

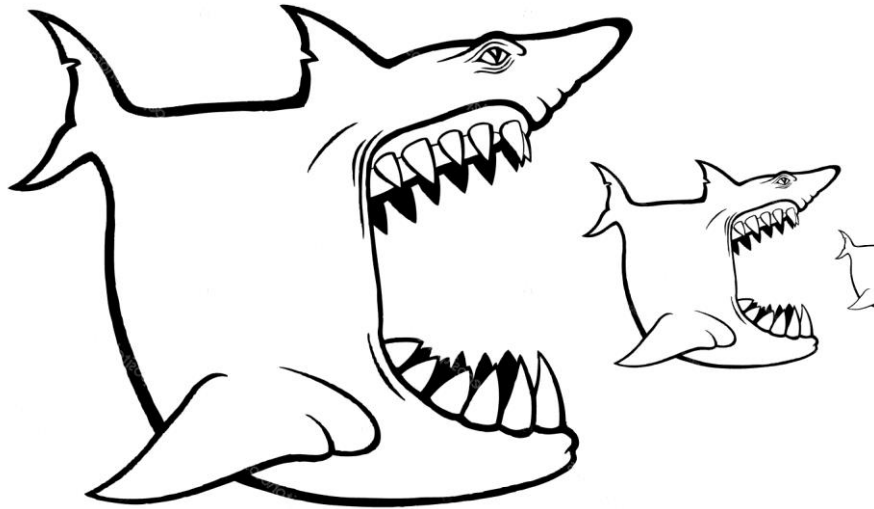
SMRT NAUČNOG METODA?

Pre gotovo deset godina, fizičar i naučni novinar Kris Anderson je u časopisu *Wired* objavio tekst koji je u najmanju ruku uznemirio klasične teoretičare naučnog metoda - najviviši njegovu smrt¹. Za ovaj „zločin“ optužio je, i pohvalio u isto vreme, razvoj informacionih tehnologija, odnosno nesagledivu količinu informacija, *big data*², koju savremeni računari bez teškoća procesiraju. Gde se *nesagledivo* uvećava velikom brzinom iz godine u godinu... Pokušajmo ipak da sagledamo, bar donekle, šta su *big data*.

¹<https://www.wired.com/2008/06/pb-theory/>

² Prema *Oxford Distionary*, big data su veoma veliki skupovi podataka koje proizvode korisnici interneta, koje je jedino moguće uskladištiti, razumeti i koristiti uz pomoć specijalnih alatki i metoda.

Petabajt generacija



Prvi komercijalni hard disk proizveo je IBM 1958. godine. Bio je težak čitavu tonu i veliki kao industrijski zamrzivač. Cena, prava sitnica, 34.000\$. Ukupni kapacitet bio mu je tek nešto veći od 4 MB, na njega je, dakle, mogla stati jedna pesma prosečnog kvaliteta ili jedna fotografija visoke rezolucije. Danas možemo kupiti disk od 8

TB, dakle *dva miliona* puta većeg kapaciteta, za 200\$, što je oko *trista miliona* puta jeftinije po jedinici informacije. Disk za dva miliona kvalitetnih fotografija? I dalje je nesagledivo. Da pokušamo s brojevima:

Jedinca	Veličina	Bajta ³
Bajt (B)	1	1
Kilobajt (KB)	1,024 ¹	1,024
Megabajt (MB)	1,024 ²	1,048,576
Gigabajt (GB)	1,024 ³	1,073,741,824
Terabajt (TB)	1,024 ⁴	1,099,511,627,776
Petabajt (PB)	1,024 ⁵	1,125,899,906,842,624
Eksabajt (EB)	1,024 ⁶	1,152,921,504,606,846,976
Zetabajt (ZB)	1,024 ⁷	1,180,591,620,717,411,303,424
Jotabajt (YB)	1,024 ⁸	1,208,925,819,614,629,174,706,176

³ Manja jedinica od bajta je bit. U njega staje binarni zapis, 0 ili 1. Bajt se sastoji od osam bita i dovoljan je da se u njega smesti jedan slovni znak.

