

Balíček  
Dáta

Karta

# Velké dáta

01



Umelá  
inteligencia  
vo vzdelávaní



MINISTERSTVO  
ŠKOLSTVA, VÝSKUMU,  
VÝVOJA A MLÁDEŽE  
SLOVENSKEJ REPUBLIKY

Metodický materiál Kurikula umelej inteligencie pre základné a stredné školy  
Informatika na 2. stupni ZŠ a SŠ – karty

# Veľké dáta

## Základné informácie o veľkých dátach

Ustálená definícia pojmu veľké dáta (Big Data) neexistuje. Medzi spoločné charakteristiky veľkých dát však patrí ich ohromná veľkosť, rýchlosť ich vzniku, ich rôznorodosť (veľkosťou a štruktúrou) a často aj nižšia kvalita či menšia presnosť. Veľké dáta môžu zahŕňať obrázky, videá, audio, texty alebo tzv. digitálne stopy vo forme údajov o správaní používateľov. Sú dôsledkom zrýchlenia a vývoja internetu, pričom rolu tiež hrá výrazné zlacnenie ukladania a spracúvania dát. Typicky je ich toľko, že ich spracovanie si vyžaduje nové prístupy, napríklad metódy strojového učenia.

## Ak vás táto téma zaujíma...

Väčšie objemy dát začali vznikať už zhruba v 80. rokoch minulého storočia [[zdroj](#)], ale veľkými dátami ich nenazývame. Veľkú zmenu okolo roku 2002 priniesli až systémy, ktoré umožňovali veľmi rýchle spracúvanie veľkých objemov dát na bežne dostupných počítačoch, čo výrazne zlacnilo celý proces, a tým podporilo aj nárast veľkých dát.

Veľké dáta pochádzajú z rôznych zdrojov, ako sú IoT zariadenia, záznamy používateľského správania na internete, dopravné systémy, finančné transakcie, zdravotnícke dáta alebo vedecké experimenty. Ich využitie je veľmi široké – od personalizácie služieb a prediktívnej analýzy v zdravotníctve či financiách až po optimalizáciu logistiky, zlepšenie verejných služieb a riadenie inteligentných miest (smart cities). Okrem technologických spoločností využívajú veľké dáta aj štátne organizácie, nemocnice či vedecké inštitúcie.

Strojové učenie zohráva v oblasti veľkých dát zásadnú úlohu. Ľudia môžu vďaka tejto metóde vytvárať modely, ktoré nachádzajú vo veľkých dátach vzory, a na základe toho predpovedať budúci vývoj situácií, personalizovať rôzne produkty či služby, a to často v reálnom čase.

## Balíček Dáta

### Veľké dáta

Interpretácia dát pomocou číbotov



Materiál vytvoril tím neziskovej organizácie AI deťom v rámci projektu AI Kurikulum a podlieha licencií [Creative Commons 4.0 – Medzinárodná](#).

Preklad do slovenčiny zabezpečilo Ministerstvo školstva, výskumu, vývoja a mládeže Slovenskej republiky.



**AI deťom**

[Formulár na pripomienky.](#)



Pozn.: Rodová rovnosť je pre AI deťom kľúčová, ale na zostručenie využívame v našich metodikách formulácie v mužskom rode.

Aktivita na hodinu

# Vzory v ľudskom správaní

30 – 45  
minút

## Opis aktivity

Cieľom úvodnej aktivity balíčka Dáta je, aby si žiaci uvedomili, že veľké dáta obsahujú informácie o správaní ľudí v bežnom živote a je možné z nich na základe analýzy (napr. vizualizácie pomocou grafov) interpretovať naše minulé správanie, prípadne vytvárať predpovede budúceho správania.

Žiaci v aktivite najskôr pomocou aplikácie The New Normal skúmajú vizualizácie veľkých dát z vyhľadávača Google a porovnávajú zmeny nákupného správania používateľov na internete. Potom pomocou stránok What Are We Searching For hľadajú vzory v ľudskom správaní v rôznych regiónoch. Nakoniec hľadajú a interpretujú zmeny v správaní používateľov internetu pred, počas a po období pandémie COVID-19 v nástroji Google Trends.

