epitweetr: وثائق المستخدم

Arabic translation of the following document: **epitweetr : user documentation**

This document is a translation provided by ECDC under the EU Initiative on Health Security. The original document was drafted in English and is available here https://www.ecdc.europa.eu/en/publications-data/epitweetr-tool ECDC is not responsible for the accuracy of the translation

الوصف

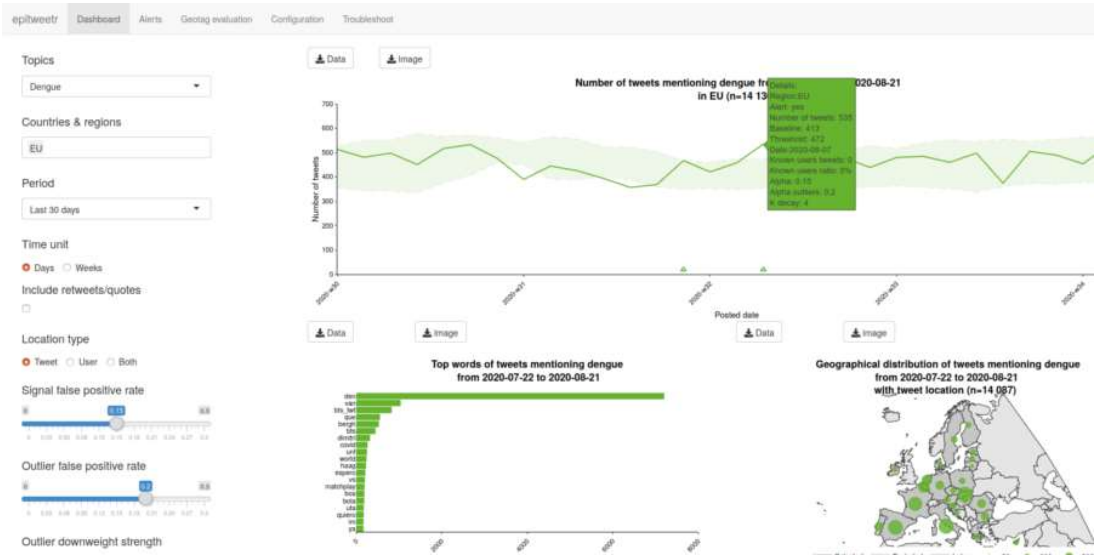
تساعدك حزمة `epitweetr` على رصد الاتجاهات السائدة للتغريدات بصورة آلية بحسب الزمان والمكان والموضوع المطروح. وتهدف عملية الرصد الآلية هذه إلى الكشف المبكر عن المخاطر التي تحدق بالصحة العامة من خلال اكتشاف الإشارات (على سبيل المثال، حدوث زيادة غير عادية في عدد التغريدات التي تنشر في وقت معين وذات مصدر محدد وتتناول موضوعًا معينًا). كما صُممت حزمة`epitweetr` بحيث تركز في آليتها على الأمراض المعدية، ويمكن توسيع نطاقها لتشمل جميع المخاطر أو ميادين أخرى للدراسة عن طريق تعديل الموضوعات والكلمات الرئيسية.

ويستند `epitweetr` إلى مبدأ عام واحد في آلية تشغيله ألا وهو جمع التغريدات والبيانات الوصفية ذات الصلة من واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر الإصدار 1.1 (https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/overview/standard) وفقاً لمواضيع محددة ثم تخزّن هذه التغريدات بشكل مجمع على جهازك الخاص. كما يحدد `epitweetr` موقع التغريدات الجغرافي ويجمع معلومات عن الكلمات الرئيسية التي احتوتها التغريدة. ثم تُصنف التغريدات وفقاً للموضوع والموقع الجغرافي. وبعد ذلك، تحدد خوارزميات قراءة الإشارات عدد التغريدات (حسب الموضوع والموقع الجغرافي) التي تتجاوز الحد المتوقع في يوم معين. ومن ثم يرسل `epitweetr` تنبيهاً عبر البريد الإلكتروني للقائمين على عملية التقصي لإخطارهم بضرورة مواصلة التحقيق في هذه الإشارات بعد تنفيذ عمليات الاستخبارات الوبائية (لفرزها والتحقق من صحتها وتحليلها وإخضاعها للتقييم الأولي).

وتتضمن الحزمة تطبيقًا تفاعليًا على الشبكة (Shiny app) مكوّنًا من خمس صفحات: صفحة dashboard، حيث يمكن للمستخدم مشاهدة البيانات المرئية للتغريدات واستكشافها (الشكل 1)، صفحة alerts، والتي تتيح استعراض التنبيهات الحالية والمعلومات المرتبطة بها (الشكل 2)، صفحة geotag evaluation، تساعد على تقييم خوارزمية تحديد الموقع الجغرافي في حقول تغريدة مختلفة لاختيار لحد الأدنى لتحديد الموقع الجغرافي يدوياً (الشكل 3)، وصفحة configuration، ويمكنك من خلالها تغيير الإعدادات والتحقق من حالة العمليات الأساسية (الشكل 4)، وصفحة troubleshoot، التي تمنحك عمليات فحص تلقائية وبعض الإرشادات النافعة لاستخدام `epitweetr` بجميع وظائفه (الشكل 5). في لوحة dashboard، يمكن للمستخدمين استعراض العدد الكلي للتغريدات على مدى فترة من الزمن، والموقع الجغرافي لهذه التغريدات على الخريطة والكلمات الأكثر شيوعاً في نصوص التغريدات. ويمكن فرز هذه البيانات المرئية بحسب المواضيع والمواقع والفترات الزمنية التي تتقصى عنها. كما تتوفر عوامل تصفية أخرى والتي تتضمن إمكانية ضبط الوحدة الزمنية للجدول الزمني، وما إذا كان ينبغي إدراج التغريدات المكررة/ المقتبسة، ونوع نظام تحديد الموقع الجغرافي الذي يحظى باهتمامك، وحساسية مجالات التنبؤ لاكتشاف الإشارة، وعدد الأيام المستخدمة لحساب الحد الأدنى للإشارات. يمكن تنزيل هذه المعلومات مباشرة من الواجهة إما على شكل بيانات أو صور أو تقارير:

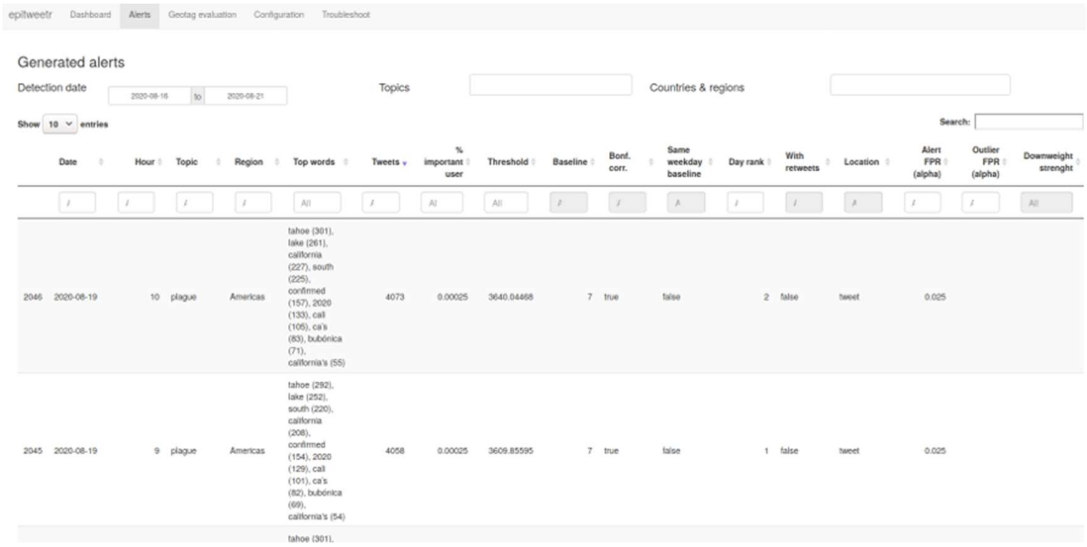
Shiny app dashboard:

الشكل 1: شكل Shiny app dashboard:



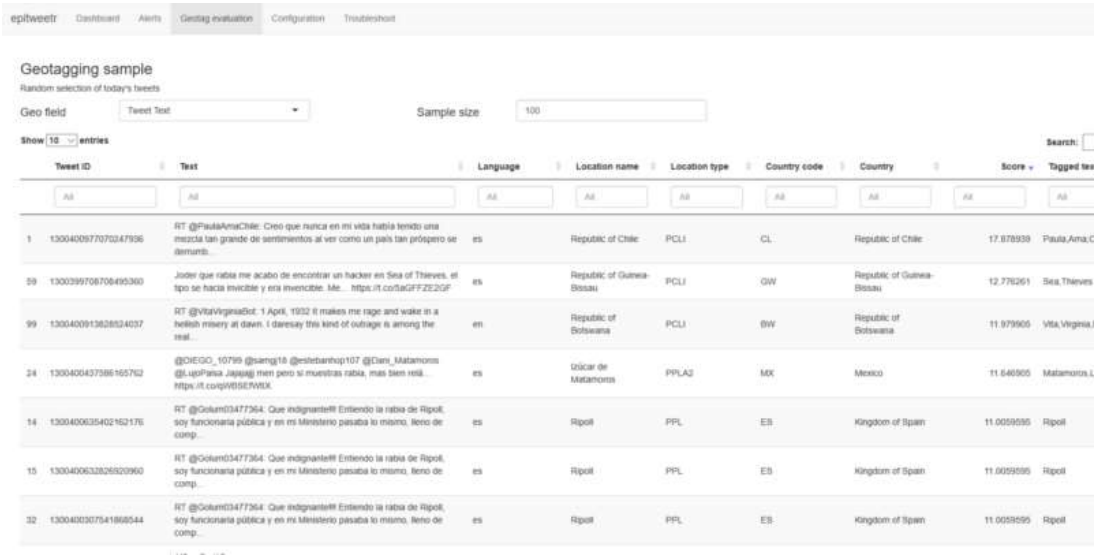
صفحة **Shiny app alerts**:

الشكل 2: صفحة Shiny app alerts



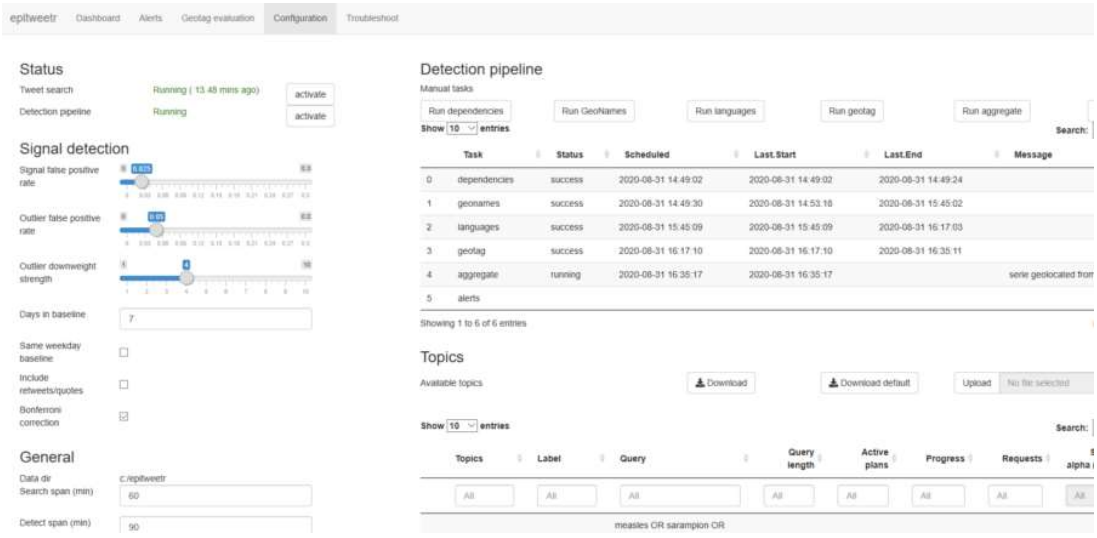
صفحة **Shiny app geotag evaluation**:

الشكل 3: صفحة Shiny app geotag evaluation



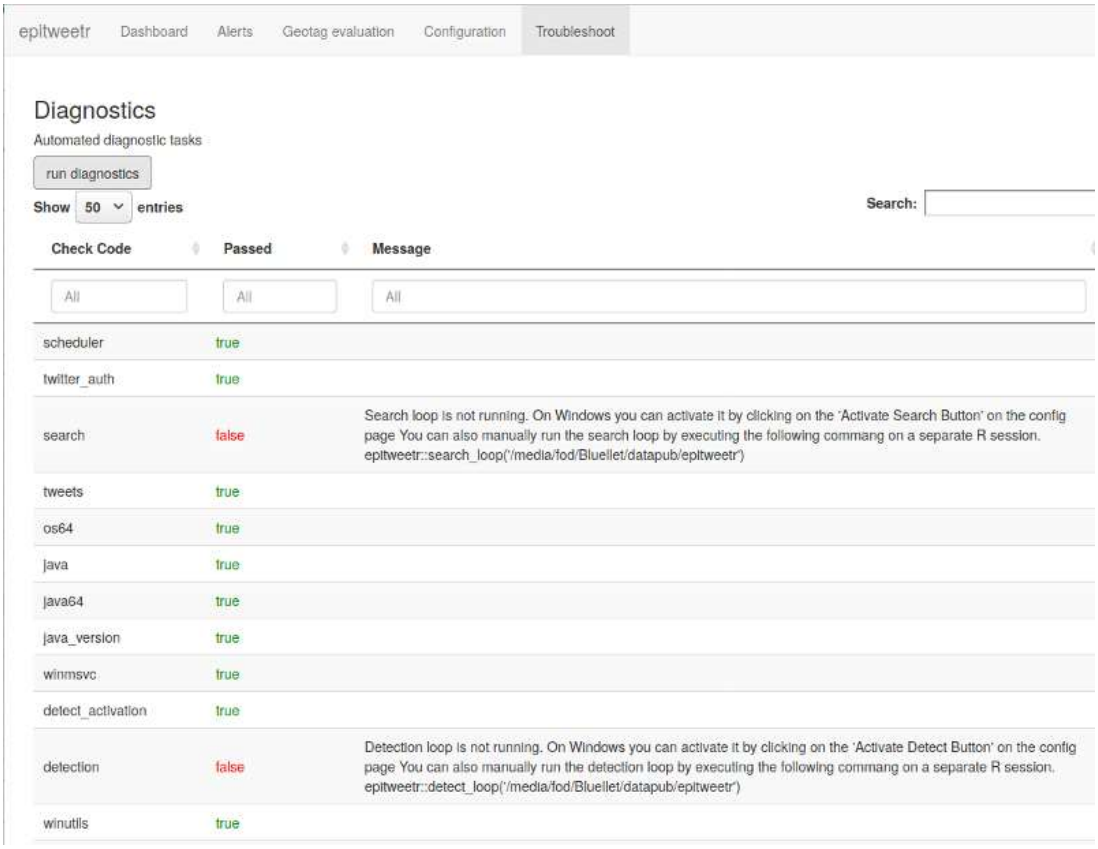
صفحة **Shiny app configuration**:

الشكل 4: صفحة Shiny app configuration



صفحة **Shiny app troubleshoot**:

الشكل 5: صفحة Shiny app troubleshoot



معلومات أساسية

الاستخبارات الوبائية في المركز الأوروبي للوقاية من الأمراض ومكافحتها (ECDC)

حددت المادة 3 من لائحة تمويل المركز الأوروبي للوقاية من الأمراض ومكافحتها (ECDC) والقرار رقم 1082/2013/EU الذي تناول المخاطر الحقيقية العابرة للحدود والتي تحدق بالصحة إجراءات الكشف عن المخاطر التي تهدد الصحة العامة كنشاط أساسي للمركز.

ويضطلع المركز بأنشطة الاستخبارات الوبائية (EI) التي تستهدف في آليتها الكشف السريع عن المخاطر التي تهدد الصحة العامة وتقييمها، مركزةً على الأمراض المعدية، لضمان الأمن الصحي للاتحاد الأوروبي. ويوظف المركز وسائل التواصل الاجتماعي كأحد مصادره لتساعده في عملية الكشف المبكر عن الإشارات التي توحي بوجود تهديد يحدق بالصحة العامة. وحتى عام 2020، تولى خبراء أو منظمات تم اختيارهم مسبقًا عملية مراقبة وسائل التواصل الاجتماعي والتي أجريت بصورة أساسية عبر إخضاع المنشورات لعمليات فحص وتحليل مستمرة، ولا سيما تلك المنشورة على موقعَي تويتر وفيسبوك.

تفضل بالاطلاع على المزيد من المعلومات والمشاركة في برامج تدريبية مباشرة على الشبكة عبر المواقع التالية:

مصادر الاستخبارات الوبائية

البرنامج التعليمي للاستخبارات الوبائية

رصد الاتجاهات السائدة على وسائل التواصل الاجتماعي

يقول الواقع إن بعض الإشارات لم يتم كشفها بالمطلق أو إنها اكتشفت ولكن ليس بصورة مبكرة بما فيه الكفاية بالطرق الموضحة أعلاه. ويسمح الرصد الآلي للبيانات الوصفية المأخوذة من وسائل التواصل الاجتماعي (مثل تحليل اتجاهات وسائل التواصل الاجتماعي) باكتشاف الإشارات التي قد يحدث أن لا تُكتشف عبر مراقبة حسابات في وسائل التواصل الاجتماعي وقع عليها الاختيار مسبقًا، كما تعمل على تحسين توقيت الكشف عن الإشارات.

يولد تحليل الاتجاهات على وسائل التواصل الاجتماعي بحسب الموضوع والزمان والمكان إشارات ذات صلة بعملية الكشف المبكر.