Tek sad je nesagledivo. Doduše, i terabajt je već zastareo - eksterni hard disk od 1 TB pristupačan je čak i u našim prodavnicama, podsećam - više stotina hiljada, ako ne i nekoliko miliona fotografija, u zavisnosti od toga koliku rezoluciju preferirate!

Živimo u petabajt eri. Sa diska⁴ od 1 PB možemo slušati muziku dve hiljade godina. Ili gledati film dugačak 3,4 godine. Toliko „pamćenja“ može da stane u glavu oko 800 ljudi, procenjuju futuristi. Google *dnevno* isprocesira (samo) 20 PB⁵ kroz 5,2 milijarde pretraživanja. Ako na to dodamo 1.32 milijarde aktivnih korisnika Facebook-a koji za sobom svakoga dana ostave 4.3 milijarde postova i komentara, 5.75 milijardi lajkova i 22 milijarde tekstualnih poruka, pa dodamo Twiter i ostale društvene mreže...

Odustala bih, ali tu nije kraj. Eksabajt (EB), jedinica 1028 puta veća od petabajta, uveliko je zaživela. Isti taj Google ne čuva baš sve što kroz njega „prođe“, uskladišteno je tek nekih 10 do 15 EB.

Pri tom Google nije ni izbliza najveći svetski magacioner informacija. Istraživanja iz 2014. pokazala su da kompletan internet već živi u zetabajt eri⁶, odnosno da će do 2019. globalni saobraćaj na internetu dostići 2 ZB godišnje. Otprilike 72.000 godina visokokvalitetnog video zapisa, ukoliko je to ikom zamisliva veličina.

⁴ Tačnije, cloud-a, „oblaka“. To je model čuvanja podataka u logičkim bazama, u kojem se digitalni podaci skladište na više servera, često fizički lociranih na različitim mestima.

⁵ Radi se o proceni, nijedan pretraživač ne objavljuje egzaktne podatke.

⁶ *Cisco's Visual Networking Index initiative*

ANDERSONOVE BAJKE

Ali vratimo se Andersonu⁷ i njegovoj najavi smrti naučnog metoda. Najkraće, on je ustvrdio da trenutno na internetu raspoložemo sa tolikom količinom podataka da više nemamo nikakve potrebe za istraživanjima *zašto* se nešto događa, već je dovoljno da dokažemo da se *događa* statističkom obradom ogromne mase raspoloživih podataka. Korelacija pobeđuje kauzalitet!? Navodi primer Google, koji je osvojio svet oglašavanja ničim drugim do korišćenjem primenjene matematike. Bez pretvaranja da zna išta o kulturi i konvencijama oglašavanja, Google uvodi jednostavnu pretpostavku da sa boljim podacima i boljim analitičkim sredstvima - dobija. Što se i pokazalo kao tačno. Bazična pretpostavka pretraživača je da ne moramo da znamo zbog čega je jedna internet stranica bolja od druge, ako je broj poseta stranice veći to je dovoljno. Nikakva semantička ili kauzalna analiza nije nam neophodna. I eto kako Google nepogrešivo usklađuje reklame sa sadržajem stranice na koju ih postavlja, ne znajući pri tom ništa ni o sadržaju stranice ni o sadržaju reklame. Tako funkcioniše i Google translate. Da banalizujemo, uskladišteni su samo rečnici, prvo se tekst bukvalno prevede reč po reč, a zatim se iz bezbroj tekstova na ciljanom jeziku “prebrojavaju” situacije i fraze u kojima prevedene reči stoje jedne uz druge i bira najčešće rešenje. Google ne zna *jezike*, ne zna ni značenja pojedinačnih reči, već samo skupove slova i njihove relacije na različitim jezicima, otud u istoj rečenici možete naći imenice pobrkane po rodu, pogrešne padeže, homonime “padobrance”...

Ukratko, koga briga zašto ljudi rade to što rade? Poenta je da rade, i da to možemo pratiti sa ogromnom pouzdanošću. Sa dovoljno podataka, brojevi govore sami za sebe.