## Ako aktivita súvisí s veľkými dátami

Podľa článku na stránke [seo.ai](https://seo.ai) používatelia internetu každú sekundu zadajú do vyhľadávača Google cca 99 000 vyhľadávacích dopytov, čo je v prepočte zhruba 8,5 miliárd vyhľadávacích dopytov denne. Tieto veľké dáta sa následne upravujú do štruktúry, na základe ktorej ich potom nástroj Google Trends vizualizuje. Veľké dáta však nevznikajú iba vyhľadávaním na internete, ale napríklad aj pri používaní sociálnych sietí, pri zaznamenávaní transakcií v online obchodoch a bankách či pri zbere dát z inteligentných zariadení, ako sú senzory či nositeľná elektronika (wearables).

# Informácie o lekcii

**Vstupné znalosti/ročníky, dĺžka lekcie**

8. a 9. ročník ZŠ a študenti SŠ, 45 minút.

**Stavebné kamene**

Veľké dáta, Big Data.

**Čo sa žiaci učia?**

Veľké dáta môžu byť nástrojom na porozumenie javom.

**Prečo sa to učia?**

Rozvíjajú si kritické myslenie a schopnosť analýzy a interpretácie komplexných javov.

**Ako spoznáme, že sa to naučili?**

Analyzujú a interpretujú vzory v nákupnom správaní používateľov internetu.

**Pomôcky**

Pedagóg: Projektor, prezentácia na premietnutie.  
Žiaci: Digitálne zariadenia do skupín alebo dvojíc.

**Digitálne kompetencie**

Informácie a komunikácia.

**Bloomova taxonómia**

Porozumenie: Žiaci porovnávajú a interpretujú grafické vyjadrenia na základe veľkých dát.

Analýza: Analyzujú vizualizácie veľkých dát.

## Súlad so štátnym vzdelávacím programom

---

### ŠVP ZV 2023: Informatika

**3. cyklus:**

1. cieľ: Analyzovať problémy a rozvíjať abstrakciu a logiku.

5. cieľ: Zoznámí sa s digitálnou reprezentáciou informácií rôznych typov (výkonový štandard: dekodovať informáciu z jednoduchých reprezentácií, vyhľadať a získať informácie v informačnom systéme a databáze).

6. cieľ: Spoznať princípy a koncepty informatiky, vrátane reprezentácie a spracovania údajov v tabuľkách, stromoch a grafoch (výkonový štandard: interpretovať údaje zo štruktúr).

8. cieľ: Používať sieťové zariadenia a nástroje na prácu v počítačovej sieti (výkonový štandard: diskutovať o správnosti a kvalite vyhľadaných informácií).

9. cieľ: Analyticky posudzovať digitálne technológie (výkonový štandard: diskutovať o digitálnych technológiách a ich vplyve na ostatných).