وفي عام 2019، طوّر المركز نموذجًا أوليًا لأداة يدوية قائمة على لغة البرمجة R للكشف المبكر عن الأخطار التي تهدد الصحة العامة والتي تفرزها بيانات تويتر. وتعد حزمة `epitweetr` امتداداً لنطاق هذا النموذج الأولي، فتتيح بذلك إمكانية أكبر لتحديد المواقع الجغرافية للتغريدات وقدراً أكبر من الأتمتة.

أهداف epitweetr

يتمثل الهدف الرئيسي من تطوير حزمة `epitweetr` في استخدام تطبيقات البحث القياسي الخاص بتويتر الإصدار 1.1 بغية رصد الإشارات المبكرة التي تُنذر بوجود تهديدات محتملة حسب الموضوع المطروح والوحدة الجغرافية.

بينما يتمثل هدفه الثانوي في منح المستخدمين القدرة، من خلال تزويدهم بواجهة Shiny التفاعلية، على استطلاع ليس فقط الاتجاهات السائدة للتغريدات بحسب توقيتها ومواقعها الجغرافية ومواضيعها، بل تمدهم أيضاً بمعلومات عن أهم العبارات الواردة في التغريدات وأعداد تغريدات المستخدمين الموثوق بهم، باستخدام الرسوم البيانية والجداول.

متطلبات الأجهزة

نورد في الجدول أدناه الحد الأدنى والمقترح لمواصفات المعدات الحاسوبية:

متطلبات الأجهزة	الحد الأدنى	المقترحات
ذاكرة الوصول العشوائي المطلوبة	8 غيغابايت	يُوصى باستخدام 16 غيغابايت
وحدة CPU المطلوبة	4 نويات	12 نواة
المساحة اللازمة للتخزين لمدة 3 سنوات	3 تيرابايت	5 تيرابايت

يمكنك تهيئة ذاكرة الوصول العشوائي (RAM) ووحدة المعالجة المركزية (CPU) من صفحة Shiny app configuration (انظر القسم تطبيق المستخدم التفاعلي (*Shiny app*)> صفحة التكوين). وقد تعتمد ذاكرة الوصول العشوائي (RAM) ووحدة المعالجة المركزية (CPU) والمساحة المطلوبة على مقدار وحجم الموضوعات التي تطلب الاستعلام عنها في عملية التجميع.

التثبيت

صُممت حزمة `epitweetr` لتكون منصة مستقلة، تعمل على أنظمة التشغيل Windows وLinux وMac. بيد أننا ننصحك باستخدام حزمة `epitweetr` على جهاز كمبيوتر يمكن له أن يعمل باستمرار ودون توقف. بإمكانك إيقاف تشغيل الجهاز، ولكن في حال امتدت فترة الإيقاف لساعات طويلة، قد يفوتك استقبال بعض التغريدات، الأمر الذي سيخلّف أثراً على آلية الكشف عن تنبيهات. كما يتوجب عليك تثبيت العناصر التالية قبل استخدام `epitweetr`:

الشروط الأساسية اللازمة لتشغيل epitweetr

- لغة R إصدار 3.6.3 أو إصدار أعلى منه

- جافا 1.8 مثلاً إصدار "1.8"openJDK . https://www.java.com/download/ننصحك باستخدام الإصدار 64 بت بدلاً من الإصدار 32 بت، لتجنب وجود قيود في الذاكرة. عند استخدامك نظام Mac، ننصحك بتثبيت عدة تطوير جافا [https://docs.oracle.com/javase/9/install/installation-jdk-and-jre-macos.htm

- وفي حال رغبت في تشغيله في نظام تشغيل Windows، ستحتاج لتثبيت Microsoft Visual C++، إلا أنك ستجده مثبّتاً مسبقاً في نظام جهازك في معظم الحالات:

  - حزمة Microsoft Visual C++ 2010 قابلة لإعادة التوزيع(x64) https://www.microsoft.com/en-us/download/details.aspx?id=14632

الشروط الأساسية اللازمة ليعض الخصائص الوظيفية في epitweetr

- برنامج Pandoc، لتصدير مستندات بصيغة PDF وMarkdown

  - https://pandoc.org/installing.html

- تثبيت Tex (نظام TinyTeX أو MiKTeX) (أو تثبيت TeX آخر) لتصدير مستندات بصيغة PDF

  - Easiest: https://yihui.org/tinytex/ يرجى تثبيته من لغة برمجيات R، وتسجيل الدخول/الخروج مطلوب بعد الانتهاء من التثبيت.

  - يستلزم إجراء عملية تثبيت كاملة من https://miktex.org/download، وتسجيل الدخول/الخروج مطلوب بعد الانتهاء من التثبيت.

- تحسين التعلم الآلي للنحو الأمثل (لأغراض المستخدمين المتقدمين فقط)

  - برنامج أوبن بلاز (BLAS optimizer) الذي سيزيد من سرعة بعض عمليات تحديد المواقع الجغرافية: https://www.openblas.net/ تعليمات التثبيت على الوصلة التالية: https://github.com/fommil/netlib-Java

  - أو Intel MKL (https://software.intel.com/content/www/us/en/develop/tools/math-kernel-library/choose-download.html)

- برنامج جدولة

  - في حال كنت تستخدم نظام تشغيل Windows، فيتوجب عليك تثبيت حزمة taskscheduleRR:

  - وفي حال كنت تستخدم نظام تشغيل Linux، فيتوجب عليك التخطيط للمهام يدوياً

  - أما إذا كنت تستخدم نظام تشغيل Mac، فيتوجب عليك تثبيت حزمة cronRR:

شروط أساسية إضافية لمطوّري لغة البرمجيات R

في حال رغبت في مواصلة تطوير epitweetr، فستحتاج لأدوات التطوير التالية:

- Git (تحكم الشفرة المصدرية) https://git-scm.com/downloads

- Sbt (شفرة تجميع سكالا) https://www.scala-sbt.org/download.html

- وفي حال كنت تستخدم نظام تشغيل Windows، فبالإضافة لما سبق، ستحتاج إلى Rtools: https://cran.r-project.org/bin/windows/Rtools/

تبعيات خارجية

بغية تشغيل epitweetr، يتعين تحميل بعض التبعيات الخارجية. وستعمل هذه الأداة من تلقاء ذاتها بمجرد انطلاق عملية إصدار التنبيهات لأول مرة. وستتيح لك صفحة Shiny app configuration إمكانية تغيير عناوين URLs الخاصة بهذه التبعيات، والتي تتألف مما يلي:

- *CRAN JARs*: وهي تبعيات متنقلة لتشغيل سبارك ولوسين وشفرة سكالا المدمجة. [*https://repo1.maven.org/maven2*]
- *Winutils.exe* (لنظام Windows فقط) هو هدوب ثنائي وضروري عند تشغيل سبارك محلياً على نظام تشغيل Windows. [*http://public-repo-1.hortonworks.com/hdp-win-alpha/winutils.exe*]

عملية تثبيت epitweetr من CRAN

بعد انتهائك من تثبيت جميع التبعيات المطلوبة المدرجة في قسم "الشروط الأساسية لتشغيل "epitweetr، يمكنك الآن البدء بتثبيت epitweetr:

```
install.packages(epitweetr)
```

متغيرات البيئة

بالإضافة إلى ذلك، تحتاج بيئة لغة البرمجيات R إلى معرفة المكان الذي سيتم تثبيت جافا فيه. ولتتحقق من هذا الأمر، اكتب في وحدة التحكم R ما يلي:

```
Sys.getenv("JAVA_HOME")
```

فإذا عاد الأمر إليك مُلغى أو فارغاً، يتوجب عليك تعيين متغير بيئة Java Home، لنظام التشغيل الذي تعمل عليه، يرجى الاطلاع على تعليمات نظام التشغيل الذي تعمل عليه. لكن في بعض الحالات، قد تجد epitweetr يعمل بدون تعيين متغير بيئة Java Home.

وفي حال فشلت الأداة في أول مرة تطلق فيها تشغيل التطبيق في تحديد مخزن آمن لكلمات مرور والذي يوفره نظام التشغيل، فستنبثق أمامك نافذة تطلب منك إنشاء حلقة مفاتيح كلمة مرور (لدى نظامي Linux وMac). وهو إجراء ضروري لتخزين بيانات اعتماد تويتر المشفرة. يرجى اختيار كلمة مرور قوية والاحتفاظ بها في ذاكرتك. فسُيطلب منك إدخال كلمة المرور هذه في كل مرة تقوم بتشغيل الأداة. بيد أنه يمكنك تجنب ذلك عبر تعيين متغير بيئة نظام يُسمى ecdc_twitter_tool_kr_password والذي يحتوي على كلمة المرور المُختارة.

