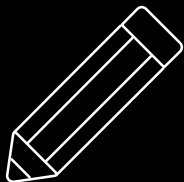


11. Konference o šedé literatuře a repozitářích
24. října 2018, Praha



JSME PŘIPRAVENI NA BUDOUCNOST?

Vliv umělé inteligence na management šedé literatury

Dr. Dobrica Savić

[linkedin.com/in/dobricasavic](https://www.linkedin.com/in/dobricasavic)



**Proč se
zajímáme o vliv
umělé
intelligence na
šedou literaturu?**



- **SAP CoPilot – Digital assistant for Enterprise (Digitalní asistent pro podnikání)**
Uživatel může pokládat dotazy ohledně podnikání po telefonu a systém sám nabídne odpověď.
- **Deloitte - Automated document review (Automatizovaný souhrn dokumentu)**
S využitím přirozeného jazyka pročítá tisíce komplexních dokumentů, extrahuje a strukturuje textové informace pro lepší analýzu.
- **AI Sense: Call and meeting transcriptions (Přepis hovorů a schůzek)**
Zaznamenává hlasové konverzace a umožňuje je prohledávat a zpřístupňovat pomocí automatického rozpoznávání řeči, identifikace řečníků, synchronizace řeči a textu a zpracování přirozeného jazyka.
- **UK Press Assoc.: RADAR (Reporters and Data and Robots)**
Roboti napíší 30,000 lokálních zpráv měsíčně, při jejich tvorbě využívají nejruznější data poskytovaná vládou, veřejnými službami a lokálními úřady. Tyto zprávy nejsou kontrolovány lidmi.
- **Google: Global Fishing Watch (Globální sledování rybolovu)**
Každý den zaznamená 22 milionů datových bodů, které vypovídají o tom, kde se pohybují lodě po trasách po celém světě, proč daná loď vyplula a tím pomáhá odhalovat nelegální rybolov.



- 1 **Lyrebird: vytvořte digitální kopii svého hlasu**
Zaznamenejte si alespoň minutu a vygenerujte jakoukoliv větu vlastním digitálním hlasem.

- 2 **Sidekick.ai: vytvořte AI verzi sebe sama**
Pomáhá vám lépe se rozhodovat, zjednodušit svůj život a je jedinečný jako vy.

Who are you?

I am an AI version of you

Whoah, what can you do?

I can help you make better decisions and simplify your life

How?

By analysing and bringing your data alive!

- 3 **Google's AlphaGo** - v roce 2016 porazil mistra ve hře Go Lee Sedola (18x mistr světa) 4: 1. Existuje 10^{170} možných tahů, zatímco v pozorovatelném vesmíru je pouze 10^{80} atomů. O rok později AlphaGo Zero porazí původní AlphaGo 100: 0. Naučil se hrát bez lidské interakce sám proti sobě pomocí nové formy **posilování učení**. Na začátku je systém s neuronovou sítí, která o hře Go neví nic. Poté hraje hry proti sobě a dramaticky se zlepšuje, protože není omezen hranicemi lidského poznání!

Prezentace - přehled

Definice

- Šedá literatura (ŠL)
- Umělá inteligence (AI)

Vliv AI

Tvorba ŠL

- Různorodost
- Objem
- Důvěryhodnost
- Rychlost
- Hodnota

Management ŠL

- Zpracování
- Udržitelnost
- Využitelnost

Závěry



Definice: Šedá literatura



*ŠL je jakýkoliv zaznamenaný,
odkazovatelný a udržitelný datový nebo
informační zdroj se současnou nebo
budoucí hodnotou, který neprošel
tradičním recenzním řízením a je veřejně
dostupný.*

Definice: Umělá inteligence



“

- “Věda o výrobě strojů zvládajících věci, které by vyžadovaly inteligenci, kdyby je měli dělat lidé.” *Marvin Minsky, 1968*
- “AI je zabývá tím, jak vyrobit počítače, které by dělaly věci, které v současné době lidé zvládnou lépe.” *Elaine Rich, 2010*

Definice: Umělá inteligence

- **AI** – Širší koncept, kde stroje jsou schopné vykonávat úlohy „chytrým“ způsobem
- **≠**
- **Strojové učení** – uplatnění AI kde se stroje učí díky přístupu k velkému objemu dat
- Rozvíjení a nasazení bývá nákladné a složité
- Kombinace mnoha technologií a technik, např. paralelní výpočty, hluboké učení, neuronové sítě, zpracování přirozeného jazyka
- Původně expertní systémy založené na pravidlech
- Dnešní algoritmy mohou pochopit, učit se, předvídat, přizpůsobit se a fungovat autonomně