### ŠVP 2015: Informatika - nižšie stredné vzdelávanie

Reprezentácie a nástroje – informácie

Reprezentácie a nástroje – štruktúry

Komunikácia a spolupráca – práca s webovou stránkou

Komunikácia a spolupráca – vyhľadávanie na webe

Informačná spoločnosť - digitálne technológie v spoločnosti

## ŠVP ZV 2023: AI gramotnosť

Súčasť balíčka	Dáta
Vzdelávací cyklus	3. cyklus
Stručný opis metodiky	Metodika uvádza žiakov do konceptu veľkých dát (Big Data). Prostredníctvom aplikácie The New Normal žiaci skúmajú zmeny nákupného správania používateľov internetu počas pandémie COVID-19. Pomocou nástroja What Are We Searching For hľadajú vzory v ľudskom správaní v rôznych regiónoch sveta a nakoniec využívajú Google Trends na analýzu konkrétnych vyhľadávacích dopytov. Cieľom je pochopiť, že veľké dáta obsahujú informácie o správaní ľudí a možno ich interpretovať pomocou vizualizácií.
Komponenty AI gramotnosti	<b>Komponent 1: Princípy fungovania AI</b> <b>Komponent 3: Kritické myslenie</b>
Ciele AI gramotnosti	<b>3. cyklus</b> <ul style="list-style-type: none"> <li>Porozumieť princípom fungovania AI (LLM, asistenti, agenti) a vplyvu kvality dát na zaujatosť výstupov.</li> <li>Kriticky overovať výstupy AI, rozpoznať manipuláciu a dezinformácie.</li> <li>Analyzovať vplyv AI na jednotlivca, spoločnosť a životné prostredie.</li> </ul>
Kľúčové obsahy a očakávané spôsobilosti	<b>3. cyklus</b> <b>Rozpoznávanie AI systémov</b> <ul style="list-style-type: none"> <li>Mapovať prítomnosť AI v rôznych oblastiach spoločnosti vrátane systémov, ktoré fungujú v pozadí a nie sú priamo interaktívne.</li> </ul> <b>Dáta ako základ AI</b> <ul style="list-style-type: none"> <li>Analyzovať úlohu dát pri tréňovaní modelov a vysvetliť vzťah medzi kvalitou dát, zaujatosťou a spoľahlivosťou výstupov.</li> <li>Posúdiť dôležitosť ľudského dohľadu nad dátami a výstupmi AI modelov.</li> </ul> <b>Overovanie výstupov AI</b> <ul style="list-style-type: none"> <li>Systematicky overovať správnosť a relevantnosť výstupov AI pomocou viacerých overených zdrojov.</li> </ul> <b>Ochrana súkromia a bezpečnosť</b> <ul style="list-style-type: none"> <li>Opísať, ako algoritmy AI na pozadí (napr. na sociálnych sieťach) zbierajú, kombinujú a vyhodnocujú údaje o správaní používateľa bez jeho priameho vedomia.</li> </ul> <b>Environmentálne a sociálne vplyvy</b> <ul style="list-style-type: none"> <li>Analyzovať vplyvy AI na spoločnosť.</li> <li>Analyzovať vplyv algoritmov a mikrocielenia na polarizáciu spoločnosti, informačné bubliny a na rozhodovanie (napr. spotrebiteľské správanie alebo verejná mienka).</li> </ul>
Zdôvodnenie prepojenia	Aktivita s Google Trends priamo ilustruje, ako veľké množstvá dát o ľudskom správaní umožňujú AI systémom nachádzať vzory a predpovedať budúci vývoj. Žiaci si zažijú, že dáta, na ktorých sa AI učí, odrážajú reálne správanie ľudí v konkrétnych podmienkach. Schopnosť interpretovať vizualizácie dát a kriticky posudzovať ich výpovednú hodnotu priamo naplňa spôsobilosť systematicky overovať správnosť a relevantnosť výstupov AI. Téma veľkých dát zároveň súvisí s ochranou súkromia: vyhľadávacie dáta, nákupné správanie aj aktivity na internete sú vyhodnocované algoritmi na pozadí – bez toho, aby si to používatelia väčšinou uvedomovali. Diskusia o zdrojoch veľkých dát (sociálne siete, online obchody) sa preto spája so spôsobilosťou opísať, ako algoritmy AI na pozadí zbierajú, kombinujú a vyhodnocujú údaje o správaní používateľa bez jeho priameho vedomia. Analýza zmien vo vyhľadávaní v čase (pandémia, sezónne výkyvy, virálne udalosti) ilustruje aj to, ako agregované dáta o správaní miliónov ľudí umožňujú algoritmom ovplyvňovať spotrebiteľské rozhodovanie a verejnú mienku. Tým sa metodika prepája aj so spôsobilosťou analyzovať vplyv algoritmov a mikrocielenia na polarizáciu spoločnosti, informačné bubliny a rozhodovanie.
Poznámky a odporúčania	Metodika naplňa viaceré spôsobilosti z rôznych komponentov AI gramotnosti - okrem analýzy dát aj ochranu súkromia a kritické porozumenie vplyvu algoritmov na spoločnosť. Pre ešte priamejšie prepojenie s komponentom Princípy fungovania AI odporúčame doplniť diskusiu o tom, ako práve tieto typy veľkých dát slúžia na tréňovanie AI modelov (napr. odporúčacie systémy na sociálnych sieťach). Na metodiku nadväzuje pokračovanie z balíčka Dáta – metodika O2: Interpretácia dát pomocou četbotov, v ktorej žiaci rovnaké dáta z Google Trends interpretujú pomocou AI asistentov. Táto dvojica metodík spolu tvorí jeden didaktický celok. Metodika má tiež čiastočné prepojenie s metodikou Ju a Pí O4: Dáta a cloud, kde sa rozoberajú typy dát a princíp ich zberu, a s metodikou Ju a Pí O9: Ambientná inteligencia, ktorá sa venuje zberu dát zariadeniami na pozadí a ochrane súkromia.

