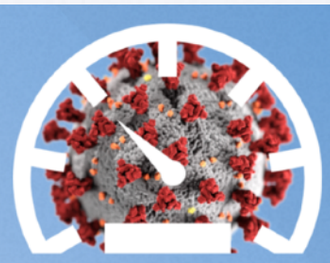
Università degli Studi di Padova

Overall Introduction to Covid-19 MLIA @ Eval

Khalid Choukri
ELRA, France

Nicola Ferro
University of Padua, Italy



http://eval.covid19-mlia.eu/
@covid19mlia

Covid-19 MLIA Eval

Round 1 Virtual Meeting
12-14 January 2021, Online

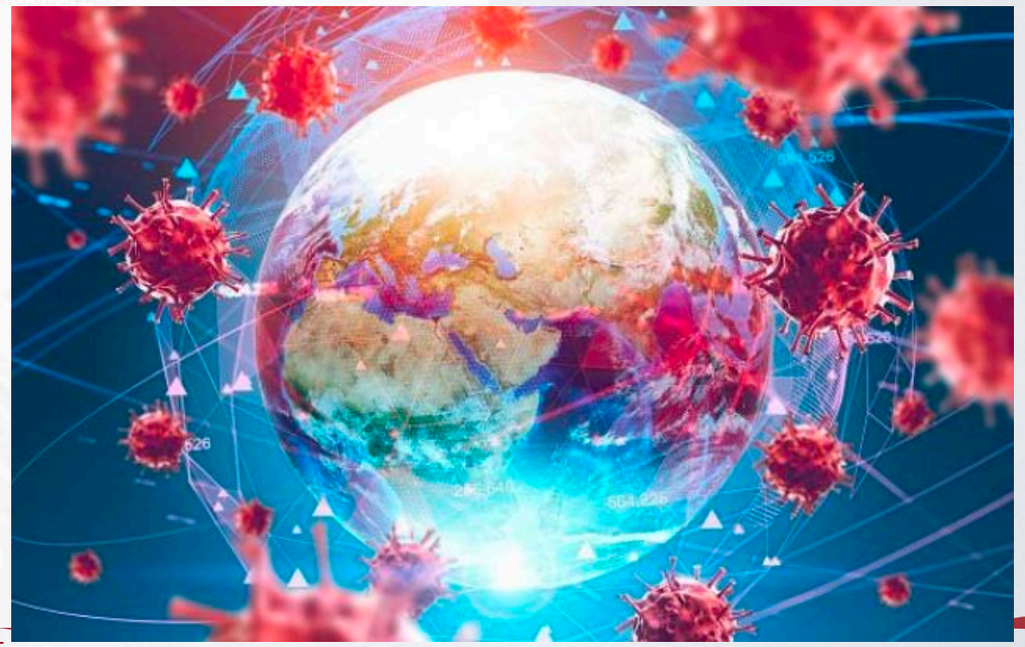
Global Pandemics & Information Access

Khalid Choukri, Nicola Ferro
@covid19mlia

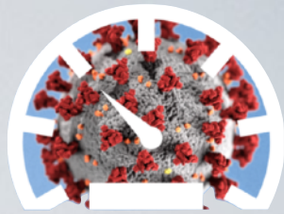
Global Pandemics & Information Access

> “GETTING INFORMATION OFF THE INTERNET IS LIKE TAKING A DRINK FROM A FIRE HYDRANT

Mitch Kapor



We already have search engines, right?

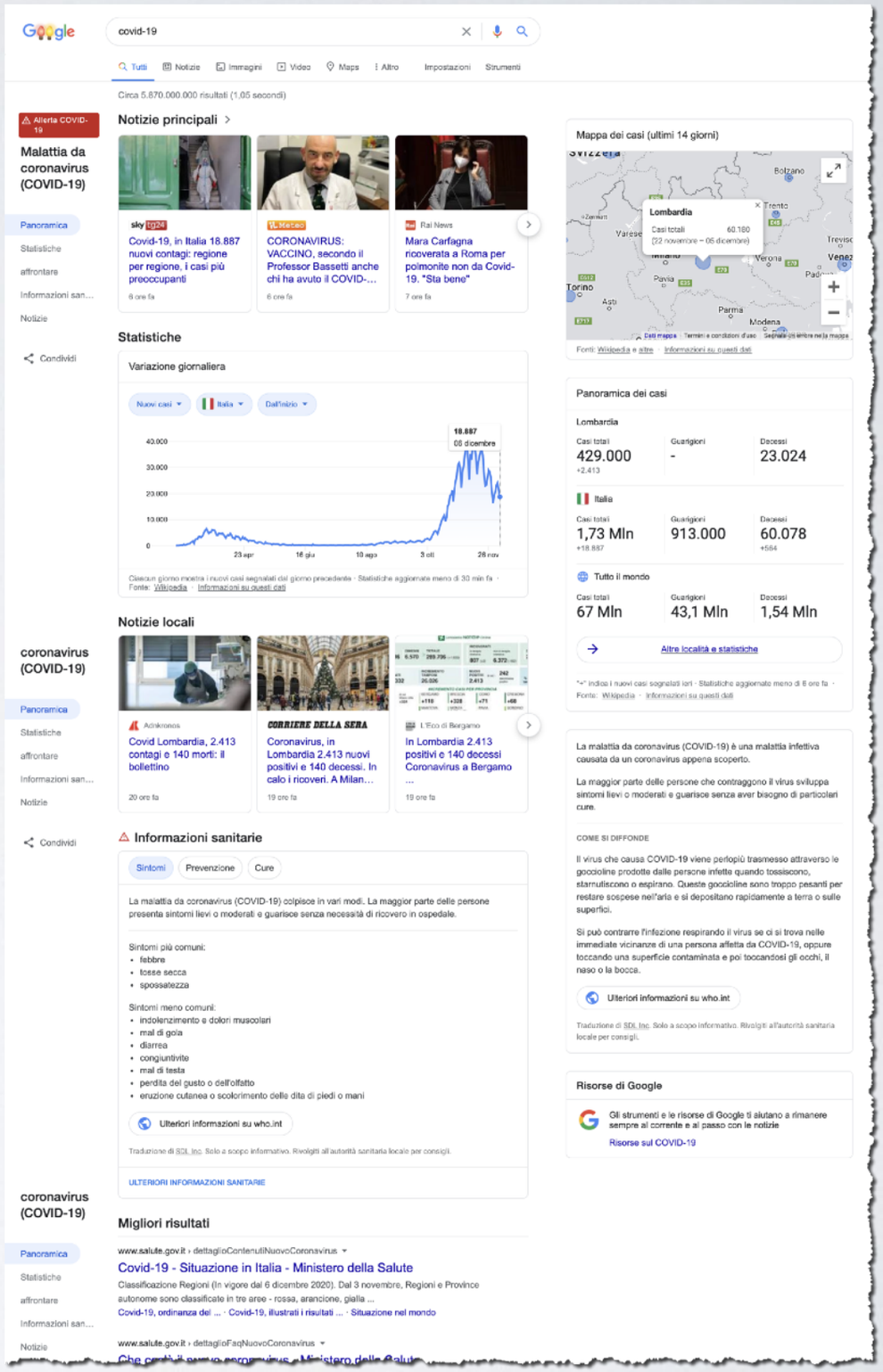


Google

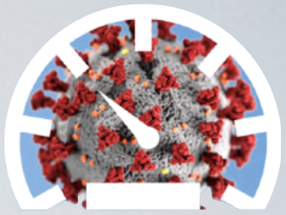
Bing

Яндекс

Baidu 百度



Well, yes… but…



HARVARD KENNEDY SCHOOL | SHORENSTEIN CENTER ON MEDIA, POLITICS, AND PUBLIC POLICY CONTACT

HARVARD KENNEDY SCHOOL
Misinformation Review

EXPLORE *the review* SUBMIT *an essay* REVIEW *for Misinfo* ABOUT *the journal*

MAY 11, 2020 SHARE DOWNLOAD PDF

PEER REVIEWED

How search engines disseminate information about COVID-19 and why they should do better

Access to accurate and up-to-date information is essential for individual and collective decision making, especially at times of emergency. On February 26, 2020, two weeks before the World Health Organization (WHO) officially declared the COVID-19's emergency a "pandemic," we systematically collected and analyzed search results for the term "coronavirus" in three languages from six search engines. We found that different search engines prioritize specific categories of information sources, such as government-related websites or alternative media. We also observed that source ranking within the same search engine is subjected to randomization, which can result in unequal access to information among users.

