



národní
úložiště
šedé
literatury

Are We Ready for the Future? Impact of Artificial Intelligence on Grey Literature Management

Savić, Dobrica
2018

Dostupný z <http://www.nusl.cz/ntk/nusl-387433>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

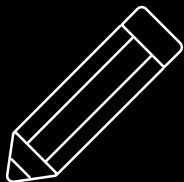
Licence Creative Commons Uveďte původ-Zachovejte licenci 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 09.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz.

11th Conference on Grey Literature and Repositories
24 October 2018, Prague



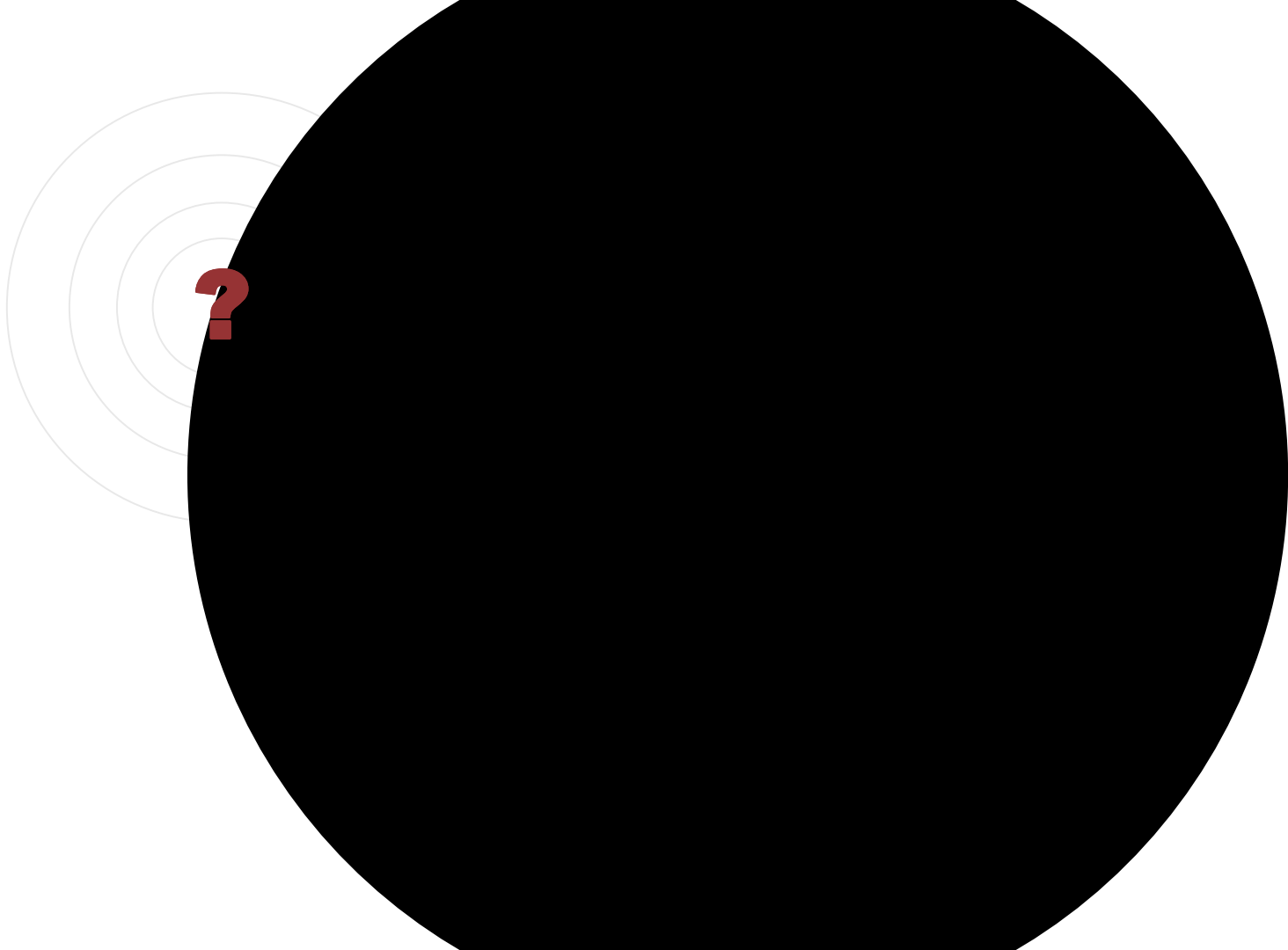
ARE WE READY FOR THE FUTURE?