## Evokácia

10 minút

Zdieľaj

### Prezentácia strana 02

**Spomeniete si, kedy Svetová zdravotnícka organizácia vyhlásila epidémiu COVID-19 za celosvetovú pandémiu?**

Odpoveď: Bolo to 11. marca 2020.

**Vyberte si začiatok obdobia pandémie. Ako sa zmenil život ľuďom okolo vás?**

Možné odpovede: Školy, obchody, reštaurácie a kiná sa zavreli. Ľudia museli väčšinou zostať doma a mohli ísť von len na nákupy alebo na prechádzky. Museli sme začať nosiť rúška. Veľa štátov zavrelo hranice a ľudia nemohli ľahko cestovať do iných krajín.

**Zmenili sa nejakým spôsobom návyky ľudí v oblasti nakupovania? Čo podľa vás ľudia nakupovali viac?**

Možné odpovede: Ľudia vstávali neskôr, pretože nebola škola a nemohli ani chodiť do práce. Často nakupovali online, pretože obchody boli zatvorené. Občas si objednávali jedlo domov. Začali nakupovať viac vecí ako rúška, dezinfekciu a napríklad aj počítače alebo tablety na online výučbu.

**Myslíte si, že rôzne krajiny a svetadiely sa líšili tým, aké výrobky tam ľudia nakupovali?**

Možná odpoveď: Pravdepodobne áno. Ľudia potrebovali rôzne veci, napríklad na základe toho, aké opatrenia boli zavedené.

## Uvedomenie

25 minút

Aktivita 1

### Prezentácia strana 03

**Premietnite žiakom stránku The New Normal.**

Vizualizuje zmeny nákupného správania pomocou vyhľadávacích dopytov používateľov vo vyhľadávачi Google pred, počas a po období pandémie koronavírusu.

Vzorce správania zaraďujú používateľov do troch kategórií:

**New Normal:** Počet vyhľadávani sa počas pandémie zmenil a zmena sa stala novou normou.

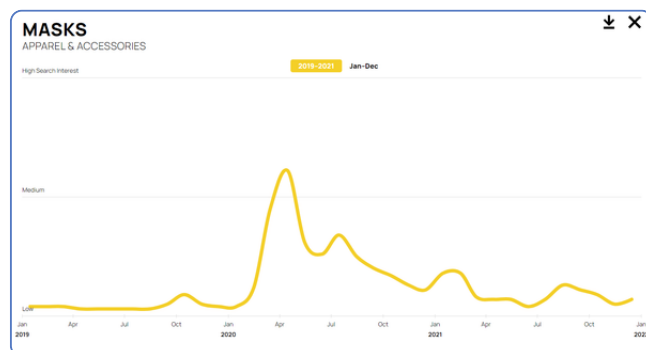
**Unusual:** Počet vyhľadávani sa počas pandémie zmenil, ale potom sa vrátil k predošlým hodnotám.

**Normal:** Počet vyhľadávani zostal pandemiou relatívne nedotknutý.

Prejdite so žiakmi všetky tri kategórie. Zamerajte sa na také výsledky (grafy), ktoré sú žiakom blízke. Odporúčame predstaviť napríklad graf vyhľadávania pojmu „masks“ (rúška).