Definice: Umělá inteligence

- **Vestavěné do fyzických zařízení**

Roboti, auta, spotřební elektronika, bezpečnostní systémy

- **Zabudované do aplikací a služeb**

- Vyhledávače, cílení reklam
- Virtuální osobní asistenti
- Chatboty, chytří poradci
- Doporučování obsahu
- Rozpoznávání hlasu, počítačové vidění
- Překladače, kontrola pravopisu a gramatiky
- Finance
- Přizpůsobivé uživatelské rozhraní
- Hry



Definice: Umělá inteligence

Typy AI podle complexity

- **Úzká AI**

System vyvinutý a vyškolený pro konkrétní úkol ve vymezené oblasti

- **Obecná AI**

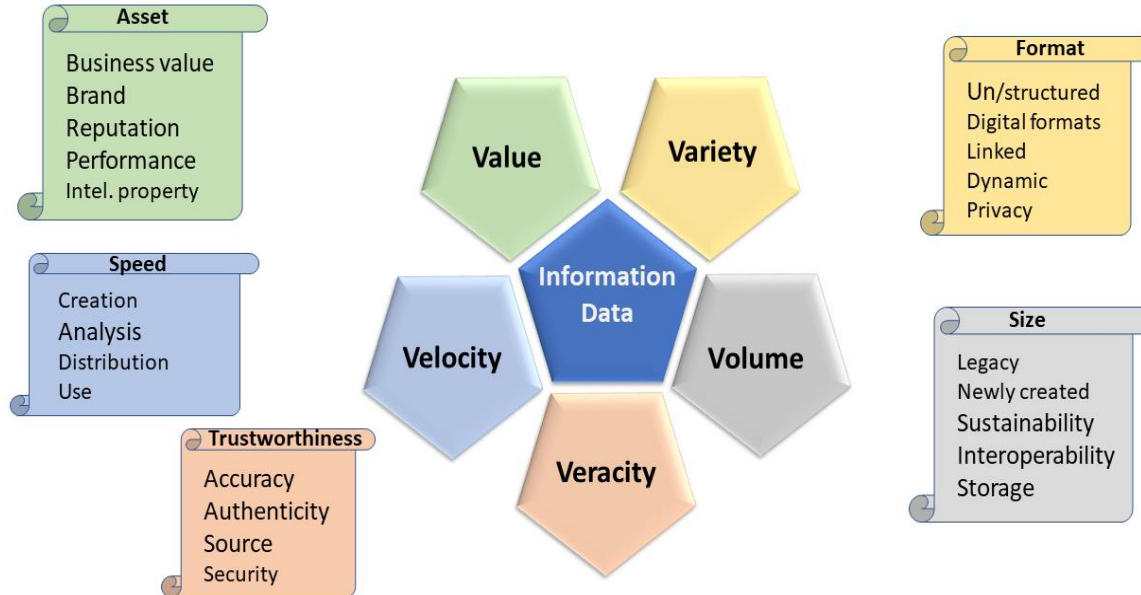
System, který může chápat a rozumět svému okolí stejně jako člověk

- **Super AI**

System mnohem chytřejší než nejlepší lidské mozky prakticky v každé oblasti



5 Vs of Data/Information



Vliv AI na tvorbu ŠL - Různorodost

Types of grey literature

Bibliographies	Rejected manuscripts	Publications from NGOs and consulting firms
Discussion papers	Un-submitted manuscripts	Videos
Newsletters	Conference abstracts	Wiki articles
PowerPoint presentations	Book chapters	Emails
Program evaluation reports	Personal correspondence	Blogs and social media
Technical notes	Newsletters	Data sets
Publications from governmental agencies	Informal communications	Committee reports
Reports to funding agencies	Census data	Working papers
Unpublished reports	Pre-prints	Company reports
Dissertations	Standards	Catalogues
Policy documents	Patents	Speeches
	Webinars	Reports on websites

The GreyNet International website lists over 150 document types specific to GL.