Impact of Artificial Intelligence on
Grey Literature Management

Dr. Dobrica Savić

This presentation is licensed under the Creative Commons:
CC-BY-SA-4.0, via <http://repozitar.techlib.cz/record/1287>

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





- **SAP CoPilot - Digital Assistant for the enterprise**
Using phones, users can ask business related questions and the system offers an answer.
- **Deloitte - Automated document review**
Using natural language processing reads thousands of complex documents, extracts and structures textual information for better analysis.
- **AI Sense: Call and meeting transcriptions**
Records voice conversations and makes them searchable and accessible using automatic speech recognition, speaker identification, speech-and-text synchronization, and natural language processing.
- **UK Press Assoc.: RADAR (Reporters and Data and Robots)**
Robots write 30,000 local news stories each month fed with a variety of data from government, public services, and local authorities. These news stories are not covered by humans.
- **Google: Global Fishing Watch**
On any given day, 22 million sea data points are created that show where ships are in the world's waterways, why a vessel was at sea, and when fishing is happening illegally.



- 1 Lyrebird: Create a digital copy of your voice**
Record yourself for at least one minute and generate any sentence you like with your own digital voice.

- 2 Sidekick.ai: Make an AI version of yourself**
Helps you make better decisions, simplify your life, and it is unique like you.

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** - in 2016 beats Lee Sedol (18-time world champion) 4:1. There are 10^{170} possible moves, while there are 10^{80} atoms in the observable universe. Only one year later **AlphaGo Zero** beats original AlphaGo 100:0. It learned to play without human interaction by playing against itself! It was able to do this by using a novel form of **reinforcement learning**. The system starts off with a neural network that knows nothing about the game of Go. It then plays games against itself and dramatically improves since it is not constrained by the limits of human knowledge!

Presentation at a Glance

Definitions

- Grey literature (GL)
- Artificial intelligence (AI)

AI impact

GL creation

- Variety
- Volume
- Veracity
- Velocity
- Value

GL management

- Processing
- Sustainability
- Usability

Conclusions



Definition: Grey literature

C

Definition: Artificial intelligence

•

•

•

Definition: Artificial intelligence

- **AI** – A broader concept where machines are able to carry out tasks in a way that we would consider “smart”
- **Machine learning** – an application of AI where machines learn by accessing big data
- Costly and complex to develop and deploy
- Combination of many technologies and techniques, e.g. parallel computing, deep learning, neural networks, natural language processing
- Initially rule-based expert systems
- Today's algorithms can understand, learn, predict, adapt, and operate autonomously



Definition: Artificial intelligence

- **Built into physical devices**

Robots, cars, consumer electronics, security systems

- **Built into apps and services**

- Search engines, ads targeting
- Virtual personal assistants
- Chatbots, smart advisors
- Content recommendation
- Voice recognition, computer vision
- Translation, spelling, grammar checking
- Finance
- Adaptive user interfaces
- Games



Definition: Artificial intelligence

Complexity based types of AI

- **Narrow AI**

A system developed and trained for a particular single task within a limited domain

