Open Science and Exploitation

Policy and Practice

Dr. Prodromos Tsiavos

Athena RC/ EPLO IPRI

ptsiavos@athenarc.gr/ p.tsiavos@eplo.int

What is research?

What we think as research

July 5, 2020

`Dataset` `Open Access`

A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration

ⓘ Banda, Juan M.; ⓘ Tekumalla, Ramya; Wang, Guanyu; Yu, Jingyuan; Liu, Tuo; Ding, Yuning; Artemova, Katya; Tutubalina, Elena; ⓘ Chowell, Gerardo

NEW in Version 17: Besides our regular update, we now have included the tweet identifiers and their respective tweet location place country code for the clean version of the dataset. This is found on the clean_place_country.tar.gz file, each file is identified by the two-character ISO country code as the file suffix.

Due to the relevance of the COVID-19 global pandemic, we are releasing our dataset of tweets acquired from the Twitter Stream related to COVID-19 chatter. Since our first release we have received additional data from our new collaborators, allowing this resource to grow to its current size. Dedicated data gathering started from March 11th yielding over 4 million tweets a day. We have added additional data provided by our new collaborators from January 27th to March 27th, to provide extra longitudinal coverage. Version 10 added ~1.5 million tweets in the Russian language collected between January 1st and May 8th, gracefully provided to us by: Katya Artemova (NRU HSE) and Elena Tutubalina (KFU). From version 12 we have included daily hashtags, mentions and emojis and their frequencies the respective zip files. From version 14 *we* have included the tweet identifiers and their respective language for the clean version of the dataset. This is found on the clean_languages.tar.gz file, each file is identified by the two-character language code as the file suffix.

The data collected from the stream captures all languages, but the higher prevalence are: English, Spanish, and French. We release all tweets and retweets on the full_dataset.tsv file (468,169,539 unique tweets), and a cleaned version with no retweets on the full_dataset-clean.tsv file (115,262,201 unique tweets). There are several practical reasons for us to leave the retweets, tracing important tweets and their dissemination is one of them. For NLP tasks we provide the top 1000 frequent terms in frequent_terms.csv, the top 1000 bigrams in frequent_bigrams.csv, and the top 1000 trigrams in frequent_trigrams.csv. Some general statistics per day are included for both datasets in the statistics-full_dataset.tsv and statistics-full_dataset-clean.tsv files. For more statistics and some visualizations visit: http://www.panacealab.org/covid19/

More details can be found (and will be updated faster at: https://github.com/thepanacealab/covid19_twitter) and our

Indexed in

OpenAIRE

Publication date:
July 5, 2020

DOI:
DOI 10.5281/zenodo.3930903

Keyword(s):
`social media` `twitter` `nlp` `covid-19` `covid19`

Related identifiers:
Continued by
http://www.panacealab.org/covid19/ (Other)

Supplement to

Process Behavior Template provided by Mark Graban, author of "Measures of Success"