⁷ A ne Hansu Kristijanu Andersenu koji je odista pisao bajke.



Vilijem Dejvis, britanski sociolog, autor knjige *Industrija sreće*, komentariše⁸: “[...] Oglašavači su verovatno prvi koji su komercijalizovali psihološka istraživanja zarad profita. Otkako su psiholozi počeli da izučavaju pažnju krajem 19.veka, marketinški stručnjaci su usvajali njihova znanja da bi shvatili kako da prodaju što više proizvoda. Danas,

u dobu društvenih medija i aplikacija, imaju mnogo više prilika da prate kako se ljudi osećaju i ponašaju. Sektoru usluga potrebno je da zna kako se ljudi osećaju da bi mogao da održava zadovoljstvo i sreću svojih mušterija do izvesnih granica. [...] U Australiji je Fejsbuk obećao oglašavačima da će ih direktno povezati sa tinejdžerima koji su na mreži i koji pate od anksioznosti i manjka samopoštovanja, jer je procenjeno da su takvi adolescenti podložniji uticaju reklama.”

Bez obzira na to što se ponekad nasmejemo Google prevodima ili neumesnoj reklami, ova filozofija daje odlične rezultate u svetu marketinga. Međutim, Anderson svoju futurističku teoriju ležerno proširuje i na nauku smelom postavkom da petabajt era i „cloud“ računarstvo zaobilaze klasični naučni metod hipoteza - model - eksperiment

⁸ „Koliko košta sreća u kapitalizmu“, intervju, *Politika*, 25.novembar 2017.

jer naučno teoretisanje jednostavno ne može da se nosi sa poplavom podataka. „Petabajti nam dozvoljavaju da kažemo korelacija je dovoljna. I možemo prestati da pravimo modele.⁹“

Međutim, ako prestanemo da postavljamo hipoteze i kreiramo modele, da li ćemo se i dalje baviti naukom? Nauka se, za razliku od marketinga, ne bavi pronalaženjem obrazaca ponašanja - mada je i to deo naučnog procesa, već pronalaženjem objašnjenja tih obrazaca. Anderson je zapravo pogrešio čak i kada je oglašavanje u pitanju. Stručnjake za marketing možda ne zanimaju teorije ljudskog ponašanja ali oni i dalje prikupljaju i organizuju podatke na određeni način, uz pomoć Google clouda, gde ovo mora uključivati i formulaciju i testiranje određenih hipoteza. Zašto prikupljamo jedne podatke a ne druge? Zašto koristimo jednu ključnu reč za pretragu a ne drugu? Svaki načinjeni izbor odražava često neistraženi skup hipoteza o tome šta želimo i očekujemo od prikupljenih informacija. „Bez modela, matematičkih ili konceptualnih, podaci su samo buka.“ kaže genetičar i filozof nauke Massimo Pigliucci¹⁰

Primer koji je Anderson naveo kao podršku svom viđenju “nove nauke” tiče se istraživanja molekularnog biologa Krega Ventera, koji je vršio velika pretraživanja *genoma*¹¹ iz uzoraka vode iz okeana uz pomoć superbrzih uređaja za sekvencioniranje genoma. Između ostalog, otkrio je da u okeanu živi više hiljada nepoznatih vrsta bakterija. Međutim, osim toga da *postoje*, Venter o njima ne zna ništa drugo - kako žive, kako izgledaju, kakva im je morfologija. Jedino što je otkrio bili su jedinstveni genomi, različiti od svih poznatih u njegovoj bazi podataka - to

⁹<https://www.wired.com/2008/06/pb-theory/>

¹⁰<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711825/>

¹¹ U biologiji, *genom* nekog organizma su svi njegovi nasledni podaci kodirani u DNK.

dakle *moraju* biti *novi* organizmi. Jeste zanimljivo, ali sa naučne tačke gledišta to je tek sami početak, otkrivena vrata za brojna istraživanja. Puka indukcija.