BY MYKOLA MAKHORTYKH
University of Bern, Institute of Communication and Media Studies

ALEKSANDRA URMAN
University of Bern, Institute of Communication and Media Studies

ROBERTO ULLOA
GESIS – Leibniz Institute for the Social Sciences

https://misinforeview.hks.harvard.edu/article/how-search-engines-disseminate-information-about-covid-19-and-why-they-should-do-better/

Well, yes… but…

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

RUTGERS TODAY Explore Topics

Research & Innovation

Online Autocompletes Are More Likely to Yield COVID-19 Misinformation in Spanish than in English

Rutgers researchers find Spanish autocompletes are more likely to yield harmful, negative results

By

Megan Schumann

Date

June 16, 2020

Media Contact

Megan Schumann

848-445-1907

megan.schumann@rutgers.edu

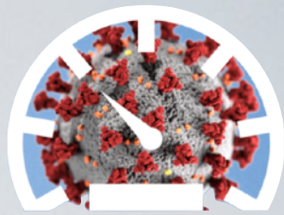
Share

Online autocomplete results for COVID-19 related information are more likely to yield misleading results if the user types in Spanish than in English, according to a new Rutgers report.

The difference may harm Spanish speakers by connecting them with misinformation about handwashing, sanitizers, masks or the disease itself, according to lead author Vivek Singh, an assistant professor at Rutgers-New Brunswick's School of Communication and Information.

https://www.rutgers.edu/news/online-autocompletes-are-more-likely-yield-covid-19-misinformation-spanish-english

Well, yes… but…



MIT Technology Review

Sign in Subscribe

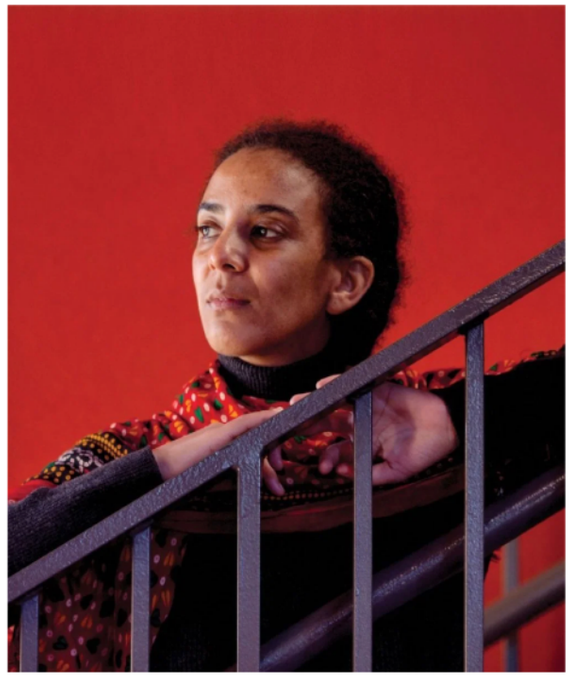
Topics Magazine Newsletters Events

Artificial intelligence / Machine learning

We read the paper that forced Timnit Gebru out of Google. Here's what it says.

The company's star ethics researcher highlighted the risks of large language models, which are key to Google's business.

by **Karen Hao** December 4, 2020



COURTESY OF TIMNIT GEBRU

On the evening of Wednesday, December 2, Timnit Gebru, the co-lead of Google's ethical AI team, announced via Twitter that the company had forced her out.

Gebru, a widely respected leader in AI ethics research, is known for coauthoring a groundbreaking paper that showed facial recognition to be less accurate at identifying women and people of color, which means its use can end up discriminating against them. She also cofounded the Black in AI affinity group, and champions diversity in the tech industry. The team she helped build at Google is one of the most diverse in AI and includes many leading experts in their own right. Peers in the field envied it for producing critical work that often challenged mainstream AI practices.

Popular

01. The coming war on the hidden algorithms that trap people in poverty

02. Logging in to get kicked out: Inside America's virtual eviction crisis

03. This is the most precise 3D map of the Milky Way ever made

MIT Technology Review

Topics Magazine Newsletters Events Sign in Subscribe

Massive data, inscrutable models

Large language models are also trained on exponentially increasing amounts of text. This means researchers have sought to collect all the data they can from the internet, so there's a risk that racist, sexist, and otherwise abusive language ends up in the training data.

An AI model taught to view racist language as normal is obviously bad. The researchers, though, point out a couple of more subtle problems. One is that shifts in language play an important role in social change; the MeToo and Black Lives Matter movements, for example, have tried to establish a new anti-sexist and anti-racist vocabulary. An AI model trained on vast swaths of the internet won't be attuned to the nuances of this vocabulary and won't produce or interpret language in line with these new cultural norms.

It will also fail to capture the language and the norms of countries and peoples that have less access to the internet and thus a smaller linguistic footprint online. The result is that AI-generated language will be homogenized, reflecting the practices of the richest countries and communities.

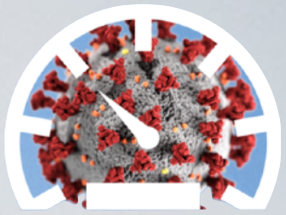
Moreover, because the training data sets are so large, it's hard to audit them to check for these embedded biases. "A methodology that relies on datasets too large to document is therefore inherently risky," the researchers conclude. "While documentation allows for potential accountability, […] undocumented training data perpetuates harm without recourse."

Research opportunity costs

The researchers summarize the third challenge as the risk of "misdirected research effort." Though most AI researchers acknowledge that large language models don't actually *understand* language and are merely excellent at *manipulating* it, Big Tech can make money from models that

https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

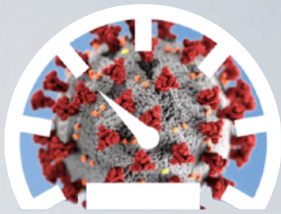
What do we need?



