorý zachytáva vzájomné prepojenie chorôb (môžeme vyskúšať sledovať rozdiely u mužov a žien aj pre rôzne kategórie/skupiny chorôb, pravdepodobnosť výskytu podľa veľkosti farebného kruhu, ako aj choroby, ktoré sa často vyskytujú spolu, resp. sa môžu pridružiť, čo je zaujímavé zobrazené aj pri zmene zobrazenia v časti LAYOUT). Táto „mapa“ vznikla na základe analýzy dát zo 7,2 milióna medicínskych záznamov.

## ROZPRACOVANIE (CCA 10 MIN.):

V tejto časti hodiny žiaci opäť pracujú vo dvojiciach s pracovnými listami (úlohu 3), pričom budú využívať analytický nástroj **Connect the Dots** na stránke **Databasic.io** ([databasic.io](http://databasic.io)). Najprv žiakov vyzveme, aby si otvorili súbor **pizza.xlsx** v tabuľkovom kalkulátore z pracovného priečinka (**I\_SS\_33\_Naco\_su\_nam\_data\_pracovne.zip**). Obsahuje výsledky simulovanej ankety medzi žiakmi

jednej triedy, keď mali odpovedať na otázku, ktorú pizzu majú najradšej (každý žiak mohol navrhnúť 1-3 druhy). Položíme žiakom otázku, aké informácie vedia vyčítať pri pohľade na tieto dáta (napr. koľkokrát zahlasoval, ktorá pizza je najobľúbenejšia, koľko žiakov má rado ktorý z uvedených druhov píz...). Upozorníme žiakov, že v tomto prípade je veľmi malá vzorka respondentov v ankete a taktiež malý počet odpovedí. Pokiaľ by sme robili podobný prieskum medzi všetkými návštevníkmi konkrétnej pizzérie (resp. všetkých pizzérií), rozsah dát by bol podstatne väčší. Cieľom úlohy 3 bude vizualizovať tieto dáta pomocou siete a pokúsiť sa zodpovedať pomocou výsledného grafu niekoľko otázok. Učiteľ len monitoruje prácu žiakov, v prípade potreby im poskytne konzultáciu alebo vysvetlí časť, v ktorej im niečo nie je jasné.

**Úloha 3** Otvorte súbor **pizza.xlsx** v tabuľkovom kalkulatore. Na stránke **Databasic.io** ([databasic.io](https://databasic.io)) zvolte mástroj **Connect the Dots** a kliknite na voľbu **Upload a file**, pomocou ktorej nahrajte na analýzu súbor **pizza.xlsx**. Vytvorte sieťový graf kliknutím na tlačidlo **Graph**:



Graf pozostáva z uzlov a hrán, ktorými sú poprepájané.

- Zistite, čo predstavujú uzly vyfarbené:  
červenou farbou ( )  
modrou farbou ( )
- Vedľa grafu máte tabuľku – zistite ktorý uzol je v nej s najvyšším stupňom (angl. **degree**) a čo predstavuje tento uzol (tzv. spojovateľ alebo angl. **connector**): \_\_\_\_\_
- Nájdite a vyznačte v grafe uzol, ktorý predstavuje tohto najlepšieho spojovateľa!
- Skóre v stĺpci **centrality** označuje do akej miery je ten-ktorý uzol spojovateľom, t.j. ako rýchlo je možné cez tento uzol sa spojiť s ostatnými uzlami v sieti. Niektoré uzly majú nulové skóre v stĺpci **centrality** – nájdite a vyznačte ich v grafe! Čo majú spoločné? \_\_\_\_\_
- V hornej časti grafu zvolte možnosť **color by community** – graf sa rozpadne na menšie, farebne odlišené časti, tzv. komunity. Koľko komunit je v grafe? \_\_\_\_\_
- Preskúmajte červenú komunitu – čo ju spája? \_\_\_\_\_
- Preskúmajte zelenú komunitu – čo ju spája? \_\_\_\_\_
- Nájdite v grafe uzol predstavujúci Miša – pravdepodobne si objednáva stále tú istú pizzu. Vedeli by ste mu pomocou grafu navrhnúť/odporučiť, ktorú ďalšiu pizzu by mal vyskúšať? \_\_\_\_\_
- Nájdite v grafe uzol predstavujúci Evku – ktorú ďalšiu pizzu by ste jej odporučili? \_\_\_\_\_
- Vložte do pôvodného súboru **pizza.xlsx** ďalší záznam Evke, ktorá vyskúšala a obľúbila si jednu z navrhovaných píz. Vytvorte v novom okne prehliadača nový graf a porovnajte komunity – došlo k zmene? \_\_\_\_\_
- Vložte do pôvodného súboru nového respondenta Tomáša, ktorý ako jediný obľubuje pizzu Carbonara. Odhadnite, ako sa zmení graf a overte svoj predpoklad vytvorením nového grafu!

## HODNOTENIE (cca 5 min.):

Vyzveme žiakov, aby postupne prezentovali výsledky, ktoré získali. Potom im položíme otázky na zamyslenie:

- pomohol nám takýto spôsob zobrazenia dát a súvislostí medzi nimi?
- boli by sme schopní vyčítať tieto informácie aj z (pôvodnej) tabuľkovej podoby dát?
- vedeli by sme v grafe „vyčítať“ aj iné informácie (okrem tých, ktoré boli v otázkach v úlohe 3)?
- vedeli by sme navrhnúť inú „výskumnú“ oblasť alebo otázku, ktorú by sme chceli/mohli pomocou tohto postupu preskúmať?

Následne žiakov vyzveme, aby vyskúšali zhodnotiť svoje skúsenosti z hodiny:

- ako sa Vám pracovalo? pomohla Vám práca vo dvojici pri riešení úloh?
- objavili ste, niečo, čo Vás zaujalo?
- videli ste niekde už podobné/zaujímavé výsledky analýzy/vizualizácie dát?

### **Poznámka:**

Žiakom môžeme dať formou domácej úlohy navrhnúť a zrealizovať vlastný individuálny alebo skupinový výskumný mikroprojekt, pri ktorom si pripraví vlastné dáta (najlepšie pomocou ankety na nejakú tému) a pokúsia sa ich analyzovať pomocou siete.

Na stránke **Databasic.io** sú aj ďalšie jednoduché nástroje (aj s inštruktážnym videom, ako ich použiť), pomocou ktorých môžu žiaci analyzovať iné dáta. Nástroj **Word Counter** slúži na analýzu textu a vytvorenie mraku slov (žiaci si môžu vytvoriť textový súbor, ktorý bude obsahovať texty piesní ich obľúbeného interpreta – čím viac, tým lepšie – a nechať ho analyzovať, čím získajú predstavu, o čom najčastejšie v týchto piesňach spieval, prípadne môžu porovnať takýmto spôsobom dvoch interpretov a ich repertoár). Nástroj **SameDiff** taktiež analyzuje texty, ich podobnosť (žiaci môžu vo dvojici vytvoriť dva textové dokumenty s ich referátmi alebo slohovými prácami na rovnakú tému a porovnať, do akej miery sa tieto ich texty zhodujú alebo líšia<sup>3</sup>).

---

<sup>3</sup> Tento nástroj môže zaujímavým spôsobom viesť k diskusii ohľadom plagiátorstva a kontroly originality textov (záverečných prác a pod.)