[The New Normal](#)



Diskutujte so žiakmi o dôvodoch, prečo mali ľudia v období od januára 2020 približne do júna 2022 zvýšený záujem o vyhľadávanie pojmu „masks“ cez vyhľadávач Google. Všimajte si tiež, ako sa záujem prepadol po dvoch rokoch, teda v januári 2022.

## Aktivita 2

## Prezentácia strana 04

**Premietnite žiakom stránku What Are We Searching For.**

Stránka analyzuje správanie ľudí v priebehu dňa v rôznych častiach sveta. Zistíte napríklad, že obyvatelia Austrálie vyhľadávajú informácie o krížovkách predovšetkým ráno, podobne ako ľudia v Nemecku, Finsku alebo Francúzsku. Alebo sa dozvieme, že záujem o pizzu rastie v niektorých štátoch predovšetkým v podvečerných alebo večerných hodinách. (Pozri obrázky nižšie.)



[What Are We Searching For](#)



## Dokonči vetu

## Prezentácia strana 05

**Nedokončené vety**

Žiaci dokončia vety: 1) Vedeli sme, že... 2) Zaujalo nás, že... 3) Prekvapilo nás, že... 4) Mýlili sme sa v tom, že... 5) Nerozumeli sme...

Rozdeľte žiakov do skupín (každá skupina bude potrebovať jedno zariadenie). Žiaci si prezrú stránku What Are We Searching For a na papier zapišu doplnenie viet, ktoré vidia v prezentácii na strane 05. Následne medzi sebou v diskusii zdieľajú tie, ktoré považujú za najzaujímavejšie.

## Aktivita 3

## Prezentácia strana 06

**Analýzujte údaje z obdobia pandémie COVID-19 pomocou nástroja Google Trends.**

Pracujte s nástrojom Google Trends: [trends.google.com/trends/explore](https://trends.google.com/trends/explore)

Nástroj vizualizuje vyhľadávanie výrazov vo vyhľadávачi Google. Môžete študovať informácie o mieste a čase rôznych vyhľadávacích dopytov vrátane ich porovnaní s inými.

Postupne hľadajte výrazy uvedené v nasledujúcom zozname. Nezabudnite upraviť obdobie vyhľadávania na cca február 2020 až december 2022 (obmedzíme časové rozpätie na obdobie pandémie koronavírusu a pridáme niekoľko mesiacov navyše, aby sme videli prípadnú zmenu v ďalšom vývoji). Nechajte nastavený región Slovensko. Čísla vyhľadávania sú relatívne (nepoznáme teda presný počet vyhľadávanií, ale iba ich pomer).

**Hľadajte napríklad:** gumičky, šijací stroj, Zoom, respirátor, potraviny domov, alkohol, bežky

Nechajte žiakov interpretovať, prečo práve tieto výrazy boli takto vyhľadávané.

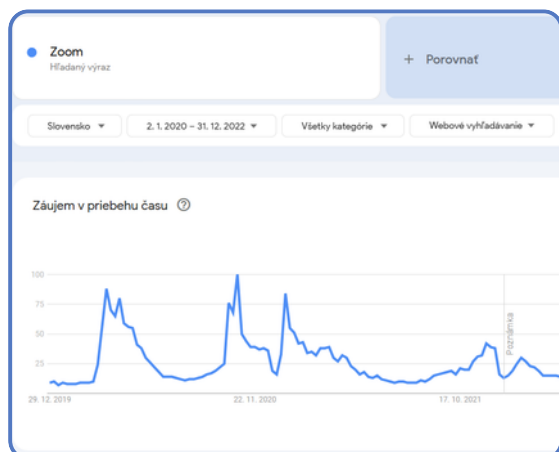
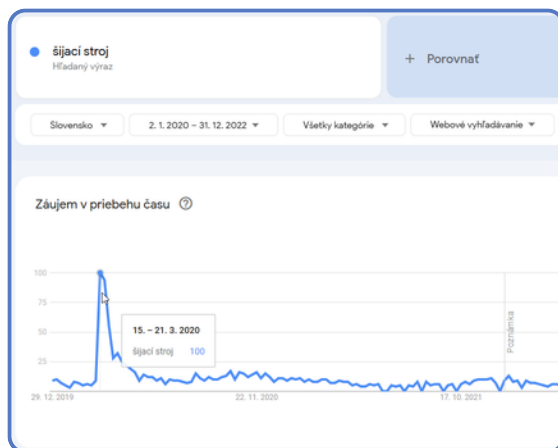
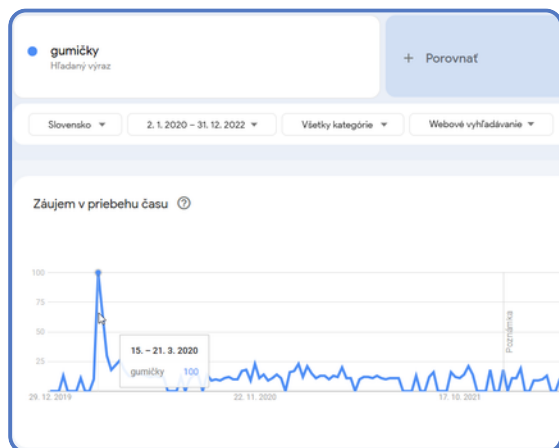