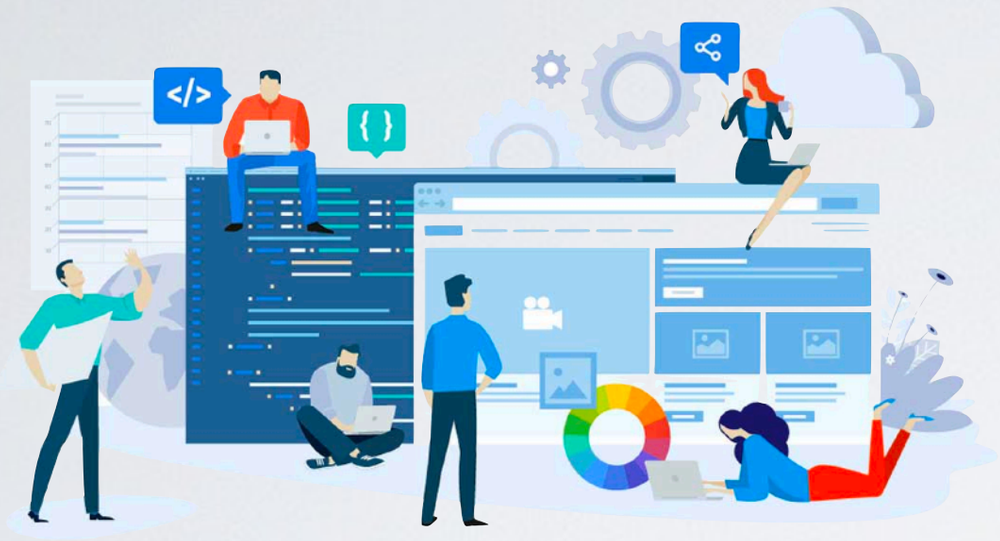
Systems

Resources

What do we need?



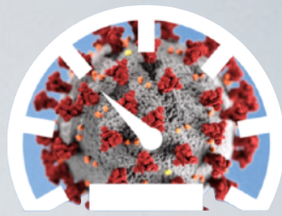
Systems



Resources

Evaluation

What do others do? CORD-19



AI2 Allen Institute for AI

Research

CORD-19: COVID-19 Open Research Dataset

Semantic Scholar • 2020

CORD-19 is a free resource of tens of thousands of scholarly articles about COVID-19, SARS-CoV-2, and related coronaviruses for use by the global research community.

Download | Read Paper | View Website

License: Semantic Scholar Dataset License

AI2

AI FOR THE COMMON GOOD

Email us: ai2-info@allenai.org

Call us: 206.548.5600

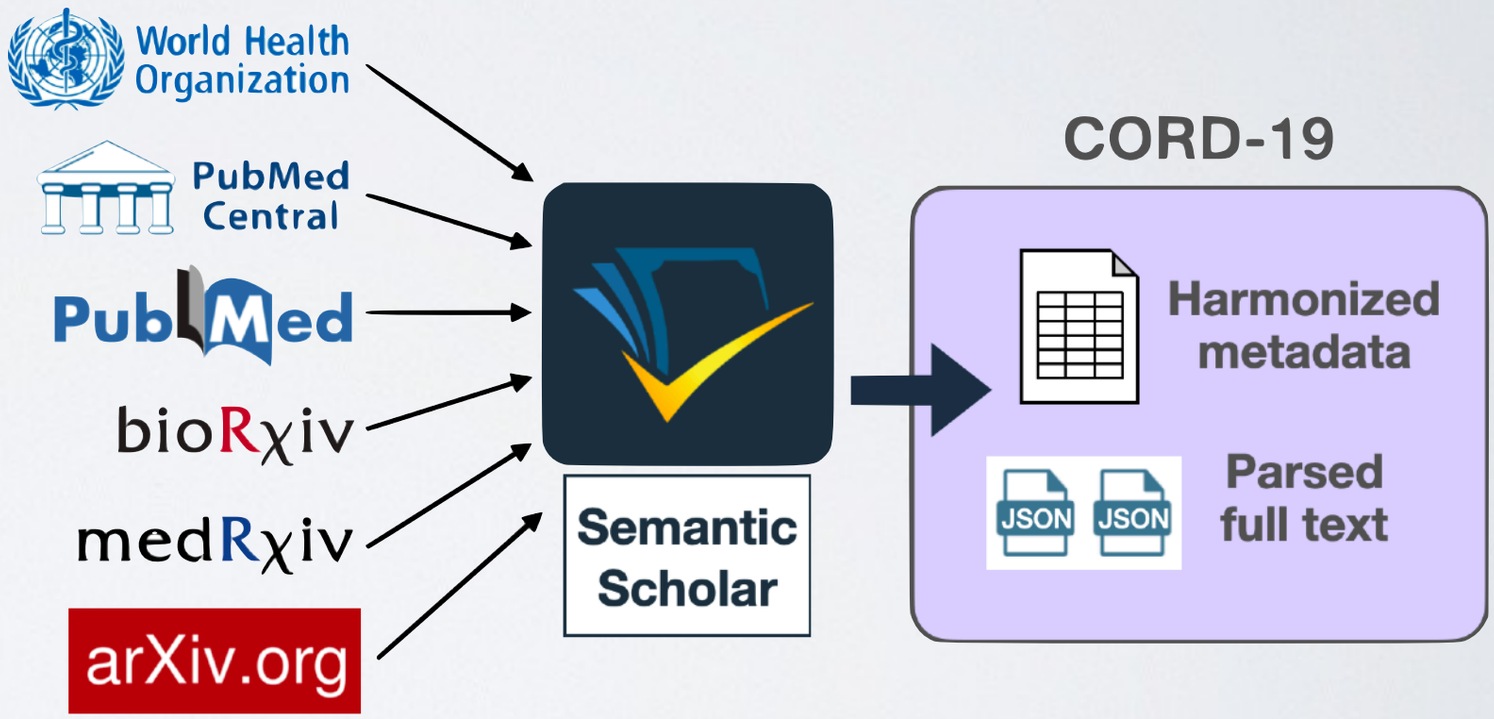
Follow us: @allen_ai

Subscribe to the AI2 Newsletter

Privacy Policy | Terms and Conditions

© The Allen Institute for Artificial Intelligence - All Rights Reserved.

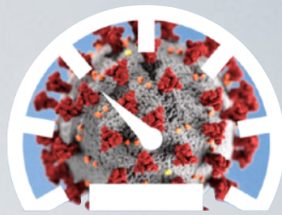
https://allenai.org/data/cord-19



Wang, L. L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Burdick, D., Eide, D., Funk, K., Katsis, Y., Kinney, R., Li, Y., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z., Stilson, B., Wade, A., Wang, K., Ru Wang, N. X., Wilhelm, C., Xie, B., Raymond, D., Weld, D. S., Etzioni, O., and Kohlmeier, S. (2020). CORD-19: The COVID-19 Open Research Dataset. *arXiv.org, Digital Libraries (cs.DL)*, arXiv:2004.10706.

