IARIA

NexTech 2020 Congress

October 25 - October 29, 2020 - Nice, France

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

Chair

Javier Fabra - Universidad de Zaragoza, Spain (jfabra@unizar.es)

Panelists

Gregor Grambow - Aalen University, Germany (gregor.grambow@hs-aalen.de)

Tim vor der Brück - FFHS, Lucerne University of Applied Sciences and Arts, Switzerland (vdb1@gmx.de)

Minoru Sasaki - Ibaraki University, Japan (minoru.sasaki.01@vc.ibaraki.ac.jp)

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech
2020

Chair

Javier Fabra, Universidad de Zaragoza

jfabra@unizar.es

- Data analysis & process mining

- Service-oriented computing

- Cloud architectures

- eHealth systems & services

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech
2020

- Three interesting panels covering hot topics:
  - Big Data Analytics
    - Ethical questions and problems
  - The dominarion of Google and Facebook in Data Science
    - Frameworks for neural networks (Tensorflow/PyTorch)
    - Ethics & implications
  - Utility of dictionaries to understand the meaning of words
    - Sentence extensions & sense definition expansion
    - Identification of phrases with one meaning

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

Panellist Position

Big Data Analytics: Ethical Questions and Problems

Gregor Grambow, Aalen University, Germany

gregor.grambow@hs-aalen.de

- Big data

- Massive-parallel data processing

- Modern database technologies

- Data analytics

- Data and software modelling

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

IARIA

Panellist Position

Curse or Blessing? Domination of Google and Facebook in Data Science

Tim vor der Brück, Lucerne University of Applied Sciences and Arts, tim.vorderbrueck@hslu.ch

FFHS (Distance University of Switzerland), tim.vorderbrueck@ffhs.ch

A lot of data scientists are using software and data from Google and Facebook

• Google and Facebook dominate the frameworks for neural networks with Tensorflow and PyTorch

• Pretrained Bert models provided by Google are widely employed by NLP applications

• Do both campanies act for the good of mankind?

• Became both companies just too powerful nowadays?

• Should the governments implement counter-measures or not?

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

Panellist Position

Understanding the Meaning of Words Using Dictionary

Minoru Sasaki, Ibaraki University, Japan

minoru.sasaki.01@vc.ibaraki.ac.jp

- Find out the meaning of words

- Example Sentence Extensions for Words

- Sense Definition Expansion

- Identification of Phrases with One Meaning

→ Dictionaries will be more useful.

→ Humans and computers will be able to capture the characteristics of word meanings and differences between them.

→ The improved dictionary will be useful for downstream tasks such as machine translation and information Extraction.

Big Data Analytics: Ethical Questions and Problems

Gregor Grambow,

Department of Computer Science,

Aalen University, Germany

Big Data Analytics:
Ethical Questions and Problems

- Ethics - derived from the ancient greek word *ēthikós* (ἠθικός), meaning "relating to one's character", which itself comes from the root word *êthos* (ἦθος) meaning "character, moral nature"

- Information Ethics - the branch of ethics that focuses on the relationship between the creation, organization, dissemination, and use of information, and the ethical standards and moral codes governing human conduct in society

Joan, Reitz M. "Information Ethics." Online Dictionary For Library And Information Science. N.p., 2010

Big Data Analytics:
Ethical Questions and Problems

- Information ethics governs how companies treat and analyze our data
  - Influenced and changed by ongoing developments