[Google Trends](#)



Prezentácia strana 07

Výsledky vyhľadávania výrazov „gumičky“, „šijací stroj“ a „Zoom“. (prezentácia strana 07).



Žiaci pracujú vo dvojiciach a hľadajú ďalšie výrazy spojené s pandémiou COVID-19.

**Lístočková metóda**  
 Žiaci zapisujú vyhľadávacie dopyty na lístočky a potom lístočky spájajú do skupín na základe podobností.

V následnej diskusii spoločne interpretujte tie najčastejšie alebo najzaujímavejšie lístočky alebo ich skupiny.

# Reflexia

10  
minút

Zamysli sa

Prezentácia strana 08

**Napadajú ti ďalšie témy alebo obdobia, ktoré by bolo zaujímavé interpretovať z veľkých dát?****Čo si predstavíš pod pojmom veľké dáta?**

Ustálená definícia pojmu veľké dáta (Big Data) neexistuje. Medzi spoločné charakteristiky veľkých dát však patrí ich ohromná veľkosť, rýchlosť ich vzniku, ich rôznorodosť (veľkosťou a štruktúrou) a často aj nižšia kvalita či menšia presnosť. Veľké dáta môžu zahŕňať obrázky, videá, audio, texty alebo tzv. digitálne stopy vo forme údajov o správaní používateľov. Sú dôsledkom zrýchlenia a vývoja internetu, pričom rolu tiež hrá výrazné zlacnenie ukladania a spracúvania dát. Typicky je ich toľko, že ich spracovanie si vyžaduje nové prístupy, napríklad metódy strojového učenia.

**Koľko vyhľadávaní na Google podľa teba ľudia uskutočnia za jeden deň?**

Podľa článku na stránke seo.ai používatelia internetu každú sekundu zadajú cca 99 000 vyhľadávacích dopytov, čo je v prepočte zhruba 8,5 miliárd vyhľadávacích dopytov denne.

**Čo si myslíš, kde inde ešte vznikajú veľké dáta?**

Možné odpovede: Keď interagujeme na sociálnych sieťach. Pri transakciách v online obchodoch alebo bankách. Vo verejnej správe a pri monitorovaní dopravy a bezpečnosti. V zdravotníctve, pri zbere dát z lekárskeho záznamov a lekárskeho zariadení. Vo verejnej správe a pri monitorovaní dopravy a bezpečnosti. Z rôznych senzorov a kamier umiestnených napríklad na verejných priestoroch alebo v budovách.

**Na čo by sme mohli ďalej využívať veľké dáta?**

Možné odpovede: Na predpovedanie počasia (senzory na celom svete merajú teplotu a pod.) alebo na predpovedanie a zmiernenie dopadov prírodných katastrof. Na zväčšenie poľnohospodárskej úrody vďaka analýze pôdných dát, klímy a používania vody...