https://arxiv.org/abs/2004.10706v4

What do others do? Google on CORD-19



Google AI Blog

The latest news from Google AI

An NLU-Powered Tool to Explore COVID-19 Scientific Literature

Monday, May 4, 2020

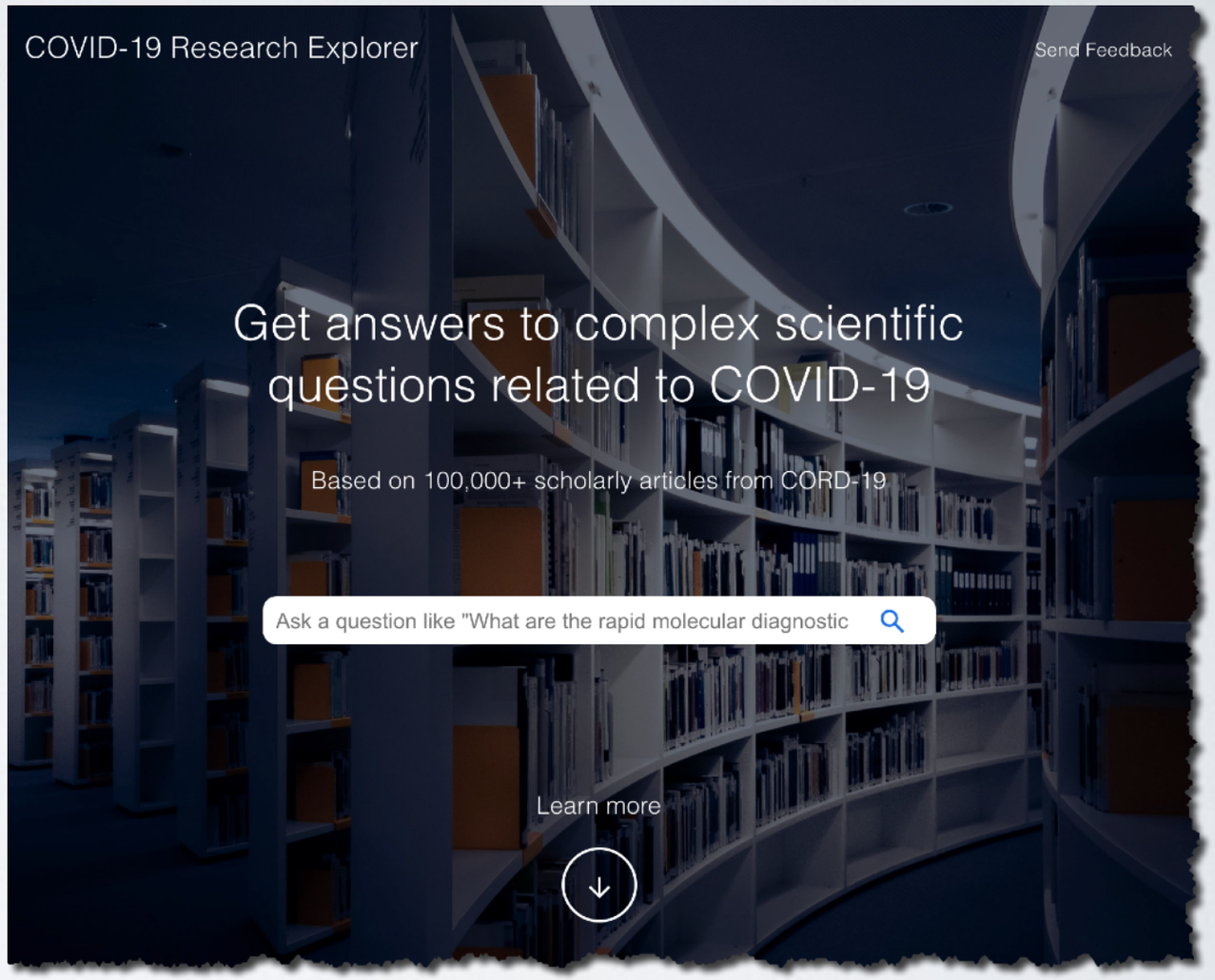
Posted by Keith Hall, Research Scientist, Natural Language Understanding, Google Research

Due to the COVID-19 pandemic, scientists and researchers around the world are publishing an immense amount of new research in order to understand and combat the disease. While the volume of research is very encouraging, it can be difficult for scientists and researchers to keep up with the rapid pace of new publications. Traditional search engines can be excellent resources for finding real-time information on general COVID-19 questions like "How many COVID-19 cases are there in the United States?", but can struggle with understanding the meaning behind *research-driven* queries. Furthermore, searching through the existing corpus of COVID-19 scientific literature with traditional keyword-based approaches can make it difficult to pinpoint relevant evidence for complex queries.

To help address this problem, we are launching the COVID-19 Research Explorer, a semantic search interface on top of the COVID-19 Open Research Dataset (CORD-19), which includes more than 50,000 journal articles and preprints. We have designed the tool with the goal of helping scientists and researchers efficiently pore through articles for answers or evidence to COVID-19-related questions.

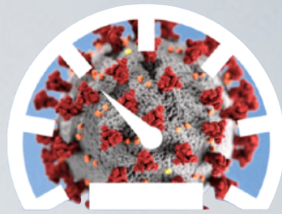
When the user asks an initial question, the tool not only returns a set of papers (like in a traditional search) but also highlights snippets from the paper that are possible answers to

https://covid19-research-explorer.appspot.com/



https://ai.googleblog.com/2020/05/an-nlu-powered-tool-to-explore-covid-19.html

What do others do? TREC-COVID



‹ NIST

TREC-COVID: Building a Pandemic Retrieval Test Collection

About Guidelines Data Tools Archive Bibliography

TREC-COVID

Researchers, clinicians, and policy makers involved with the response to COVID-19 are constantly searching for reliable information on the virus and its impact. This presented a unique opportunity for the information retrieval (IR) and text processing communities to contribute to the response to this pandemic, as well as to study methods for quickly standing up information systems for similar future events. The results of the TREC-COVID Challenge identify answers for some of today's questions and create infrastructure to improve tomorrow's search systems.

TREC-COVID followed the TREC model for building IR test collections through community evaluations of search systems. The document set used in the challenge is the COVID-19 Open Research Dataset (CORD-19). This is a collection of biomedical literature articles that is updated regularly. Accordingly, TREC-COVID consisted of a series of rounds, with each round using a later version of the document set and a larger set of COVID-related topics. Participants in a round created ranked lists of documents for each topic ("runs") and submitted their runs to NIST. Based on the collective set of participants' runs, NIST created sets of documents to be assessed for relevance by human annotators with biomedical expertise. The results of the human annotation, known as relevance judgments, were then used to score the submitted runs.

The final document and topic sets together with the cumulative relevance judgments comprise a COVID test collection called TREC-COVID Complete. The incremental nature of the collection as viewed through the successive rounds supports research on search systems for dynamic environments.

Learn more »

Organizers

The TREC-COVID Challenge was organized by the Allen Institute for Artificial Intelligence (AI2), the National Institute of Standards and Technology (NIST), the National Library of Medicine (NLM), Oregon Health and Science University (OHSU), and the University of Texas Health Science Center at Houston (UTHealth).
NIST press release.

Community Participation

TREC-COVID had strong international participation.

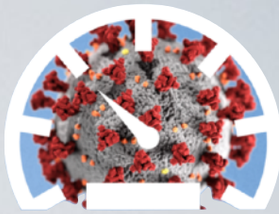
The cumulative test collection for ad hoc retrieval called TREC-COVID Complete is now available for download from the Data page. Topic sets and relevance judgments from previous rounds are also available from the data page. Retrieval results from prior rounds are stored in the open archive of submissions; and a bibliography of papers resulting from TREC-COVID is being maintained.

Next Steps

You can join the trec-covid Google group to discuss the challenge, follow #COVIDSearch on Twitter, or contact the TREC group at NIST for more information. See also the companion COVIDSearch page.

https://ir.nist.gov/covidSubmit/

What do others do? Summing Up



One language

Scientific Literature
Medical Professionals

One task

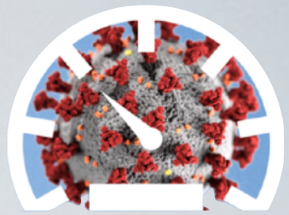
What do we aim at?