- **Zdrojem dat** a informací jsou internet věcí (IoT), počítačová komunikace (M2M), samořízené automobily, roboty, senzory, bezpečnostní systémy, kamery a mnoho dalších systémů nebo aplikací využívajících AI.
- **Počet připojených zařízení** vytvářejících specifická data se pohybuje kolem miliardy.
- Množství dat, různých formátů, nestrukturovaná data, specifické aplikace - **výzva** pro vězkumníky ŠL, odborníky z praxe a manažery.
- Data a informace, která velmi **závisí na kontextu a použitém softwaru**, se těžko shromažďují a zpracovávají, a dokonce je těžší je pochopit a zachovat pro budoucí použití.
- Vzájemně propojené data sety (**linked data**), hlavní složka sémantického webu, musí být srozumitelné lidem a počítačům.
 - Dobrým příkladem rozsáhlé propojené data setů je DBpedia, která zpřístupňuje obsah Wikipedie v RDF (Resource Description Framework) a obsahuje odkazy na jiné data sety na webu (např. GeoNames). Poskytnutím těchto dodatečných odkazů nabízí aplikace mnohem lepší přístup ke znalostem a uspokojivější uživatelskou zkušenost.
- Identifikace a nalezení role pro **klasický management ŠL** se stává výzvou.

Vliv AI na tvorbu ŠL - Objem



Jsme připraveni
zvládnout petabyty
informací a dat

	Amount per minute
Forecast requests received by The Weather Channel	18,055,555
Text messages sent	12,986,111
Videos watched by YouTube users	4,333,560
Google searches conducted	3,788,140
GB of internet traffic generated by Americans	3,138,420
Snaps shared by Snapchat users	2,083,333
GIFs served by GIPH	1,388,889
Songs streamed on Spotify	750,000
Tweets sent by Twitter users	473,400
Calls made by Skype users	176,220
Hours of video streamed on Netflix	97,222
Posts published by Tumblr users	79,740
Dollars processed via Venmo P2P transactions	68,493

- Každý den vzniká 2,5 exabytů dat, ekvivalentem je 250 000 Kongresových knihoven.
- Lidský mozek má odhadovanou kapacitu 1000 terabytů nebo 1 petabyte.
- 90% všech dat na světě bylo vytvořeno za poslední 2 roky.
- Na světě existuje 130 milionů publikovaných knih, ročně přibývá 800,000 nových titulů.
- Každý den je odesláno a doručeno 269 miliard emailů, 60% je spam.
- 56% veškerého internetového provozu tvoří automatizované nástroje – hackerské nástroje, scrapery, spammy, boty.
- Svět hladoví po informacích! Společnost Google zpracovává denně více než 6,6 miliard dotazů, z čehož 15% dotazů nebylo nikdy předtím položeno.

1 EXABYTES = 1,000 PETABYTES
1 PETABYTES = 1,000 TERABYTES
1 TERABYTES = 1,000 GIGABYTES
1 EXABYTES = 1,000,000,000 GIGABYTES

Vliv AI na tvorbu ŠL - Důvěryhodnost



Jsme připraveni
garantovat přesnost a
pravdivost

Platnost a důvěryhodnost

- Definována jako "**shoda s fakty; přesnost; obvyklá pravdivost**".
- Zabývá se přesností, autentičností, zdroji informací a bezpečností.
- Odhalování podvodů a odhadování pravdivosti informací a údajů je nyní obtížné a bude v budoucnu ještě obtížnější.

Kritická místa dat a informací vytvořených pomocí AI

- **Černá skříňka**
Neexistují žádné zdokumentované postupy nebo kroky, neznáme způsob odvození cesty, žádné odůvodnění rozhodnutí, žádná stopa použité logiky nebo odůvodnění.
- **Algoritmické zkreslení**
Zdánlivě neutrální matematická podstata, není ale objektivnější než lidé. Algoritmy jsou napsané lidmi. Lidé shromažďují a dodávají údaje požadované pro ML a přenášejí na ně své hodnoty, přesvědčení, kulturu, inteligenci.
Mohou algoritmy přinést správné a spravedlivé výsledky?

- ✓ spam email
- ✓ fake news
- ✓ computer bots
- ✓ botnets
- ✓ web spiders
- ✓ crawlers
- ✓ viruses
- ✓ misinformation
- ✓ disinformation

Vliv AI na tvorbu ŠL - Rychlost

Rychlost vzniku, zpracování, analýzy, šíření a využití informací.

Umíme se vyrovnat s
takovou rychlostí
vytváření dat?