	Data	Average	Moving Range (MR)	MR Bar	Lower Natural Process Limit	Upper Natural Process Limit	MR Upper Range Limit	Target	Halfway to Lower Limit	Halfway to Upper Limt
Point 1	14.89	15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 2	15.83	15.98	0.94	2.17	10.22	21.75	7.08		13.10	10.87
Point 3	16.32	15.98	0.49	2.17	10.22	21.75	7.08		13.10	10.87
Point 4	18.32	15.98	2.00	2.17	10.22	21.75	7.08		13.10	10.87
Point 5	14.58	15.98	3.74	2.17	10.22	21.75	7.08		13.10	10.87
Point 6	15.90	15.98	1.32	2.17	10.22	21.75	7.08		13.10	10.87
Point 7	18.90	15.98	3.00	2.17	10.22	21.75	7.08		13.10	10.87
Point 8	12.76	15.98	6.14	2.17	10.22	21.75	7.08		13.10	10.87
Point 9	15.50	15.98	2.74	2.17	10.22	21.75	7.08		13.10	10.87
Point 10	15.80	15.98	0.30	2.17	10.22	21.75	7.08		13.10	10.87
Point 11	16.90	15.98	1.10	2.17	10.22	21.75	7.08		13.10	10.87
Point 12	14.30	15.98	2.60	2.17	10.22	21.75	7.08		13.10	10.87
Point 13	16.50	15.98	2.20	2.17	10.22	21.75	7.08		13.10	10.87
Point 14	15.30	15.98	1.20	2.17	10.22	21.75	7.08		13.10	10.87
Point 15	16.30	15.98	1.00	2.17	10.22	21.75	7.08		13.10	10.87
Point 16	18.60	15.98	2.30	2.17	10.22	21.75	7.08		13.10	10.87
Point 17	15.00	15.98	3.60	2.17	10.22	21.75	7.08		13.10	10.87
Point 18		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 19		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 20		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 21		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 22		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 23		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 24		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 25		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 26		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 27		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 28		15.98		2.17	10.22	21.75	7.08		13.10	10.87
Point 29		15.98		2.17	10.22	21.75	7.08		13.10	10.87

X Chart (Individuals Data)

MR Chart
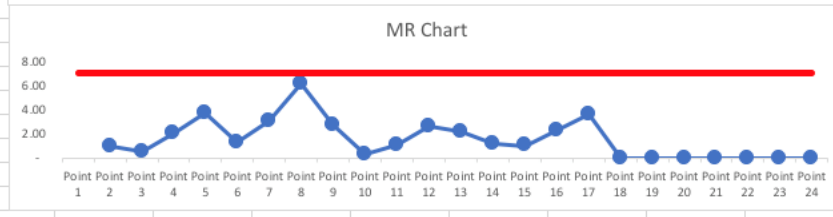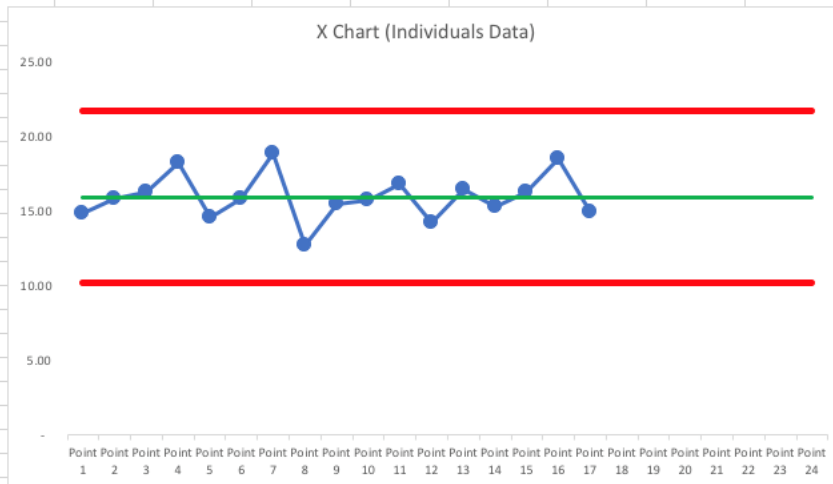
What research actually is

Dramatis Personæ.

Julius Cæsar
Octavius Cæsar
Antony
Brutus
Cassius
Casca
Trebonius
Ligarius
Decius Brutus
Metellus Cimber
Cinna
Antidoteus
Messala
Titinius
Cinna the Poet
Flavius

Conspirators

Plebeians

Women.

Calphurnia
Portia

Guards and Attendants.

Scene ROME.

THE

THE TRAGEDY. OF JULIUS CÆSAR.

Actus Primus. Scæna Prima.

Enter Flavius, Casca, and certain Commoners over the Stage.

Oh, okay. Let me gather some information and see if we can help. What is your first name?