الشروع بتشغيل *epitweetr Shiny app*

يمكنك الشروع في تشغيل تطبيق epitweetr Shiny app من جلسة R عبر الكتابة في وحدة التحكم R. استبدل "data_dir" بدليل البيانات المخصص وهو مجلد محلي عليك اختياره لتخزين التغريدات والسلسلة الزمنية وملفات التكوين في:

```
library(epitweetr)
epitweetr_app("data_dir")
```

لاحظ أن دليل البيانات الذي أدخلته في جلسة R يجب أن يحتوي على "/" بدلاً من "" (وبذلك يكون 'C:/user/name/Documents' مثال على المسار الصحيح). وتنطبق هذا الحالة بصورة خاصة في أنظمة التشغيل Windows في حال نسخت المسار من مستكشف الملف.

ويمكنك بدلاً من ذلك تشغيل التطبيق كما يلي: اكتب في أحد الملفات التنفيذية bat. أو shell، ما يلي (مستبدلاً "data_dir" بدليل البيانات المعين)

R –vanilla -e epitweetr::epitweetr_app("data_dir")

ويمكنك التحقق من تثبيتك لجميع المتطلبات بصورة صحيحة من صفحة استكشاف الأخطاء وإصلاحها. يمكنك الاطلاع على مزيدٍ من المعلومات في قسم تطبيق المستخدم التفاعلي *Dashboard > (Shiny app)*: واجهة المستخدم التفاعلية للبيانات المرئية> صفحة *troubleshoot*).

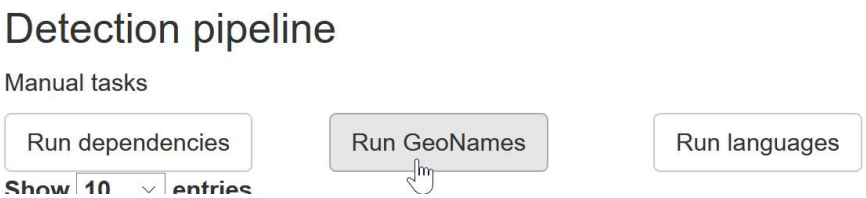
تعيين إعدادات مجموعة التغريدات وحلقة اكتشاف التنبيه

ليقدم لك epitweetrخدماته، يتعين عليك جمع التغريدات وتشغيل حلقة اكتشاف التنبيه (المسميات الجغرافية واللغات والوسم الجغرافي والتجميع وإشارات التنبيه). يمكنك الاطلاع على مزيد من التفاصيل في الأقسام اللاحقة من وثيقة المستخدم. نورد أدناه ملخصاً للخطوات المطلوبة:

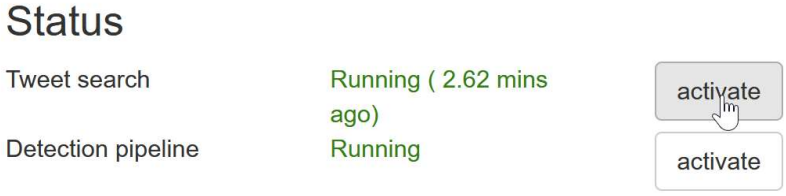
- اشرع في تشغيل Shiny app(من وحدة التحكم R)

```
library(epitweetr)
epitweetr_app("data_dir")
```

- في صفحة تكوين Shiny app، ثم من المهام اليدوية لـ ""Detection pipeline، انقر على " Run dependencies"و "Run geonames" و"Run languages" (عندها ستتغير حالتها إلى "pending"). وبذلك تسمح ل detection pipeline بتنزيل العناصر المطلوبة. وطالما أنه لم تُضاف لغات أخرى ولا تتوفر تحديثات جديدة على موقع geonames.org، ينبغي تشغيل هذه المهام فقط عند المرة الأولى التي تثبّت فيها حزمة epitweetr.



- ثم أعدْ عامل مصادقة تويتر باستخدام حساب تويتر أو تطبيق مطور تويتر، انظر قسم مجموعة التغريدات> مصادقة تويتر للاطلاع على مزيد من التفاصيل
- تنشيط مجموعة التغريدات
  - نظام تشغيل Windows: انقر على زر "Tweet search" activate"

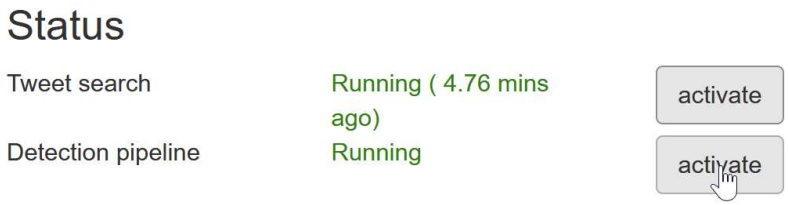


  - منصات أخرى: شغّل الأمر التالي في جلسة R جديدة

```
library(epitweetr)
search_loop("data_dir")
```

- يمكنك التأكد من أن مجموعة التغريدات مفعّلة في حال كان "Tweet search" status مُعين على وضعية "Running" وذلك على صفحة Shiny app configuration (النص المظلل باللون الأخضر في الصورة أعلاه) وإن كان في وضعية "true" في صفحة Shiny app troubleshoot.

- نشّط خط أنابيب الاكتشاف:
  - نظام تشغيل Windows: انقر على زر تنشيط "Detection pipeline"



  - منصات أخرى: شغّل الأمر التالي في جلسة R جديدة

```
library(epitweetr)
detect_loop("data_dir")
```

- يمكنك التأكد من أن خط أنابيب الكشف مفعّل في حال كان "Detection pipeline" status معين على وضعية "Running" وذلك في صفحة Shiny app configuration وفي وضعية "true" في صفحة Shiny app troubleshoot.
- ستبدأ برؤية التغريدات عقب الانتهاء من الخطوة التجميعية في جدول خط إنتاج الكشف في صفحة Shiny app configuration وفي حال تم تنشيط "Tweet search".
- وبذلك يمكنك أن تبدأ العمل مع ما يتوّلد من إشارات. نتمنى لك أوقاتاً سعيدة من الاكتشاف!

للاطلاع على مزيد من التفاصيل، تصفّح قسم كيف تعمل؟ البنية العامة الكامنة وراء تصميم *epitweetr*، والذي يصف العمليات الأساسية التي تتطلبها جمع التغريدات واكتشاف الإشارات. وأيضاً اطلع على" قسم تطبيق شايني التفاعلي (*Shiny app*)<صفحة *configuration*"، والذي بدوره يصف الإعدادات المختلفة على صفحة التكوين.

كيف تعمل؟ البنية العامة الكامنة وراء تصميم epitweetr

يقدّم هذا القسم وصفاً مفصلاً عن المبادئ العامة الواردة أعلاه. يمكن تكوين إعدادات العديد من هذه العناصر في صفحة تكوين Shiny app، الأمر الذي يوضحه قسم تطبيق شايني التفاعلي (*Shiny app*)<صفحة *configuration*.

تجميع التغريدات

استخدام واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر الإصدار 1.1

يستخدم epitweetr واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر ذات الإصدار 1.1. وتتميز هذه الواجهة بكونها خدمة مجانية مقدمة من تويتر فتُمكّن مستخدمي epitweetrمن الوصول إلى التغريدات مجاناً. لكن لا يُقصد بواجهة برمجة تطبيقات البحث أن تكون مصدراً شاملاً للتغريدات. فهي تجري عملية البحث عن طريق مضاهاة عينة من أحدث التغريدات المنشورة في الأيام السبعة الماضية مركّزة بحثها على مدى الأهمية وليس لغرض الاكتمال. ما يعني أننا قد نخرج بنتائج بحث تكون فيها بعض التغريدات والمستخدمين مفقودة.

ومع أن ذلك قد يبدو للوهلة الأولى تقييداً لمجالات أخرى من الصحة العامة أو البحث، يرى فريق تطوير epitweetr أن أخذ عينة من التغريدات لغرض الكشف عن الإشارة كافٍ لاكتشاف التهديدات المحتملة ذات الأهمية إذا ما اقترنت مع أنواع أخرى من المصادر.