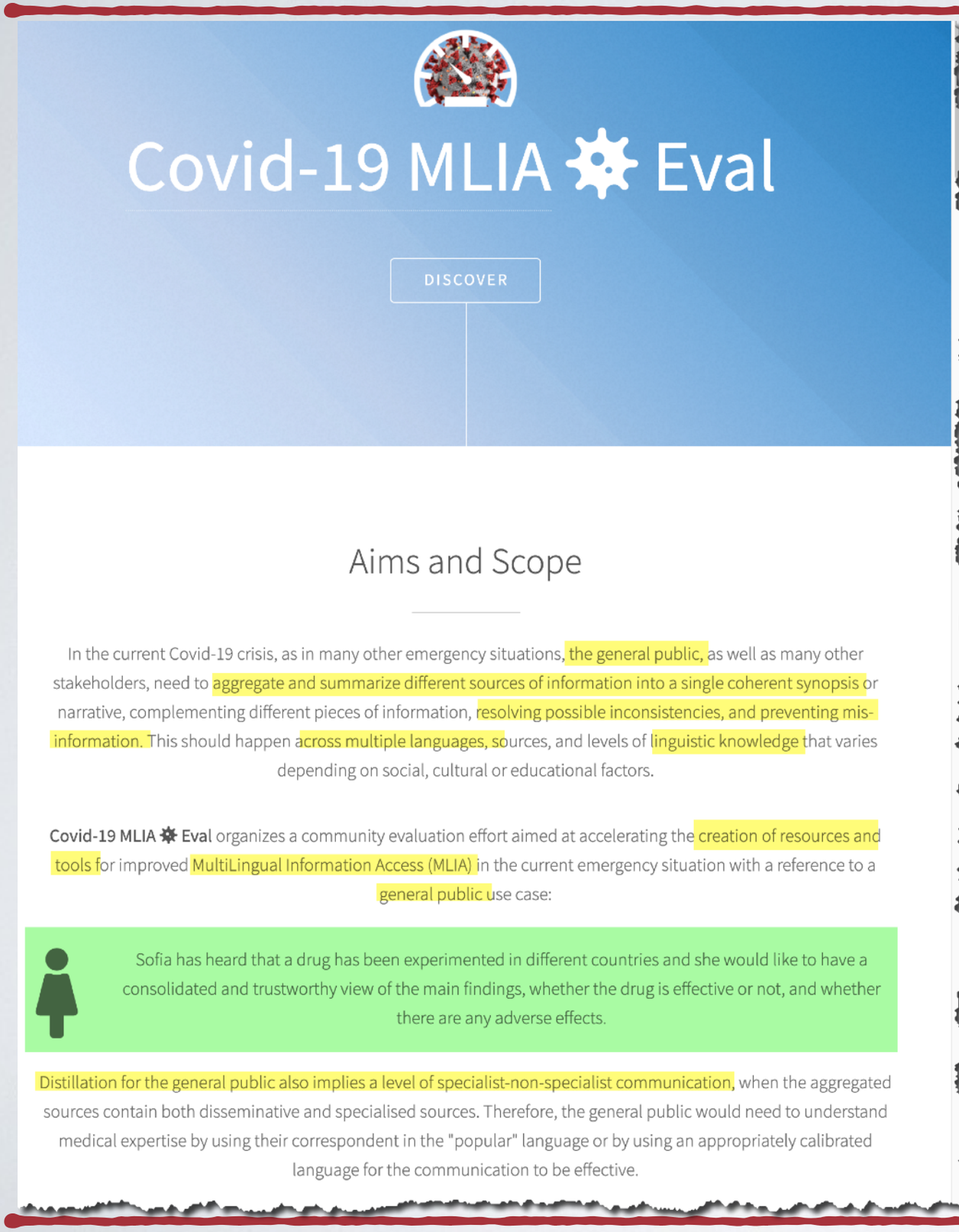
Multiples languages

General Public

Three tasks

A community-based voluntary evaluation effort





European Commission

European Language Resource Coordination

CLEF INITIATIVE

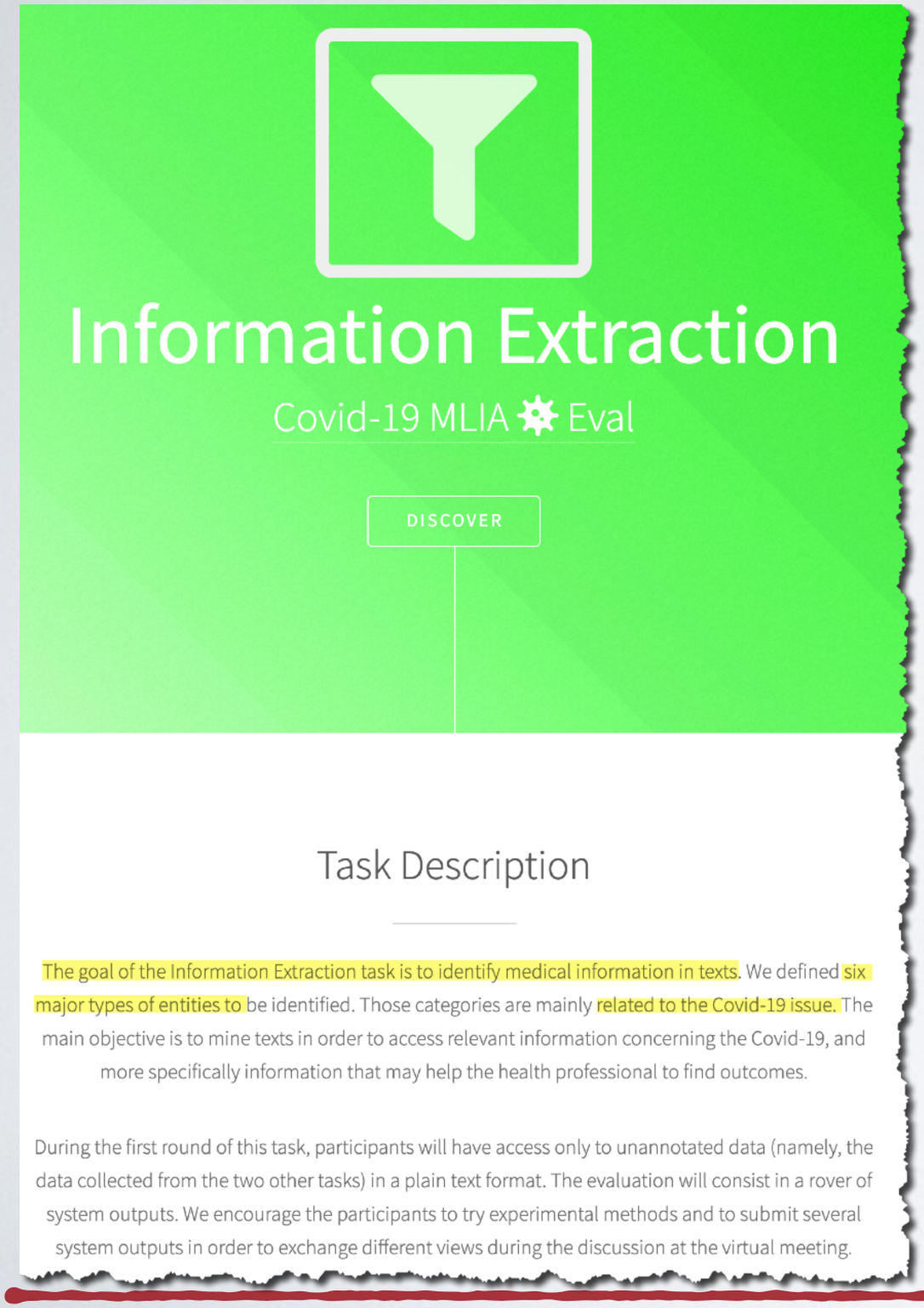
EUROPEAN LANGUAGE RESOURCES ASSOCIATION ELRA

CLARIN
Common Language Resources and Technology Infrastructure

http://eval.covid19-mlia.eu/

Information Extraction

http://eval.covid19-mlia.eu/task1/



- Categories
  - **drug names, treatments, general intervention**: this category concerns both commercial and generic names of drugs, as well as general intervention in the health domain; elements from this category usually come from advices from a professional (medical doctor, pharmacist) or from self-medication
  - **signs, symptoms, diseases**: this category deals with medical problems and merges together all signs, symptoms, and diseases shortness of breath, extreme fatigue, fever, skin infection, weight loss
  - **findings, efficacy of treatments**: this category is more complex since it concerns all elements related to positive or negative effects of treatments, including non expected stuff
  - **tests**: this category concerns all tests performed to diagnose medical problems such as blood sample, physical exam, serological test
  - **behaviors, everyday life actions**: this category concerns all actions performed by each of us such as to wash one's hands, to cough into his elbow, to self-confine, use of face masks, physical distancing