Fred

Fred.

Emily

And your last name; would you spell it for me please? Fred Sure. It's C-H-A-M-B-E-R-S, Chambers.

Emily

Okay. And your company name?

Fred

I'm with GoldStar Environmental.

Emily

GoldStar Environmental?

Fred

Yes ma'am.

Emily

system?

Fred

No, actually, I can't even get an Internet connection.

Emily

Okay. According to my records, the AmStar700 is a voice over Internet protocol phone. It appears that because your Internet isn't working, your phones are not working as well. Do you know who your Internet provider is?

Fred

I have Verizon.

Emily

Okay. Mr. Chambers, I'm going to get a hold of Randy, and have him return your call. Is the 610-265-1715 a good number to reach you at right now?

Fred

Yeah that's my cell. That is working.

Ο παρών ιστότοπος χρησιμοποιεί για τη λειτουργία του cookies. Ενημερωθείτε για την πολιτική χρήσης τους **εδώ**. ✔ Ενημερώθηκα

Δι@ύγεια
διαφάνεια στο κράτος

Υπουργείο Ψηφιακής Διακυβέρνησης
ΠΡΟΓΡΑΜΜΑ ΔΙΑΥΓΕΙΑ

👤 Είσοδος

Εισάγετε κείμενο για αναζήτηση αποφάσεων... 🔍

A-- A A+ A++

Σύνθετη αναζήτηση... 🛈 Πληροφορίες αναζήτησης...

🏢 ΠΡΟΕΔΡΙΑ ΤΗΣ ΚΥΒΕΡΝΗΣΗΣ ⚙▾

Συνολικός αριθμός πράξεων: 27

XML JSON XLS RSS

🔲 Κατηγοριοποίηση πράξεων φορέα 🛈

« ‹ 1 2 3 › »

Υπογράφοντες | Μονάδες | Τύποι Πράξεων | Θεματικές

ΑΔΑ: ΩΗΜ346ΜΓΨ7-Ρ0Ο - ΑΠΟΣΠΑΣΗ ΥΠΑΛΛΗΛΟΥ ΣΤΗΝ ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΣΥΝΤΟΝΙΣΜΟΥ ΕΣΩΤΕΡΙΚΩΝ ΠΟΛΙΤΙΚΩΝ

- ΓΕΡΑΠΕΤΡΙΤΗΣ ΓΕΩΡΓΙΟΣ (4)
- ΔΗΜΗΤΡΙΑΔΗΣ ΓΡΗΓΟΡΙΟΣ (6)
- ΜΗΤΣΟΤΑΚΗΣ ΚΥΡΙΑΚΟΣ (17)

📄 Λήψη αρχείου | 📄 Προβολή αρχείου

Ημ/νία τελευταίας τροποποίησης:
24/10/2019 12:04:47

Είδος: ΛΟΙΠΕΣ ΑΤΟΜΙΚΕΣ ΔΙΟΙΚΗΤΙΚΕΣ ΠΡΑΞΕΙΣ

ΑΔΑ: 63Ν746ΜΓΨ7-ΔΟΙ - ΑΠΟΣΠΑΣΗ ΥΠΑΛΛΗΛΟΥ ΣΤΗΝ ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΣΥΝΤΟΝΙΣΜΟΥ ΕΣΩΤΕΡΙΚΩΝ ΠΟΛΙΤΙΚΩΝ

📄 Λήψη αρχείου | 📄 Προβολή αρχείου

Ημ/νία τελευταίας τροποποίησης: **Είδος:** ΛΟΙΠΕΣ ΑΤΟΜΙΚΕΣ ΔΙΟΙΚΗΤΙΚΕΣ

European Language Resource Coordination
Connecting Europe Facility
ELRC-SHARE

Home   Browse Resources   Help   About   Register   Login

2015 Calls for Tenders for Translation 🗄

Attribution details: Utrecht University

Contains monolingual Netherlands Dutch texts with the 2015 calls for tenders for translation work for the child welfare office, the office for prisoner rehabilitation and for the ministry of Economic Affairs