وتشمل واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر ذات الإصدار 1.1 خصائص من قبيل:

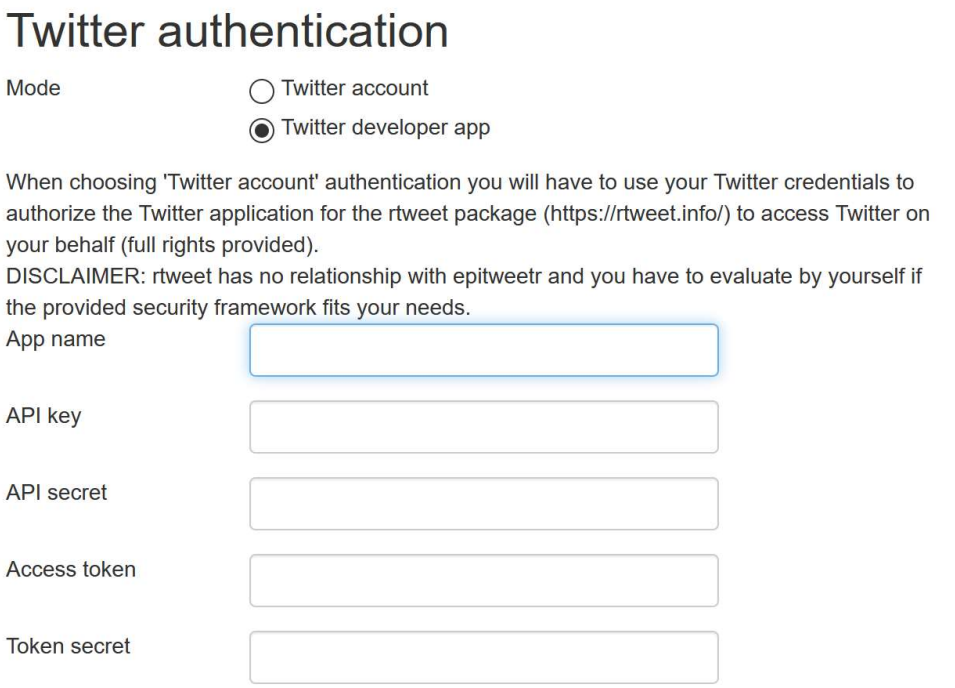
- فهرسة أحدث التغريدات المنشورة فقط في آخر 5 إلى 8 أيام بواسطة تويتر

- تدعم الواجهة 180 طلباً كحد أقصى كل 15 دقيقة (وذلك بواقع 450 طلباً كل 15 دقيقة إذا كنت تستخدم بيانات اعتماد تطبيق مطوّر تويتر، انظر القسم التالي)
- ينتج عن كل طلب 100 تغريدة و/أو تغريدة مكررة كحد أقصى

مصادقة تويتر

يمكنك مصادقة مجموعة التغريدات باستخدام **Twitter account** (يستخدم هذا النهج تطبيق حزمة رتويت) أو باستخدام **Twitter application**. وفيما يتعلق بالتطبيق الأخير، ستحتاج إلى Twitter developer account، الأمر الذي يستغرق الحصول عليه بعض الوقت، نظراً لإجراءات التحقق التي يطلبها. نوصي باستخدام حساب تويتر عبر حزمة رتويت تحقيقاً لأغراض الاختبار والاستخدام قصيرة الأجل، وتطبيق مطوّر تويتر للاستخدام طويل الأجل.

- عند استخدامك **Twitter account**: مفوّض عبر رتويت (مصادقة المستخدم)
  - ستحتاج إلى حساب تويتر (اسم المستخدم وكلمة المرور)
  - ثم ترسل حزمة رتويت طلباً إلى تويتر، ليتمكن من الوصول إلى حسابك على موقع تويتر نيابةً عنك
  - ثم تظهر نافذة منبثقة تطلب إدخال اسم مستخدم تويتر وكلمة المرور بهدف تأكيد أنك تسمح للتطبيق بالوصول إلى تويتر بالنيابة عنك. عليك أن ترسل هذا الرمز في كل مرة تدخل فيها لتستطلع التغريدات.
- عند استخدامك **Twitter developer app**: عبر `epitweetr` (مصادقة التطبيق)
  - إذا لم تفعل ذلك بعد، فستحتاج إلى إنشاء حساب مطور تويتر على الوصلة التالية: *https://developer.twitter.com/en/apply-for-access*
  - إنشاء تطبيق
  - بالنسبة لنوع الوصول، تأكّد من أن لديك حق الوصول للقراءة والكتابة
  - اكتب ملاحظة تحتوي على إعدادات OAuth الخاصة بك
    - أضفها إلى صفحة التكوين في تطبيق Shiny app (انظر الصورة أدناه)
    - باستخدامك هذه المعلومات، يستطيع `epitweetr` طلب رمز في أي وقت مباشرةً ليوفره لتويتر. تكمن ميزة هذه الطريقة في أن الرمز غير متصل بأي معلومات متعلقة بالمستخدم وتُرتجع التغريدات بصورة مستقلة عن أي سياق مستخدِم.
    - وباستخدامك هذا التطبيق، يمكنك تنفيذ 450 طلباً كل 15 دقيقة بدلاً من 180 طلباً في كل 15 دقيقة التي يسمح بها حساب تويتر.

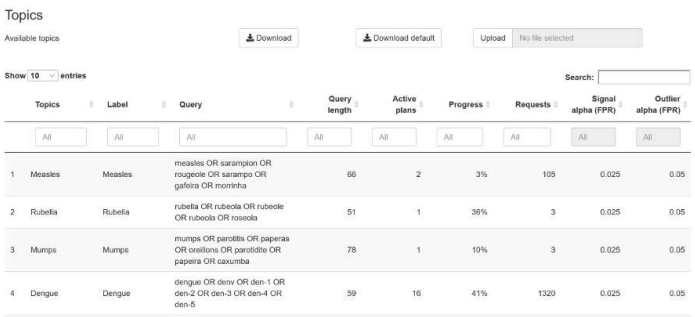


الموضوعات والاستعلامات بشأن تجميع التغريدات

بعد انتهائك من عملية مصادقة تويتر، يتعين عليك تحديد قائمة موضوعات في `epitweetr` لتُبيّن له التغريدات التي تودّ جمعها. بالنسبة لكل موضوع، يظهر لديك استعلام واحد أو أكثر يستخدمه `epitweetr` ليجمع التغريدات ذات الصلة (على سبيل المثال، عدة استعلامات عن موضوع ما باستخدام مصطلحات و/أو لغات مختلفة).

ويتكوّن الاستعلام من الكلمات الرئيسية وعامل تشغيل واللاتي تُستخدم لمطابقة سمات التغريدات. وتشير الكلمات الرئيسية المفصولة بمسافة إلى وجود عبارة \"و\". ويمكنك أيضاً استخدام معامل \ أو \. بينما يشير وجود علامة ناقص قبل الكلمة الرئيسية (بدون مسافة بين العلامة والكلمة الرئيسية) إلى أن الكلمة الرئيسية ينبغي ألا ترد في سمات التغريدة. ورغم أنه قد يمتد طول الاستعلام ليصل إلى 512 حرفاً، فإن الممارسة المُثلى في هذا الصدد أن تحصر استعلامك في 10 كلمات رئيسية وعامل تشغيل لتَحد من تعقيدات استعلامك، ما يعني أنك قد تحتاج لطرح أكثر من استعلام واحد عن كل موضوع.

تشمل حزمة `epitweetr` قائمة موضوعات افتراضية بالصيغة التي استخدمها فريق المركز الأوروبي للوقاية من الأمراض ومكافحتها المعني باستخبارات الأوبئة في تاريخ إصدار جيل الحزمة (1 سبتمبر 2020). بإمكانك الاطلاع على تفاصيل قائمة الموضوعات في صفحة تكوين تطبيق Shiny app (انظر لقطة الشاشة الواردة أدناه).



كما تتيح لك صفحة التكوين تنزيل قائمة الموضوعات وتعديلها وتحميلها إلى `epitweetr`. وبعد ذلك يجري استخدام قائمة موضوعات جديدة لتجميع التغريدات وتصبح مرئية في Shiny app. تظهر قائمة الموضوعات في ملف Excel (*.xlsx)

نظراً لأنها تتعامل مع إعدادات إقليمية خاصة بالمستخدم (مثل المحددات) ومع رموز خاصة أيضاً. ويمكنك إنشاء قائمة موضوعات خاصة بك وتحميلها أيضاً، على أن تراعي أن الهيكل يجب أن يتضمن على الأقل عنصر من العناصر التالية:

- اسم الموضوع مع عنوان "Topic" في جدول بيانات Excel. كما يجب أن يحتوي الاسم على أحرف أبجدية رقمية ومسافات وشرطات وشرطات سفلية فقط، على أن يبدأ بحرف هجائي.
- أما الاستعلام، المعنون بـ "Query" في جدول بيانات Excel. فهذا هو الاستعلام الذي تستخدمه حزمة `epitweetr` في طلباتها لتستدرج التغريدات من واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر. انظر لقطة الشاشة أعلاه لتحصل على صيغة الجملة وقيود الاستعلامات.