- **General AI**

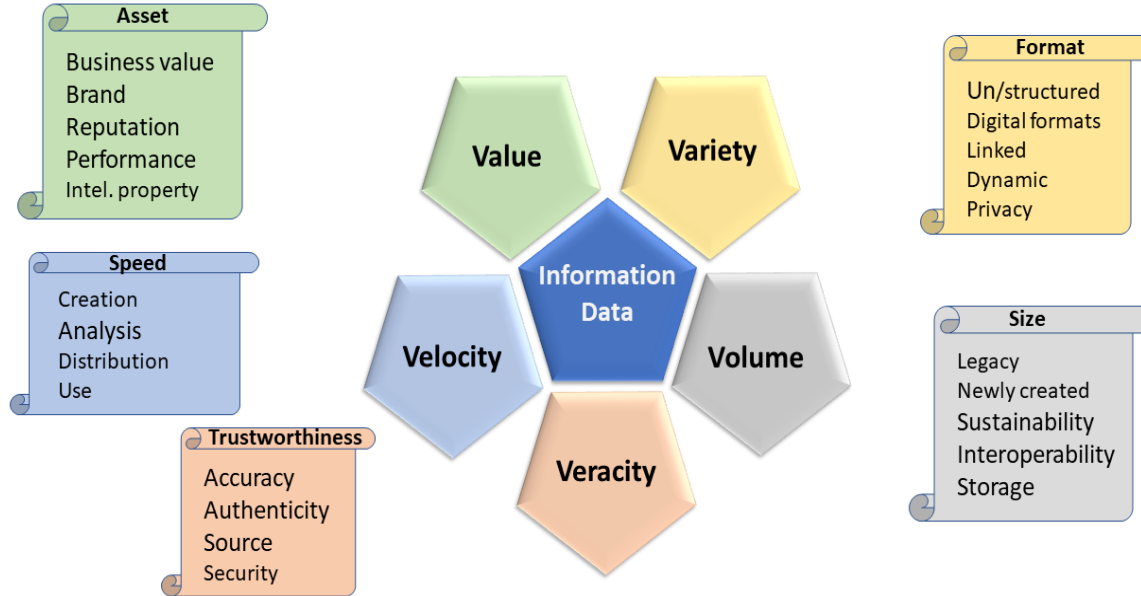
A system that can understand and reason about its environment as a human would

- **Super AI**

A system much smarter than the best human brains in practically every field



5 Vs of Data/Information



AI impact on GL creation - Variety

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.

Are we ready to manage so many new types and formats

- **Sources** of data and information are the Internet of Things (IoT), Machine to Machine communication (M2M), self-driven cars, robots, sensors, security systems, surveillance cameras, and many other systems or apps using AI.
- **Number of connected devices** creating specific data vary by billions.
- Tons of data, multiple formats, unstructured, application specific - **challenge** for GL researchers, practitioners, and managers.
- **Highly contextual and software dependent** data and information is hard to collect and process, and even harder to make sense of and preserve for future use.
- Interconnected data sets (**linked data**), the main ingredient of Semantic Web, needs to be understandable to humans and to computers.
 - A good example of a large linked data set is DBpedia, which makes the content of Wikipedia available in the Resource Description Framework (RDF), while including links to other datasets on the Web (e.g. GeoNames). By providing those extra links, the application offers much better access to knowledge and a more satisfying user experience.
- Identifying and finding a **role for classical GL management** becomes a challenge.

AI impact on GL creation - Volume



Are we ready to
manage petabytes of
information and data

	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

- 2.5 exabytes of data are produced every day, equivalent to 250,000 Libraries of Congress.
- The human brain has an estimated storage capacity of 1000 terabytes or one petabyte.
- 90% of all the data in the world has been generated over the last two years.
- There are 130 million published books around the world, with over 800,000 new titles added annually.
- 269 billion emails sent and received each day, out of which 60% is spam.
- 56% of all internet traffic is from automated sources such as hacking tools, scrapers and spammers, bots.
- World is hungry for information! Google processes over 6.6 billion queries daily, out of which 15% were never searched before.

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



Are we ready to
guarantee accuracy
and truthfulness

Validity and trustworthiness

- Defined as “*conformity to facts; accuracy; habitual truthfulness*” .
- Deals with accuracy, authenticity, information source, and security.
- Uncovering deception and estimating the veracity of information and data is difficult now and will be even more difficult in future.