← **Back**   ⬇ **Download**

text 📄

Distribution

Availability: Available

Licences

Open Under-PSI

Used for resources that fall under the scope of PSI (Public Sector

Monolingual text corpus

Languages

Dutch; Flemish (nl)

Linguality

Linguality type: Monolingual

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

IP Services   Policy   Cooperation   Resources   About IP   About WIPO

Search WIPO

Identifying Inventions in the Public Domain

A Guide for Inventors and Entrepreneurs

Author(s): | Publication year: 2020

License: CC BY

Download

English PDF
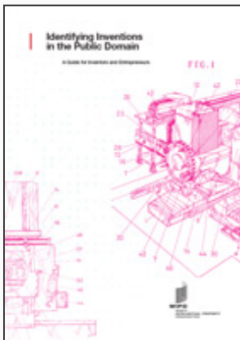
This guide aims to assist researchers, inventors and entrepreneurs in determining whether specific inventions are protected by enforceable patents or may be in the public domain, by teaching a three-stage process for searching and analyzing published patent documents using the tools of freedom to operate determination. Designed for self-study, the guide has easy-to-follow training modules that take the reader through the process step by step, including with the help of useful checklists and other tools.

Research:

- Does not happen only in the lab or the university
- Is heavily digitally mediated
- It involves multiple types of material and works
- It includes all type of Intellectual Property Rights
- It involves multiple types of actors under different regulatory regimes
- It should not be approached at the end of the research process, but rather from its outset

2 Regulating flows of information

Life-cycle essentials

- **Initiating** the research:
    - Do I have the property rights?
    - Do I have a licence?
    - Are there special regimes I have to respect?
    - Is there any funding?
    - Are there any contractual relationships?
- **Conducting** the research
    - Arrangements with my RPO/ RFO
    - Codes of Conduct
    - Confidentiality Regimes
    - Access controls
    - Consortia and partnerships
    - IPR (patents/ trademarks) registration arrangements and ownership agreements
- **Sharing** the research
    - Inter-organizational sharing
    - Opening up
    - Limited licensing
    - Exploitation (patents/ commercial licensing)

Research

- May include public domain
- May include copyrighted Works
- May lead to a patent
- May be based on a grant agreement
- May be in the context of a consortium agreement
- Sui Generis Databases
- Data that need to flow between countries
- May contain confidential information/ trade secrets
- May contain personal data
- May contain geodata
- May contain/ be public sector information
- May be protected as cultural heritage

Regimes of Property

- Copyright
- Related Rights
- Sui Generis Data Base Right (SGDB)

- Patents
- Domain Names
- Trademarks

- Confidentiality agreements
- Funding limitations (open and closed)
- Consortium limitations

- contractual

There is IPR beyond Copyright

- Copyrighted material may contain other rights:
  - Trademarks
  - Trade secrets
  - Patents

- Patented material is affected by copyrighted material
  - Time of publication
  - Copyright affects disclosure

Regimes of Protection

- Personal Data (GDPR)
- Pharmaceutical regulation/ medical regulation
- Confidential Information/ Trade Secrets
- Cultural Heritage
- Research Ethics
- National Security
- Security
- Content/ platform regulation

Regimes of Openness

- PSI / Open Data
- Open Culture (Europeana)
- Open Science (EOSC/ OpenAIRE)
- Free flow of non-personal information (links with regulatory authorities)
- Copyright II
    - Public Domain
    - Limitations and Exceptions/ Fair Use/ Fair Dealing
    - Open Licences
- Patent publications and metadata

Ownership vs. Licensing

- Who owns the material is straightforward:
  - Consortium agreements
  - Funding Agreements
  - Co-authorship
  - Institutional framework

A question of who

- Initial ownership may be transferred:
  - RFO
  - RPO
  - Publisher
  - Collecting Society
  - Collecting Organisation
  - Spin off/ out
  - Licensee

- Do we own the material?