  - **legal dispositions, regulations**: this category concerns all actions decided by local or national authorities (Government, Ministry, etc.), such as to download the employer certificate, list of authorized move, prolonged border closure, closure of educational institutions
- Languages
  - English, French, German, Greek, Italian, Spanish, Swedish

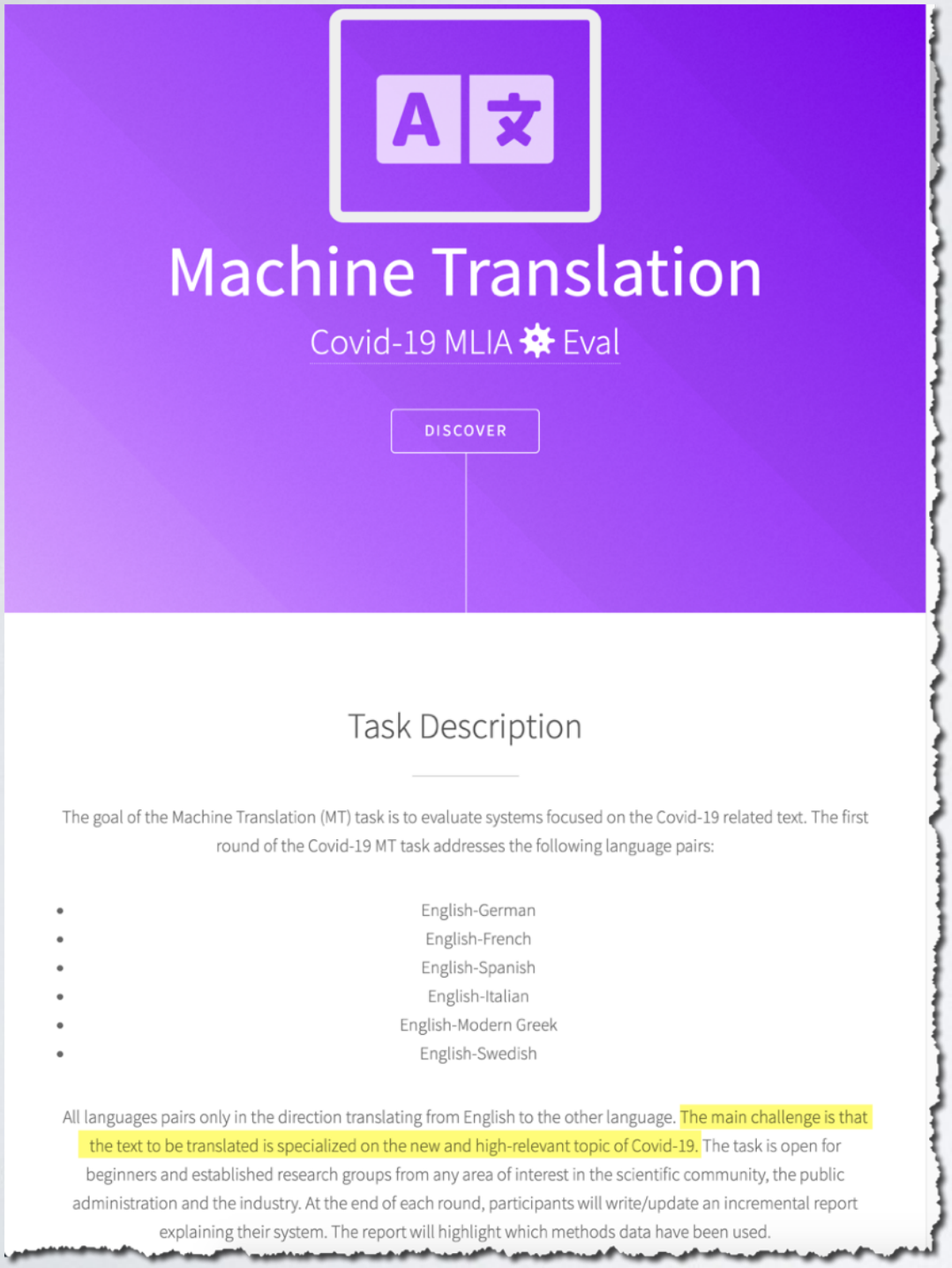
Multilingual Semantic Search

http://eval.covid19-mlia.eu/task2/

- **High Precision**: participants are required to build systems that will help the general public to retrieve the most relevant documents on the Web concerning COVID-19 efficiently. The main focus of this subtask is on the top ranked documents
- **High recall**: the focus is more on the problem of finding as many relevant documents as possible with the least effort. Given a limited amount of resources, such as a time limit and expert availability in time of crisis, there will be a limit on the maximum number of documents that can be retrieved in order to build a set of relevant documents that should be delivered to the general public.
- Both subtasks are open to **monolingual** and **bilingual** submissions
- Languages
  - English, French, German, Greek, Italian, Spanish, Swedish, and Ukrainian
- Topics
  - English, French, German, Greek, Italian, Spanish, Swedish, and Ukrainian plus Chinese and Japanese

Machine Translation

http://eval.covid19-mlia.eu/task3/



- **Constrained**: participants must submit at least a system trained only with the provided data (constrained) for each of the language pairs they would like to participate

  Nasic linguistic tools such as taggers, parsers, or morphological analyzers or multilingual systems are allowed in the constrained condition