Challenges with AI created data and information

- **Black box**
No documented procedures or steps followed, inference paths unknown, no decision justifications. No trace of the logic or reasoning used.
- **Algorithmic bias**
Seemingly neutral mathematical nature, they are not more objective than humans. They are written by people. People collect and feed the data required for ML and bring over their values, beliefs, culture, intelligence.
Can algorithms bring fair results?

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

AI impact on GL creation - Velocity

The speed of information creation, processing, analysis, distribution, and use.

Can we cope with such
a speed of data creation

- In 1993 a low-quality movie (700 MB) took 28 hours to download through a dial-up Internet connection through 56 kbps modems via phone line.
- Forecast for 2020 is speed of 1 terabit-per-second. An entire multi-season HD TV series could be downloaded in just one second.
- By 2020 about 1.7 megabytes of new information will be created every second for every human being on the planet.
- 99.5% of all data created is not currently being analysed and used.
- Technical, physical, financial, and other challenges limit the possibilities for analysing such a huge amount of data created at such a speed.
- A big financial, business, and information loss for everyone involved.
- Huge amount of information and data enables AI to turn data analysis from retrospective practice into a proactive approach and strategic decision making.
- AI can greatly increase the frequency, flexibility, and immediacy of data analysis across a range of industries and applications.

AI impact on GL creation - Value



Information asset (IA)

An identifiable collection of data stored in any manner and recognized as having value for the purpose of enabling an agency to perform its business functions thereby satisfying a recognized agency requirement.

- Information is an asset that costs millions.
 - Difficult to measure and financially evaluate.
 - Hard to tell where it sits, its quantity, its quality, or even it's origin.
 - Many claim to own the IAs but try to avoid any accountability for it.
-
- By 2020 10% of organizations will have a business unit for making and commercializing their IAs.
 - IAs have a great potential beyond the utility for which they were originally produced.
 - Unlike most of the other enterprise assets, information isn't depleted after consumed.
-
- Using AI, GL managers can be more engaged and happier, spending less time working on tedious, repetitive tasks such as summarizing articles, tagging content and adding descriptive metadata.
 - Instead, they can work on creative tasks, producing fresh, well-conceived and valuable content.
 - Teams are more productive and contribute to overall efficiency.

AI impact on GL management



1

GL processing

Demands - single document and ad-hoc management approaches are neither appropriate nor sufficient. Processing during GL creation, rather than post-processing is required. A GL management system needs to be in place all the time. No ad-hoc solutions or deviations.

2

GL sustainability

Challenges – environmental and technical; economic and financial; social or organizational.

3

GL usability

Issues - adequate IT tools, the availability of qualified human resources, and the protection of intellectual property and privacy.



Opportunities

Reliable automation of repetitive tasks, great accuracy, no fatigue, added intelligence, semantic understanding, powerful analytics, improvement of learning algorithms, creation of new knowledge, extraction of new value from existing GL.

Conclusions

- AI will have an immense impact on the way we create, process, disseminate, manage and use GL
- The number of GL types, its volume, and the speed of creation will increase. Trustworthiness and value will need to be maintained
- Processing needs to be well-thought and present from the beginning of GL creation. No ad-hoc or post-processing
- Environmental and technical; economic and financial; as well as social or organizational constraints need to be taken into consideration if long-term GL sustainability is to be provided
- Usability of GL depends on the existence of adequate IT tools, the availability of qualified human resources, the protection of intellectual property and the protection of personal privacy
- To secure the future and maintain the value of GL, intensive training, wide cooperation, and rigorous preparation need to be organized
- Only a small percent of businesses extract full value from the information they hold. Use of AI in GL management might enable organizations to get the most out of and improve their business results, efficiency gains, quality of products and services





**AI is likely to be either the best or
worst thing to happen to humanity!**

Stephen Hawking

Thanks!