A question of when

- If I am to license out:
    - Decide on commercial exploitation
    - Disclose only when I am not bound by:
        - Confidentiality
        - Patent application pending

- After I have patented:
    - Can go for open access
    - Attention to the value model

A question of what

- Need to decide:
  - What to release what to keep
  - Collaborate with the TTO
  - Collaborate with the library
  - Collaborate with the NOAD

3

Considerations and Questions

Rights

Rights

IN ≥ OUT

Who am I?

Producer/ Collector of Research

> Need to obtain all necessary rights

Disseminator of Research

> Need to use the licences that support my **value model**

1. Is it mine?

- 3rd Party Copyright/ SGDB/ patents

- Personal Data

- Confidential Information
  - Trade secrets
  - Confidential info
  - National Security limitations

- Special regimes
  - Cultural heritage
  - Medical/ Health/ pharma

2. Is it truly mine?

A. Copyright and Patents
1. Public Domain
2. Am I covered by a limitation (fair use)?
3. Is the correct licence there?
   1. Open Licence
   2. Compatible Open Licence
4. Am I risking a patent filing?
5. What does my consortium/ grant agreement say?

C. Is it clear of any confidential information?

B. GDPR
1. Scientific Research Exception
2. Public Interest/ Mission
3. Legal Obligation
4. (Legal Interest)
5. Prior Informed Consent (PIC)

D. Have I obtained all special regimes licences?
- Security clearance
- Testing and protocols (life sciences)
- Special permissions (stats/ culture)

HOUSEKEEPING

3. How to share I (housekeeping)

- Record data provenance
- Record ownership (owner/ type of rights)
- Record limitations (e.g. classified/ confidential)a
- Record personal data and their legal basis
- Record licensing information
- Include notice and takedown procedure
- Record and follow patent limitations and process of filing
- Follow access protocols

4. How to Share II (licensing)

- Ensure you do not violate:
    - Access protocols
    - Patent filing process (disclosure)
    - Other contractual arrangementa and agreements
    - Confidentiality
- Include a PSI notice if applicable in your jurisdiction
    - Notice new PSI rules!
- Use standard licences (Creative Commons)
- Use Free Cultural Heritage Works Licences
    - CC Zero
    - CC BY 4.0 int.
    - CC BY SA 4.0 int
- If you use additional restrictions (ND/ NC)
    - Always provide a way to obtain more rights (through additional licences)

- If you use custom licences:
    - Understand what your value model is
    - Ensure no other standard licences exist
    - Ensure they do not contain any copyleft elements
    - Ensure they are machine readable
- If you use restrictive licences:
    - Ensure you reduce frictions through electronic transactions

5. How to Share III (infrastructure)

- Use standard Free/ Open Source Technologies for the repositories (e.g. CKAN/ Dspace etc)
- Use the appropriate Open Data Standards (data/ metadata/ Services)
- Use (open) APIs
- Identify core data sets (see new PSI directive)
- Interoperate with major initiatives (ELRC/ EOSC/ OpenAIRE/ Europeana/ Wikimedia)
- Follow FA(I)R principles
    - Findable
    - Accessible
    - Interoperable
    - Reusable

4 conclusions

Conclusions

- Research does not only include publications and repo-data
- There are several legal regimes covering research (some permissive – some restrictive) – we need to make the best use of them
- Always clear your content/ data before your release
- Clearing is not just about copyright
- Make use of Public Domain
- Follow protocols (patents/ ethics/ access)
- Watch the new PSI implementations / Open Science / platform/ content regulations
- Follow the Keep It Stupid Simple (KISS) principle in all licensing choices
- Use existing open infrastructure
- Interoperate with major initiatives – national, international, European, public and private

Q +A