إلى جانب ذلك، يتضمن ملف topic.xlsx الحقول التالية:

- معرّف، تحت العنوان "#" في جدول بيانات Excel، يشير إلى معرّف عدد صحيح قيد التشغيل للموضوع.
- تسمية، تحت العنوان "Label" في جدول بيانات Excel، وهو ما يُعرض في قائمة الموضوعات المنسدلة لعلامات تبويب تطبيق Shiny app.
- مَعْلمة ألفا، تحت العنوان "Signal alpha (FPR)" في جدول بيانات Excel. وFPR هي اختصار " false positive rate" ومن شأن زيادة مستوى ألفا أن يُخفض الحد الأدنى لاكتشاف الإشارة، الأمر الذي سينتج عنه زيادة في الحساسية وربما استدارج المزيد من الإشارات. ويمكنك إعداد ألفا بصورة تجريبية ووفقاً لأهمية الموضوع وطبيعته.
- أما "Length_charact" فهو عبارة عن حقل يُنشأ تلقائياً تنطوي مهمته على حساب طول جميع الأحرف الأبجدية المستخدمة في الاستعلام. ويفيدك هذا الحقل ليلفت نظرك في حال تجاوز الطلب 500 حرف.
- ويعكس "Length_word" عدد الكلمات المستخدمة في الطلب، متضمناً ذلك عوامل التشغيل. والممارسة المُثلى في هذا الصدد أن تحصر استعلامك في 10 كلمات رئيسية.
- مَعْلمة ألفا، تحت العنوان "Outlier alpha (FPR)" في جدول بيانات Excel. وFPR هي اختصار " false positive rate" تعين مَعْلمة ألفا هذه المعدل الإيجابي الكاذب لتحديد القيم المتطرفة عند تخفيض ثقل القيم المتطرفة/الإشارات السابقة. وكلما انخفضت القيمة، قلّ عدد القيم المتطرفة السابقة المحتمل إدراجها. ومن المحتمل أن تتضمن القيمة الأعلى مزيداً من القيم المتطرفة السابقة.
- "Rank" هو عدد طلبات الاستعلام للموضوع الواحد

	A	B	C	D	E	F	G	H	I
1	#	Topic	Label	Alpha	Outliers Alpha	Query	Length_charact	Length_word	rank
2	1	Measles	Measles	0.025	0.05	measles OR sarampion OR rougeole OR sarampo OR gafeira OR morrinha	66	11	1
3	2	Rubella	Rubella	0.025	0.05	rubella OR rubeola OR rubeole OR rubeola OR roseola	51	9	1
4	3	Mumps	Mumps	0.025	0.05	mumps OR parotitis OR paperas OR oreillons OR parotidite OR papeira OR	78	13	1
5	4	Dengue	Dengue	0.025	0.05	dengue OR denv OR den-1 OR den-2 OR den-3 OR den-4 OR den-5	59	13	1
6	5	Haemorrhagic fever	Haemorrhagic fever	0.025	0.05	"hemorrhagic fever" OR "haemorrhagic fever" OR vhf OR "fiebre	129	18	1

عند تحميلك ملف، يرجى تعديل الموضوع وحقول الاستعلام، ولكن لا تعدّل عناوين الأعمدة.

الخطط المقرر إجراؤها لجمع التغريدات

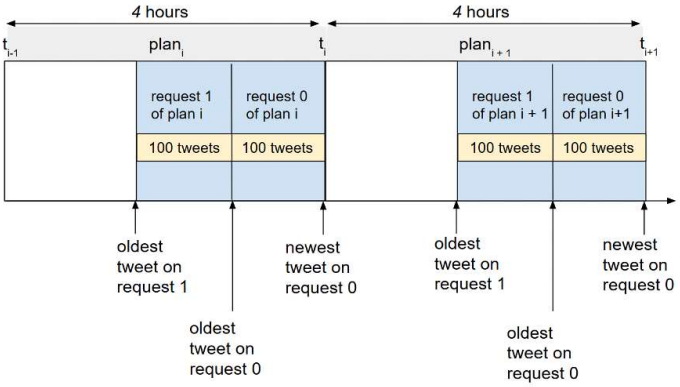
تذكّر أنه يُتوقع من حزمة `epitweetr` أن يرسل 180 طلباً (استعلاماً) إلى تويتر في كل 15 دقيقة (أو 450 طلباً في كل 15 دقيقة إذا كنت تستخدم بيانات اعتماد تطبيق مطوّر توتير). وكل استعلام يعود عليك بحصيلة 100 تغريدة. وتتنوع الحصيلة بين تغريدات وتغريدات مكررة. ويعود بتنسيق جسون، وهو صيغة تبادل بيانات خفيفة الوزن.

ولتخرج بأقصى حصيلة ممكنة من التغريدات، بالنظر لقيود واجهة برمجة تطبيقات البحث القياسية، ولكي لا تقف الموضوعات الشائعة كعائق أمام جمع الموضوعات الأخرى بصورة تحقق أغراض التقصي، يوظف `epitweetr` "خطط بحث" عن كل استعلام.

وتجمع "خطة البحث" الأولى للاستعلام التغريدات من التاريخ والوقت الحاليين وتعود في الزمان إلى الخلف حتى 7 أيام سابقة (7 أيام بسبب قيود واجهة برمجة تطبيقات البحث القياسية) قبل تنفيذه "خطة البحث" المدرجة. تعتبر "خطة البحث" الأولى الأكبر من نوعها، ذلك لأنه لم يحدث وأن جُمعت تغريدات من قبل.

بينما جميع "خطط البحث" التي ستلحق الأولى هي عبارة عن فواصل زمنية مجدولة جرى إعدادها في صفحة التكوين الخاصة epitweetr Shiny app (انظر القسم
التطبيق التفاعلي *Shiny app* > صفحة *configuration*> عام). دعنا نفترض لغرض التوضيح أن خطط البحث قد تقرر مثلاً إجراؤها كل أربع ساعات. تجمع الخطط تغريدات متعلقة باستعلام محدد من التاريخ والوقت الحاليين وحتى الأربع ساعات التي تسبق التاريخ والوقت عند تنفيذ "خطة البحث" الحالية (انظر الصورة أدناه). سيرسل epitweetrأكبر قدر ممكن من الطلبات (يصل عدد كل منها إلى 100 تغريدة) خلال مدة الأربع ساعات حسب الحاجة ليستدرج كافة التغريدات التي نُشرت خلال مدة الأربع ساعات.

فعلى سبيل المثال، إذا بدأت "خطة البحث" في الساعة 4 صباحاً من يوم 10 سبتمبر 2020، فستطلق شركة epitweetr طلبات للتغريدات تتناسب مع طلبات البحث الخاصة بها لمدة أربع ساعات وذلك من الساعة 4 صباحاً وحتى منتصف ليلة العاشر من سبتمبر 2020. أي يبدأ epitweetr بعمله بجمع أحدث التغريدات (من الساعة 4 صباحاً) بصورة زمنية عكسية. وفي حال توقفت واجهة برمجة التطبيقات عن عرض أي نتائج أخرى مستدرجة من المدة الزمنية الواقعة بين الساعة 4 صباحًا ومنتصف الليل، تعتبر "خطة البحث" لهذا الاستعلام مكتملة.



بيد أنه إذا كانت الموضوعات المُعالجة شائعة جداً (على سبيل المثال، جائحة كوفيد-19 في عام 2020)، فقد لا تُستكتمل "خطة البحث" الخاصة بالاستعلام في نافذة يمتد إطارها الزمني إلى أربع ساعات. وإذا حدث ذلك، ينتقل epitweetr إلى "خطط بحث" النافذة اللاحقة والتي يبلغ إطار مدتها الزمني أربع ساعات، ويضع أي "خطط بحث" سابقة غير مكتملة في قائمة انتظار للتنفيذ عند اكتمال "خطط البحث" لهذه النافذة الجديدة التي تبلغ مدتها أربع ساعات.

ثم تخزّن كل "خطة بحث" المعلومات التالية:

المجال	النوع	الوصف
النهاية_المتوقعة	الطابع الزمني	تاريخ الانتهاء من نافذة البحث الحالية
مجدول إجراؤه_لغاية	الطابع الزمني	التاريخ والوقت المجدول للطلب التالي. عند إنشاء الخطة، سيكون هذا هو DateTime الحالي وعقب كل طلب، يتم تعيين هذه القيمة لتصبح DateTime المستقبلية. لتحديد DateTime المستقبلية، سيخمن التطبيق عدد الطلبات اللازمة للانتهاء. إذا قدر أن الطلبات N ضرورية، فسيكون الجدول التالي في N/1 من الوقت المتبقي.
start_on	الطابع الزمني	DateTime وقت انتهاء الطلب الأول للخطة

end_on	الطابع الزمني	التاريخ والوقت الذي انتهى فيه الطلب الأخير للخطة إذا وصل هذا الطلب إلى تقدم الخطة بنسبة 100٪.
max_id	طويل	الحد الأقصى لمعرّف تويتر المستهدف من خلال هذه الخطة، والذي سيُحدد بعد معالجة الطلب الأول
since_id	طويل	آخر معرّف تغريدة جرى إرجاعه بعد معالجة آخر طلب في هذه الخطة. وسيبدأ الطلب التالي بتجميع التغريدات المنشورة قبل هذه القيمة. تُحدَّث هذه القيمة بعد تلقي كل طلب وتسمح لواجهة برمجة تطبيقات موقع تويتر باسترجاع التغريدات المنشورة قبل $min\_time(p_i)$
since_target	طويل	في حالة وجود خطة سابقة، تخزن هذه القيمة معرّف تغريدة كان قد حُمّل لتلك الخطة. ولن تجمع الخطة الحالية التغريدات المنشورة قبل هذا المعرّف. وتتيح هذه القيمة لواجهة برمجة تطبيقات موقع تويتر باسترجاع التغريدات المنشورة بعد p_i-time_back
الطلبات	كثافة الطلبات	عدد الطلبات المعالجة كجزء من الخطة
التقدم	الضعف	مدى تقدم الخطة الحالية كنسبة مئوية. وتحسب على الشكل التالي (current$max_id - current$since_id)/(current$max_id - current$since_target) -. في حال لم تُرجع واجهة برمجة تطبيقات تويتر أي تغريدة، فيُعين قيمة التقدم على 100٪. وتنطبق هذه الحالة فقط على الأخطاء غير المتعلقة بالاستجابات التي تحتوي على قائمة فارغة من التغريدات.