Gregor Grambow, Aalen University
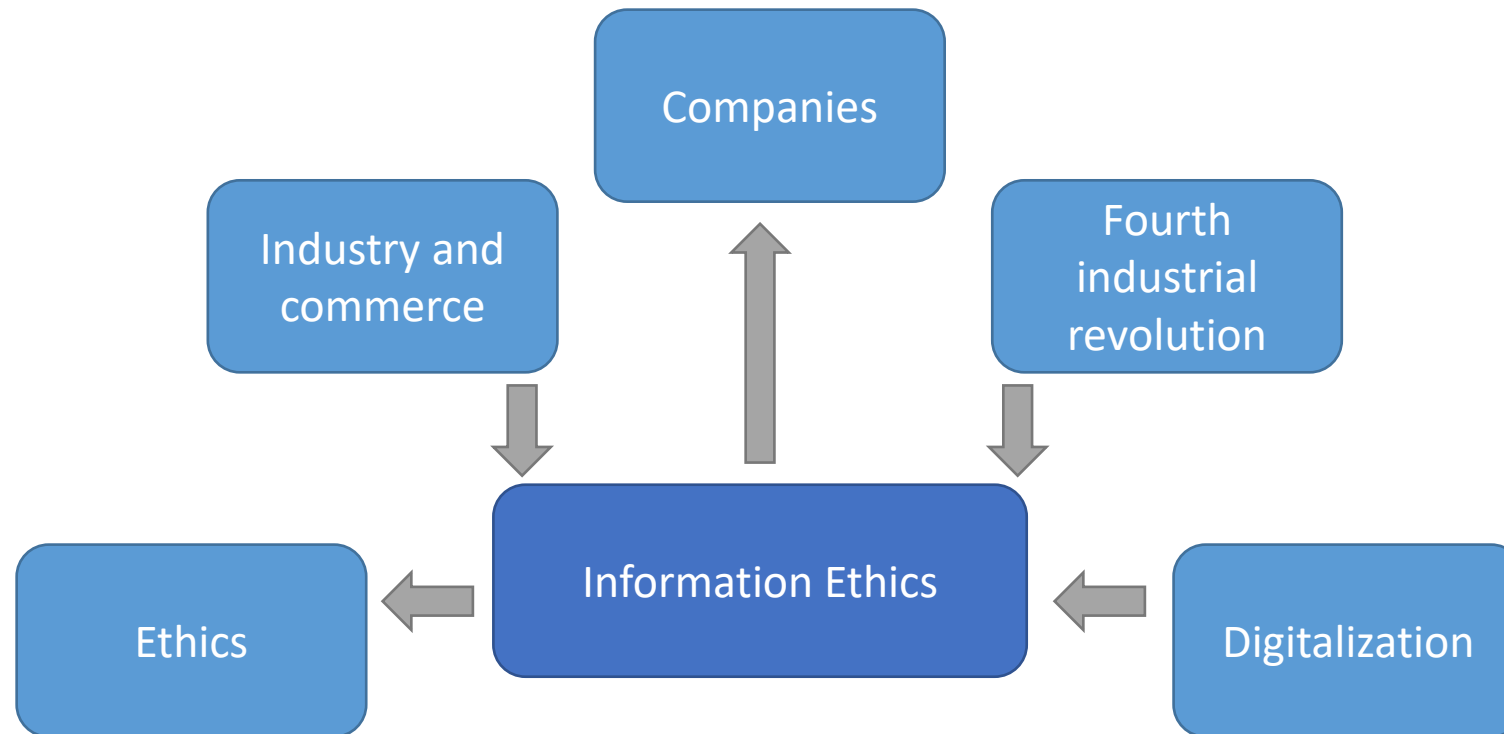
Big Data Analytics:
Ethical Questions and Problems

- Power of knowledge
  - Comes from analytics, not the data itself
  - Big Data = various sources
  - → Intransparent

- Privacy
  - Many countries have laws for protecting the individual
  - Difficult to measure and enforce

Big Data Analytics:
Ethical Questions and Problems

- Digital identities
  - Enable inference of the individuals behaviour
  - Intransparencies
    - Which data is used / combined?
    - For what?

- Manipulation
  - Manipulative information processing
    - More data – more options
  - Individuals face strong pressure to adapt to new paradigms

Big Data Analytics:
Ethical Questions and Problems

- Unknown correlations
  - Modern algorithms find new ones
  - Often intransparent
  - Relation to the source data difficult to establish

- Training data
  - Algorithms learn on the basis of human-selected data
  - Result of prior decisions
  - Can already contain discrimination, racism, …
  - Will be applied in all futures decisions

Big Data Analytics:
Ethical Questions and Problems

- Limitation of decision-making independence
  - Big data processing can alter and limit many factors for the individual
  - Pre-filtered offers for consumers
    - Limited options
    - Automatic price adaptation
  - Individual chooses – but has no choice other than using pre-computed options and following recommendations?

- Growing insecurity
  - Systems control many factors
  - Individuals don't know which ones

Big Data Analytics:
Ethical Questions and Problems

- Individuals need more transparency

- More public discussion is mandatory

- Companies should apply code of conduct for data processing

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech
2020

Curse or Blessing? Domination of Google and Facebook in Data Science

Tim vor der Brück

Lucerne University of Applied Sciences and Arts, tim.vorderbrueck@hslu.ch

FFHS (Distance University of Switzerland), tim.vorderbrueck@ffhs.ch

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

Curse or Blessing? Domination of Google and Facebook in Data Science

Nowadays, a data science researcher nowadays oftentimes:

- Communicates with other researchers using Google Mail and WhatsApp

- Searches for research paper using Google Scholoar

- Runs Tensorflow and PyTorch for conducting his machine learning research

- Employs pretrained data (e.g., Bert Word Embeddings) provided by Google

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech
2020

Curse or Blessing? Domination of Google and Facebook in Data Science

**Google+Facebook's domination regarding
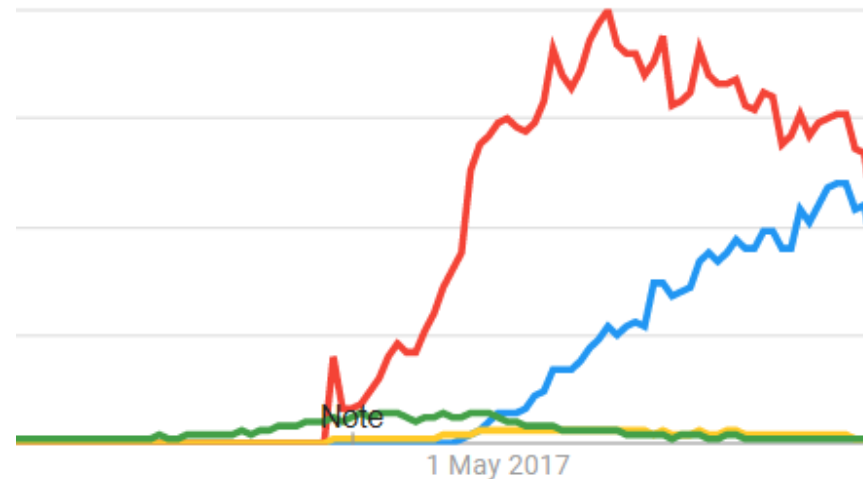neural network frameworks**

Diagram from Google Trends. Google Trends analyzes the popularity of certain search terms over time.

- Red: Tensorflow (Google)

- Blue: PyTorch (Facebook)

- Yellow: MxNet (Apache)

- Green: Theano (Montreal Institute for Learning Algorithms (MILA))

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

Curse or Blessing? Domination of Google and Facebook in Data Science

	August 2015	August 2020
Theano	100%	1%
MxNet	0% (not yet released)	2%
Tensorflow	0% (not yet released)	53%
PyTorch	0% (not yet released)	44%

Table above displays relative frequency of web searches for 4 neural network frameworks (occurrences/total searches for any of the 4 frameworks in %) in August 2015 and August 2020.

Note: Due to ambiguities, searches for Cafe and Torch (variant of PyTorch) give misleading results and therefore these framework are excluded from this table.

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

IARIA

Curse or Blessing? Domination of Google and Facebook in Data Science

Pro: Strong Dominance of Facebook and Google in Data Science Research

- They provide their software and data usually for free

- They conduct important research for advancing this field

Contra: Strong Dominance of Google and Facebook in Data Science Research

- Researchers have to rely on companies providing services, software and data for their research, where those companies might not be trustworthy

- Unclear whether Google and Facebook spy on their users

- Questionable whether Google and Facebook treat their competitors fairly regarding
    - search engine query results
    - provided data

- Academic researchers cannot keep up with financial and computing resources of Google and Facebook

- Google and Facebook can influence the direction of research

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

IARIA

Curse or Blessing? Domination of Google and Facebook in Data Science

Possible counter-measures:

- leave status quo unchanged

- tighter governmental control of Google and Facebook

- government agencies provide high computing resources to the research community

- government agencies provide web search services to the research community

- increased funding of non-profit organizations like Apache Foundation or universities

Panel:
Information, Data, and Services in Digital Society
(Citizen Science, Personal Data, Health Services, Personalized Services, Semantic Processing)

NexTech 2020

Understanding the Meaning of Words Using Dictionary

Minoru Sasaki

Ibaraki University, Japan

minoru.sasaki.01@vc.ibaraki.ac.jp

Find out the meaning of the words

- How do we behave in the digital society in order to know the meanings of words?
  - Wikipedia
  - English Dictionary, Translation Dictionary (e.g. English-Japanese)
  - WordNet
- Recently, Neural word sense disambiguation systems make use of external resources (e.g. dictionary, thesaurus)
- What is a useful dictionary for both humans and computers to find the meaning?
  - We need to further enrich the content of the existing dictionaries.

Example Sentence Extensions for Words

- Word Sense Disambiguation (WSD)
  - Identify which sense of a target polysemous word is used in an example.
  - Example sentences assigned with word senses can be added to the dictionary.
- WSD is used in many NLP tasks.
 - machine translation, question answering, information extraction, etc.

Sense Definition Expansion

- Sense Definition or glosses has been shown to be a valuable resource for improving WSD.
  - GlossBERT(Huang, 2019), EWISE(Kumar, 2019), EWISER(Bevilacqua, 2020) etc.
- GlossBERT using a Japanese dictionary
  - Average precision of WSD is about 30%
  - Japanese sense definition is not effective for WSD based on the knowledge base.

- I want to show the effectiveness of sense definition sentences that expand on the relevant information.

Identification of Phrases with One Meaning

- Using phrases or idioms in the dictionary for WSD
  - I want to use the knowledge of common phrases to identify the meaning.

- "He is just pulling your leg."
  - "He is just joking."
    - The meaning of the sentence can be captured by using knowledge of phrases.
  - "He is just taking your foot."
    - If you only consider the meaning of words, it may not be possible to understand the meaning correctly.
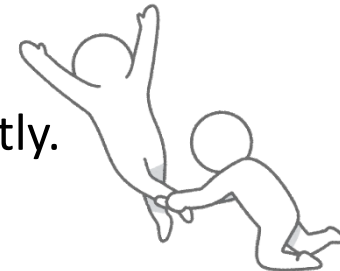
Conclusion

- Dictionaries will be more useful "for people to understand the meaning of words" and "for WSD systems to improve the performance".
  - Humans and computers will be able to capture the characteristics of word meanings and differences between them.
  - The improved dictionary will be useful for downstream tasks such as machine translation and information Extraction.