- V r. 1993 trvalo 28 hodin stažení nekvalitního filmu (700 MB) prostřednictvím modemu a telefonického připojení přes telefonní linku rychlostí 56 kb/s.
- Předpověď pro rok 2020 předpovídá rychlost stahování 1 Tb/s. Celé TV seriály s více řadami v HD by mohly být staženy za pouhou vteřinu..
- Do roku 2020 každou vteřinu vznikne zhruba 1,7 megabajtů nových informací na každého člověka na planetě.
- 99,5% všech vytvořených dat není v současné době analyzováno ani používáno.
- Technické, fyzické, finanční a další překážky omezují možnost analyzovat tak velké množství dat vytvářených takovou rychlostí.
- Velká finanční, podnikatelská a informační ztráta pro všechny zúčastněné.
- Velké množství informací a dat umožňuje, aby AI posunula datovou analýzu od retrospektivní praxe k proaktivnímu přístupu a strategickému rozhodování.
- AI může výrazně zvýšit frekvenci, flexibilitu a rychlost analýzy dat v celé řadě průmyslových odvětví a aplikací.

Vliv AI na tvorbu ŠL - Hodnota



Informační aktivum (IA)

Identifikovatelná sbírka dat uchovávaná jakýmkoliv způsobem a uznaná jako hodnota s cílem umožnit agentuře vykonávat své obchodní funkce, čímž splňuje požadavek uznaného subjektu.

- Informace jsou hodnotným majetkem, stojí miliony.
 - Obtížně měřitelné a finančně vyčíslitelné.
 - Dá se těžko určit umístění, jejich množství, kvalita nebo dokonce původ.
 - Mnozí tvrdí, že vlastní IAs, ale snaží se vyhnout odpovědnosti za ně.
-
- V roce 2020 bude mít 10% organizací obchodní oddělení pro vytváření a komercializaci svých informačních aktiv.
 - IA mají další velký potenciál kromě účelu, pro který byly vytvořeny.
 - Na rozdíl od většiny ostatních podnikových aktiv mají informace hodnotu i po jejich využití.
-
- Díky zapojení AI mohou být manažeři ŠL více angažovaní a spokojení, stráví méně času s nudnými opakujícími se úkoly, jako je shrnutí článků, označování obsahu a přidávání popisných metadat.
 - Místo toho mohou pracovat na kreativních úkolech a vytvářet nový, dobře koncipovaný a cenný obsah.
 - Týmy jsou produktivnější a přispívají k celkové efektivitě.

Vliv AI na management ŠL



1

Zpracování ŠL

Požadavky - Jednotlivé dokumenty a ad hoc přístupy k řízení nejsou ani vhodné ani dostatečné. Zpracování už během tvorby ŠL by bylo vhodnější než následné zpracování. Žádné ad hoc řešení nebo odchylky.

2

Udržitelnost ŠL

Výzvy - environmentální a technické; ekonomické a finanční; sociální nebo organizační.

3

Využitelnost ŠL

Otázky - vhodné IT nástroje, dostupnost kvalifikovaných lidských zdrojů, ochrana duševního vlastnictví a soukromí.



Příležitosti

Spolehlivá automatizace opakujících se úkolů, velká přesnost, žádná únava, přidaná inteligence, sémantické porozumění, výkonná analýza, zdokonalení učebních algoritmů, vytváření nových poznatků, extrakce nové hodnoty z existující ŠL.

Závěry

- AI bude mít obrovský dopad na způsoby, jakými vytváříme, zpracováváme, šíříme, řídíme a využíváme ŠL
- Počet typů ŠL, její objem a rychlost vzniku poroste. Musí být zachována důvěryhodnost a hodnota.
- Principy zpracování musí být dobře promyšlené a přítomné od začátku tvorby ŠL. Žádné ad-hoc nebo post-zpracování.
- Pokud má být poskytnuta dlouhodobá udržitelnost, je třeba vzít v úvahu environmentální a technické, ekonomické a finanční, sociální i organizační omezení.
- Použitelnost ŠL závisí na existenci vhodných IT nástrojů, dostupnosti kvalifikovaných lidských zdrojů, podmínkách ochrany duševního vlastnictví a ochrany osobního soukromí
- Pro zajištění budoucnosti a udržení hodnoty ŠL je třeba organizovat intenzivní školení, zajistit širokou spolupráci a dbát na přípravu
- Pouze malé procento podniků dokáže vytěžit z informací, které vlastní, plnou hodnotu. Použití AI v řízení ŠL může organizacím pomoci co nejlépe využít a zlepšit své obchodní výsledky, zajistit zvýšení efektivity, kvality produktů a služeb





AI je nejspíš buď ta nejlepší nebo ta
nejhorší věc, která lidstvo potkala!

Stephen Hawking

Děkuji!