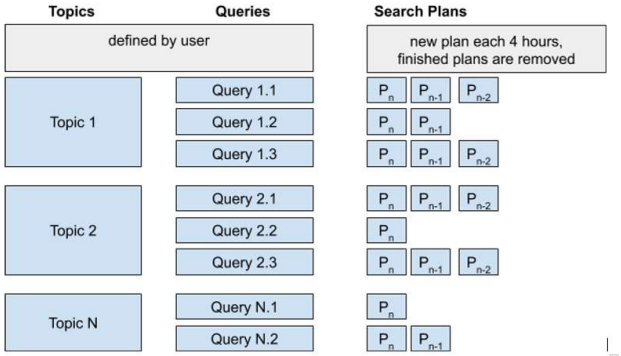
وتنفذ `epitweetr` الخطط وفقاً للقواعد التالية:

- يكتشف `epitweetr` أحدث خطة غير مكتملة لكل استعلام بحث مع المتغير Schedule_for الموجود مسبقاً.
- وينفذ `epitweetr` الخطط بأقل عدد ممكن من الطلبات التي سبق ونُفذت. ومن شأن هذا أن يضمن أن جميع الخطط المجدولة ستعالج العدد ذاته من الطلبات.
- ونتيجة للقاعدتين السابقتين، تنفذ واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر الطلبات المتعلقة بالموضوعات التي تقع ضمن حدود 180 حرف أولاً (أو 450 إذا كنت تستخدم مصادقة تطبيق مطوّر تويتر) محرزة بذلك تقدماً أعلى من ذلك المحرز فيما يتعلق بالموضوعات

ويكمن الأساس المنطقي وراء ذلك في أنه من المرجح أن هذه الموضوعات التي تتناول هذا العدد الكبير من التغريدات والتي لا يتسع لنافذة البحث التي تمتد ل 4 ساعات أن تجمعها في إطارها، هي موضوعات سبق وحظيت باهتمام. وبناءً عليه، يجب إيلاء الأولوية للموضوعات الأصغر وربما الموضوعات الأقل رواجاً.

وسنتناول جائحة كوفيد-19 في عام 2020 كمثال حيّ على ذلك. ففي أوائل عام 2020، كانت المعلومات المتاحة حول فيروس كوفيد-19 محدودة، الأمر الذي سهّل من اكتشاف الإشارات باستخدام معلومات مفيدة أو تحديثات مجدية (على سبيل المثال، البلدان الجديدة التي أبلغت عن حالات أو التي أكدت أنها إصابات نتيجة فيروس كورونا). بيد أنه وطوال مدة الجائحة، أصبح الموضوع حديث الساعة، ولم يعد الموضوع الواسع المتمثل بكوفيد-19 يجدِ نفعاً في اكتشاف الإشارات فاستغرق بذلك `epitweetr` وقتاً أطول من المتوقع وطلبات أكثر ليجري عملياته. ولتتجاوز هذا الحالة، قد يكون من الأنسب أن تحدد أولويات مجموعة الموضوعات الأصغر من قبيل موضوعات فرعية متعلقة بجائحة كوفيد-19 (كأن تستخدم اللقاح وكوفيد-19)، أو تحرص على ألا يفوتك أحداث أخرى تحظى باهتمام وسائل التواصل الاجتماعي بصورة أقل.

وفي حال تعذر استكمال خطط البحث، فقد تجد العديد من خطط البحث عن كل استعلام واقفة في قائمة الانتظار:



تحديد الموقع الجغرافي

يسعى epitweetr في عملية موازية لعملية تجميع التغريدات تحديد المواقع الجغرافية لجميع التغريدات التي جرى جمعها باستخدام عملية تعلم آلي غير مراقبة. تسير هذه العملية في نافذة الجدول المحددة بواسطة خاصية "Detect span" في صفحة configuration ضمن "General settings" (على سبيل المثال، إذا عينت نافذة مدتها أربع ساعات، فستجدها وقد تفعلّت كل أربع ساعات وستحدد المواقع الجغرافية لجميع التغريدات التي جُمعت بنجاح في آخر مرة.)

يخزن epitweetr نوعين من الأنماط التي يجري فيها تحديد الموقع الجغرافي للتغريدة: موقع التغريدة، وهي معلومات تحديد الموقع الجغرافي ضمن نص تغريدة (أو ضمن تغريدة مكررة أو مقتبسة)، وموقع المستخدم من واصفات البيانات المتاح. أما لاكتشاف الإشارة، فيُستخدم أفضل موقع مقترح بينما يمكن رؤية كلا النوعين في لوحة التحكم.

تحديد الموقع الجغرافي استناداً إلى موقع التغريدة

يستخلصepitweetr موقع التغريدة الجغرافي ويخزّن المعلومة استنادًا إلى معلومات تحديد الموقع الجغرافي الواردة في نص تغريدة. وفي حالة التغريدات المكررة أو المقتبسة، فيستخلص معلومات تحديد الموقع الجغرافي من نص التغريدة الأصلية المكررة أو المقتبسة. وفي حال عدم توافر أي منهما، فلن يحدد موقع للتغريدة بناءً على نص التغريدة.

ويُميزepitweetr ما إذا كان نص التغريدة يحتوي على إشارة إلى موقع جغرافي معين عبر تصنيف الكلمات إلى مجموعات ثم يجري عملية تقييم لتلك التي يرجح أن تكون موقعاً باستخدام نموذج التعلم الآلي. وتضيف الخوارزمية بدورها كلمات أخرى (واحدة تلو الأخرى) إلى المجموعة وإذا زادت النتيجة باستخدام المزيد من الكلمات، تحاول الخوارزمية أن تجد حد أقصى محلي للنتيجة بواسطة الاطلاع على نص أكبر. ثم تطابق هذه الكلمات مع قاعدة البيانات المرجعية، ألا وهي geonames.org. وهي قاعدة بيانات جغرافية متاحة للجميع ويمكن الاستفادة منها عبر خدمات الويب المختلفة، بموجب ترخيص المشاع الإبداعي. وتحتوي قاعدة بيانات GeoNames.org على أكثر من 25.000.000 مسمّى جغرافي.
يستخدمepitweetrضمن خياراته الافتراضية تلك التي تقتصر على الموجودة حالياً وتلك التي تتمتع بعدد سكان معروف (أي ما يزيد قليلاً عن 500000 مسمّى) ويمكنك أن تبّطل هذا الإعداد الافتراضي عبر التوجه لصفحة Shiny app configuration، ثم إلغاء تحديد "المسميات الجغرافية المبسطة". وتحتوي قاعدة البيانات أيضاً على سمات خطوط الطول والعرض للمواقع الجغرافية وتهجئات مختلفة (إحالات مرجعية)، والتي بدورها تطرح نتائج مفيدة عند البحث، بالإضافة إلى تهجئة العديد من المسميات بالحروف غير الرومانية.

ويمكن إجراء المطابقات في أي مستوى من مستويات التسلسل الهرمي الإداري. وتُدار عملية المطابقة بواسطة Apache Lucene، وهي مكتبة محرك بحث نصي مفتوح المصدر وعالية الأداء ومكتملة المواصفات.

وقد ينتج عن عملية المطابقة ربط بعض من أجزاء النص بعدة مواقع. ولكن يقع الاختيار في النهاية على الموقع الذي يحصد أعلى الدرجات فقط.

إذ ترتبط الدرجة الأعلى بقدر أكبر من احتمال صحة المطابقة. وتكون الدرجة:

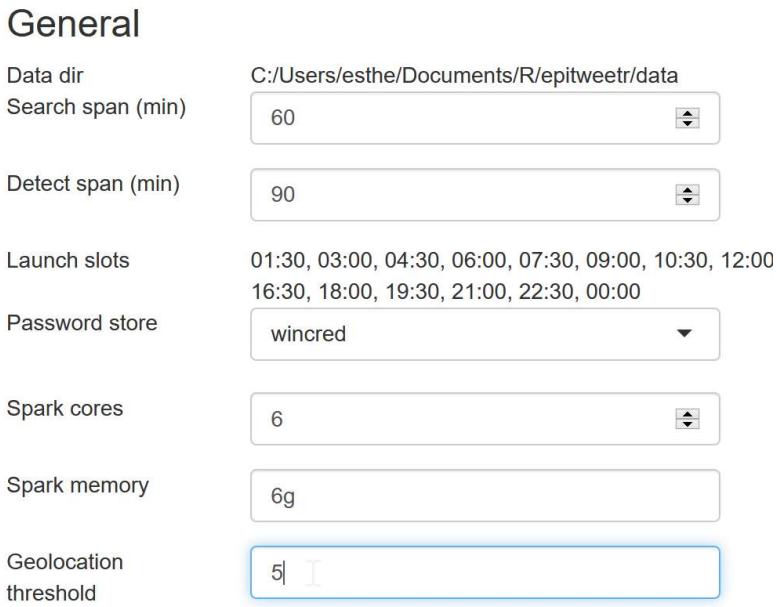
- أعلى في حالة تطابق أجزاء غير مألوفة من الاسم
- أعلى في حالة تطابق عدة مستويات إدارية
- أعلى إذا كان عدد سكان الموقع أكبر
- أعلى بالنسبة للبلدان والمدن مقابل المستويات الإدارية
- أعلى للاختصارات التي تأتي على شكل أحرف كبيرة مثل NY
- أقل بالنسبة للكلمات التي يُرجّح أنها ترتبط بأنواع أخرى للكلمات (مسميات غير جغرافية). على سبيل المثال بلدة فيربلاي في ولاية كولورادو. ويمكنك تحقيق ذلك باستخدام نماذج اللغة التي يوفرها fasttext.cc.