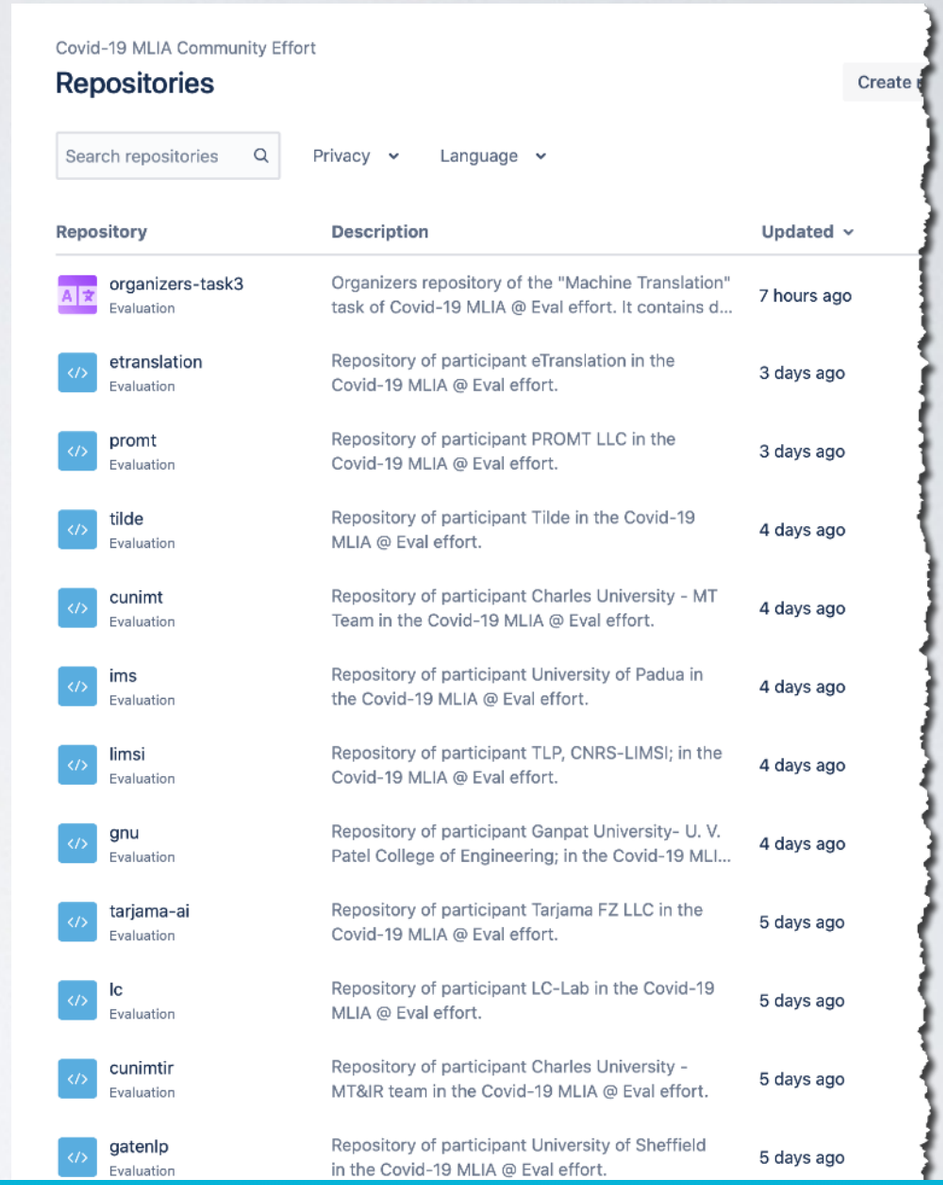
- **Unconstrained**: participants can use additional training data (not provided by the organisers) or existing translation systems specifying a flag that the system uses additional data

- Languages
  - English-German
  - English-French
  - English-Spanish
  - English-Italian
  - English-Modern Greek
  - English-Swedish

Rounds, Repositories, Reports and Meetings

https://bitbucket.org/covid19-mlia/

- We plan for **three rounds**, tentatively one-month long each, organized as follows:
  - **training (2 weeks)**: data will be released and you will develop your own systems;
  - **testing (2 weeks)**: you will submit your system runs, ground-truth will be created, and your runs will be scored;
  - **meeting (1 day)**: an interactive (remote) meeting will be organized where you will shortly present the highlights and downlights to accelerate knowledge transfer and take up for the next round.
- **Rolling technical report**: **participants** and **organizers** will keep and update a rolling technical report the techniques applied and insights gained during participation, round after round
- Participants are provided with a dedicated git **repository** where to push and share the outcomes of your participation in the different rounds, i.e. **runs**, **code**, (language) **resources**, and a **technical report**



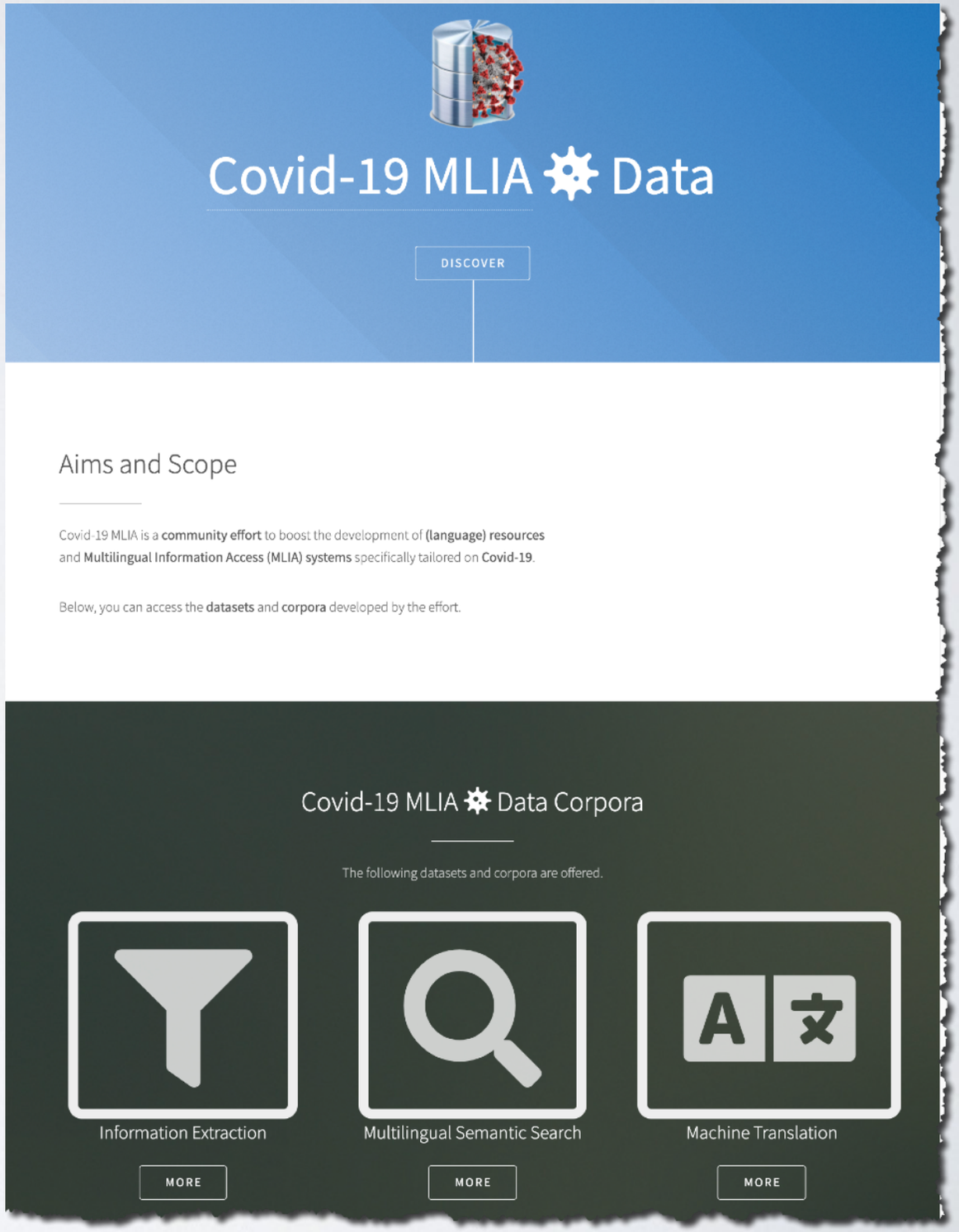
Attribution-ShareAlike 4.0 International
(CC BY-SA 4.0)