Suprotstavimo ovom primeru drugi - famozni *bozon*, „kraljevski česticu“. Britanski fizičar Piter Higs 1964. godine izneo je hipotezu o postojanju ove podatomske čestice, kao objašnjenje za mehanizam po kojem elementarne čestice dobijaju masu. Postojala su, dakako, i druga predložena objašnjenja, u kojima je hipotetički bozon bio suvišan. Tek 2012. godine, dakle četrdeset osam godina nakon iznošenja Higsove hipoteze, naučnici u CERN¹² su u simuliranom proton-proton sudaru, koji se odvija sa 99% brzine svetlosti, u dva nezavisna eksperimenta uspeli da registruju trag čestice koja odgovara Higsovom bozonu. Iako je ovo otkriće u štampi predstavljeno kao potvrda karike koja je nedostajala u teoriji elementarnih čestica, eksperimenti se nastavljaju. CERN godišnje generiše preko 30 PB podataka o sudarima elementarnih čestica tragajući, između ostalog, za onima koji mogu da budu od interesa za unapređenje trenutno prihvaćenog *Standardnog modela*¹³.

¹² Evropska organizacija za nuklearno istraživanje u Ženevi.

¹³ Standardni model je teorija elementarnih čestica koja opisuje tri od četiri fundamentalne interakcije između elementarnih čestica od kojih se sastoji materija - elektromagnetizam, slabu i jaku nuklearnu interakciju. Iako dosadašnji eksperimenti, u kojima je detektovanje Higsovog bozona nedvosmisleno igralo veliku ulogu, podržavaju Standardni model, posla za naučnike ima još mnogo. Naime, četvrta interakcija, gravitacija, još nije uzeta u obzir. Otud i popularni naziv Standardnog modela - *Teorija skoro svega* 😊



Da ne ulazimo u subatomske teorije, vredno je zabeležiti i da je naša matična, građevinska, stuka dala svoj veliki doprinos čitavom projektu. Pre svega izgradnjom podzemnih objekata u koje je smešten LHC¹⁴. Na granici Švajcarske i Francuske, u blizini Ženeve, na dubinama između 50 i 175 m izgrađen je kružni tunel dužine 27 km, sa još 6 km dodatnih

tunela, trideset ogromnih podzemnih prostorija i šest vertikalnih šaftova za pristup i instalaciju opreme. Prema današnjim zamislima, njega će zameniti novi tunel, dužine 80 do 100 km, koji će negde 2035. godine postati domaćin FCC, budućem cilindričnom „sudaraču“.

Eksperimenti koje CERN izvodi odvijaju se, dakle, po klasičnom naučnom metodi. Prvo je dedukcijom postavljena hipoteza, još 1964, koja se zatim eksperimentalno dokazuje, i to kroz decenije, indukcijom, uz jasnu svest da na Standardnom modelu još uvek mora mnogo da se radi. Pa i da se sam model obori, ukoliko se dođe do novih saznanja.

¹⁴ Large Hadron Collider, veliki hadronski „sudarač“.

UMESTO ZAKLJUČKA

Nauka napreduje jedino kada je u stanju da pruži objašnjenja, ako to nije u stanju postaje aktivnost nalik skupljanju poštanskih maraka. Postoje delovi naučnih procesa u kojima petabajti informacija mogu igrati značajnu ulogu, ali oni sami po sebi nisu nauka.

Suprotstavljanje petabajta (eksabajta, zetabajta... i svega što nam budućnost informacionih tehnologija donosi) i klasičnog naučnog metoda, u smislu dva alternativna puta moderne nauke, potpuno je pogrešno. Niti napretka informacionih tehnologija ima bez akvizicije znanja u cikličnom procesu u kojem se smenjuju deduktivne i induktivne faze, odnosno uz primenu klasičnog naučnog metoda, niti moderna nauka može da napreduje bez pomoći brzih pretraživača i izobilja informacija koje im informacione tehnologije nude.

Smrt naučnog metoda, dakle, nije na vidiku, naprotiv, za njega se razvojem informacionih tehnologija otvaraju nove mogućnosti.