Packaging and Exit Strategy

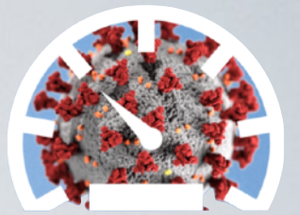
- The git repositories (previous slide) and the data site (see right) constitute our **incremental packaging** approach and our **sustainability and exit strategy** to allow other researchers and practitioners to access, re-use, and exploit everything produced during the initiative, also after its end

- These incremental packages comprise:
  - all train/dev/test sets for each task
  - all the runs submitted by participants
  - all the measures, performance scores and analyses about submitted runs
  - the reports describing these results
  - any other resources and source code shared by participants

http://data.covid19-mlia.eu/

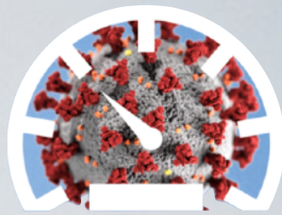


Where We are?



- **50 teams from 26 countries registered** to date, participating in multiple tasks
  - 35 for the Information Extraction task
  - 25 for the Multilingual Semantic Search task
  - 25 for the Machine Translation task
- We completed round 1
  - participants just submitted their runs
    2 December 2020
  - relevance assessments just started
    due by mid December 2020
  - preliminary participants reports
    due by 23 December 2020
  - virtual meeting [NOW]
    12-14 January 2021
- **14 teams from 10 countries actually submitted** runs
  - 4 for the Information Extraction task
    - English 4; German 1; Greek 1; Italian 1; Spanish 1
  - 4 for the Multilingual Semantic Search task
    - English 3; French 2; German 2; Greek 1; Italian 1; Spanish 3; Swedish 1; Ukrainian 1
    - {German, French, Spanish, Swedish} → English 1
    - English → German 1
    - English → Spanish 1
    - English → French 1
  - 8 for the Machine Translation task
    - English → German 5
    - English → French 8
    - English → Spanish 6
    - English → Italian 4
    - English → Greek 2
    - English → Swedish 5

Plans for the Virtual Meeting

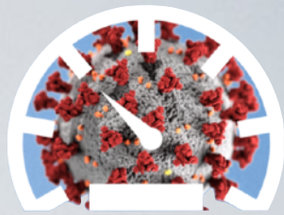


	Tuesday 12 January 2021	
14:00-14:15	Setup and Logistics	
14:15-14:30	Welcome	Philippe Gelin [European Commission, Luxembourg] Alexandru Ceausu [European Commission, Luxembourg] Khalid Choukri [ELRA, France] Nicola Ferro [University of Padua, Italy]
14:30-15:00	Overall Introduction to Covid-19 MLIA ⚙ Eval	Khalid Choukri [ELRA, France] Nicola Ferro [University of Padua, Italy]
15:00-15:20	Data Acquisition	Guillaume Jacquet [EU Joint Research Centre, Italy] Stelios Piperidis [ILSP/Athena RC, Greece]
15:20-15:30	Break	
15:30-16:00	Overview of the "Information Extraction" Task	Thierry Declerck [DFKI, Germany] Cyril Grouin [LIMSI, France] Pierre Zweigenbaum [LIMSI, France]
16:00-16:30	Overview of the "Machine Translation" Task	Francisco Casacuberta [Universitat Politècnica de València, Spain] Miguel Domingo [Universitat Politècnica de València, Spain] Mercedes García-Martínez [Pangeanic, Spain] Manuel Herranz [Pangeanic, Spain]
16:30-17:00	Overview of the "Multilingual Semantic Search" Task	Giorgio Maria Di Nunzio [University of Padua, Italy] Maria Eskevich [CLARIN ERIC]
17:00-17:15	Wrap up and planning for the next day	Khalid Choukri [ELRA, France] Nicola Ferro [University of Padua, Italy]

Plans for the Virtual Meeting

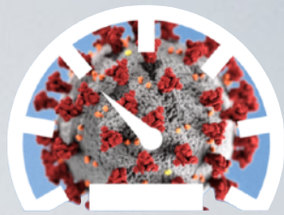
	Wednesday 13 January 2021		
14:20-14:50	University of Padua (Task 2)	Alberto Purpura [University of Padua, Italy]	
14:50-15:20	Charles University (Task 2, Task 3)	Shadi Saleh [Charles University, Czech Republic]	
15:20-15:30	Break		
15:30-16:00	GATENLP, University of Sheffield (Task 2)	Iknoor Singh [University of Sheffield, UK]	
16:00-16:30	Universidad de Jaén (Task 2)	José Alberto Mesa Murgado [Universidad de Jaén, Spain]	
16:30-16:45	Wrap up and planning for the next day	Khalid Choukri [ELRA, France] Nicola Ferro [University of Padua, Italy]	

Plans for the Virtual Meeting



Thursday 14 January 2021			
14:00-14:30	CNRS (Task 3)	François Yvon [CNRS, France]	
14:30-15:00	Charles University (Task 3)	Ivana Kvapilikova [Charles University, Czech Republic]	
15:00-15:30	PROMT(Task 3)	Alexander Molchanov [PROMT, Russia]	
15:30-15:40	Break		
15:40-16:10	SWLab - University of Cagliari (Task 1)	Diego Marcia [University of Cagliari, Italy]	
16:10-16:40	Innoradiant (Task 1)	Luca Dini [Innoradiant, France]	
16:40-17:10	Lingua Custodia (Task 3)	Raheel Qader [Lingua Custodia, France]	

A Joint Effort



Coordinators

Overall

Khalid Choukri, Evaluations and Language resources Distribution Agency (ELDA), France
choukri@elda.org

Nicola Ferro, University of Padua, Italy
ferro@dei.unipd.it

Data Acquisition and Engineering

Miltos Deligiannis, ILSP/Athena RC, Greece
mdel@athenarc.gr

Vassilis Papavassiliou, ILSP/Athena RC, Greece
vpapa@athenarc.gr

Stelios Piperidis, ILSP/Athena RC, Greece
spip@athenarc.gr

Prokopis Prokopidis, ILSP/Athena RC, Greece
prokopis@athenarc.gr

Information Extraction

Thierry Declerck, DFKI, Germany
declerck@dfki.de

Cyril Grouin, LIMSI, France
cyril.grouin@limsi.fr

Pierre Zweigenbaum, LIMSI, France
pz@limsi.fr

Multilingual Semantic Search

Giorgio Maria Di Nunzio, University of Padua, Italy
dinunzio@dei.unipd.it

Maria Eskevich, CLARIN ERIC
maria@clarin.eu

Machine Translation

Francisco Casacuberta, Universitat Politècnica de València, Spain
fcn@prhlt.upv.es

Miguel Domingo, Universitat Politècnica de València, Spain
midobal@prhlt.upv.es

Mercedes García-Martínez, Pangeanic, Spain
m.garcia@pangeanic.com

Manuel Herranz, Pangeanic, Spain
m.herranz@pangeanic.es

Topic translation

Relevance Assessment

JOIN US

http://eval.covid19-mlia.eu/registration/