يمكنك تحديد languages التي تود أن تتحقق من وجود أنواع أخرى للكلمات فيها، عبر تحديد اللغة النشطة المطلوبة في صفحة configuration في Shiny app ثم النقر على أيقونة "+":

كما يمكنك إلغاء تحديد اللغات عبر تحديد اللغة في صفحة التكوين في Shiny app والنقر على أيقونة "-". ----------

ويمكنك تعيين الحد الأدنى من الدرجات ("geolocation threshold") على نطاق عالمي في الإعدادات العامة في صفحة التكوين للحد من عدد الإيجابيات الكاذبة (انظر الصورة). عندها ستتجاهل الخوارزمية جميع المواقع الجغرافية التي تحظى بدرجة أقل من الحد الأدنى المعين لتحديد الموقع الجغرافي ولن تعتمدها كموقع للتغريدة. وفي حال خروج أكثر من نتيجة واحدة متطابقة وقد تجاوزت الحد الأدنى من الدرجات، فسيقع اختيار الخوارزمية على نتيجة التطابق ذات الدرجات الأعلى.

يُعيّن الحد الأدنى بشكل تجريبي ويمكن تقييمها في ضوء القراءات البشرية للتغريدات ومواقع التغريدات الجغرافية، على صفحة تقييم العلامات الجغرافية.



تحديد الموقع الجغرافي استناداً إلى موقع المستخدم

توفر واجهة برمجة تطبيقات البحث القياسي الخاص بتويتر عبر واصفات بياناتها أنواعاً مختلفة لمواقع المستخدمين الجغرافية، حيث يقع اختيار epitweetr على أفضل موقع للمستخدم للملفات المجمعة متبعاً الترتيب التالي:

- موقع المستخدم الدقيق أو التقريبي عند نشره التغريدة (الأمر الذي توفره واجهة برمجة التطبيقات)
- إذا كان موقع المستخدم غير متوفر وكانت التغريدة عبارة عن تغريدة مكررة أو تغريدة مقتبسة، فيُستخدم موقع المستخدم الدقيق أو التقريبي في وقت نشر إعادة تغريد التغريدة /اقتباس التغريدة (الأمر الذي يوفره واجهة برمجة التطبيقات)
- إذا لم يكن متاحاً، فيُستخدم الموقع المعلن من قبل المستخدم
- إذا كان بدوره غير متاحاً، فيُستخدم موقع "الوطن" الوارد في الملف الشخصي العام.

ولتحديد المواقع بدقة، يتم تزويد خطوط الطول والعرض. وإذا تطابق هذا مع المكان التقديري، يَحسبepitweetrخطوط الطول العرض عبر GeoNames.org.

وفي حال لم توفر واجهة برمجة التطبيقات معلومات عن موقع المستخدم، يحسب epitweetrخطوط الطول العرض من الموقع المعلن لدى المستخدم أو من اسم المكان الوارد في "الملف الشخصي العام" للمستخدم، مستخدماً GeoNames.org.

معلومات التغريدة الجغرافية المخزنة

يُخزن الموقع الجغرافي الذي نتج عن المطابقة على شكل رمز بلد (مستخدماً معيار المنظمة الدولية للتوحيد القياسي 3166) وخطوط طول وعرض المرتبطين بالتحديد الدقيق للموقع الجغرافي في البيانات المجمعة.

أكثر الكلمات تواتراً في التغريدات

نظراً لضخامة عدد التغريدات والكلمات الواردة فيها، تخضع التغريدات لتحليل يستهدف تحديد أكثر الكلمات تواتراً في مجموعات من 10000 تغريدة لاستخلاص أكثر 500 كلمة استخداماً من كل مجموعة لذات اللغة واليوم والموضوع.

ولضمان أن يعمل epitweetr بمستوى أداء معقول، تُحدد أكثر 500 كلمة أولاً على مستوى العالم، ثم تُستخدم هذه المفردات في مجموعات فرعية حسب البلد، لاستخلاص أهم المفردات حسب البلد واليوم والموضوع. وفي المواقع الصغيرة جداً التي يخرج منها عدد قليل من التغريدات المحددة جغرافياً، قد لا يتوفر لدينا مفردات متواترة.

لاحظ أنه، وعلى عكس الأرقام المرئية الأخرى، تستند المفردات الأكثر تواتراً في التغريدات إلى الموقع الجغرافي المرتبط "بموقع التغريدات" بصورة دائمة وليس المرتبط "بموقع المستخدم" بغض النظر عن عامل التصفية المحدد في لوحة التحكم.

تجميع البيانات

تنتج عملية تجميع البيانات ثلاثة ملفات Rds (صيغة R أصلية): geolocated، وcountry_counts وtopwords.

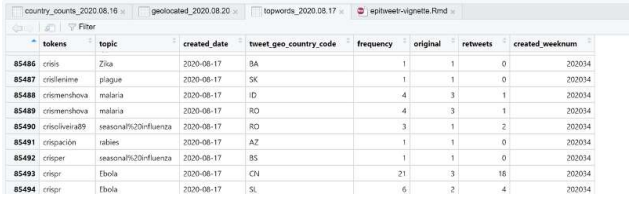
ففي ملف geolocated Rds، يُخزن عدد التغريدات أو التغريدات المكررة بحسب الموضوع والتاريخ وخطوط الطول والعرض للموقع الجغرافي لنص التغريدة والموقع الجغرافي وخطوط الطول والعرض للموقع الجغرافي للمستخدم. ويحتوي كل إدخال على البلد المرتبط بالموقع الجغرافي لنص التغريدة وعلى البلد المرتبط بالموقع الجغرافي للمستخدم (انظر لقطة الشاشة الجزئية أدناه). لاحظ أنه يجري تضمين التغريدات التي لا تحتوي على معلومات تُحدد الموقع الجغرافي.

	topic	created_date	user_geo_country_code	tweet_geo_country_code	user_geo_code	tweet_geo_code	user_geo_name	tweet_geo_name
1	COVID-19	2020-08-20	IT	AR	IT	AR	Italian Republic	Argentine Republic
2	COVID-19	2020-08-20	CA	US	6113365	4521037	Prince George	Pennsville
3	COVID-19	2020-08-20	US	US	4726206	4726206	San Antonio	San Antonio
4	gonorrhea	2020-08-20	CO	CO	CO	3687459	Republic of Colombia	El Caracol

ويُستخدم الملف ذات الصيغة country_counts Rds لإنشاء منحنى في Shiny app. وهو ملف Rds أصغر حجماً، بدون تضمين معلومات خطوط الطول والعرض، ويتضمن عدد التغريدات بحسب الساعة خلال اليوم، وبحسب البلد (وفقاً لموقع التغريدة أو موقع المستخدم)، والموضوع (انظر لقطة الشاشة)، وما إذا كانت التغريدة عبارة عن تغريدة مكررة أم لا. وتعطي الحقول known_retweets وknown_original عدد التغريدات أو التغريدات المكررة من قائمة "important users". كما يجري تضمين التغريدات التي لم تُحدد جغرافياً في هذا الملف. يتيح لك تضمين التغريدات التي لم يُحدَد فيها معلومات عن موقعها الجغرافي عرض جميع التغريدات عند اختيار "world" كمنطقة، بغض النظر عما إذا كان الموقع الجغرافي قد جرى تحديده بنجاح أم لا.

	topic	created_date	created_hour	tweet_geo_country_code	user_geo_country_code	retweets	tweets	known_retweets
33	COVID-19	2020-08-16	19	AU	US	71	13	0
34	COVID-19	2020-08-16	19	GH	PK	5	3	0
35	rabies	2020-08-16	21	PK	ES	20	0	0
36	gonorrhea	2020-08-16	01	VE	NA	1	0	0
37	COVID-19	2020-08-16	04	NA	PE	88	22	0

كما يُخزّن نتائج التجميع بحسب أكثر الكلمات استخداماً في ملف ذات الصيغة topwords.Rds، ويعرض عدد التغريدات أو التغريدات المكررة (أو كليهما) حسب الموضوع أو الكلمة الرئيسية أو التاريخ أو بلد موقع التغريدة وما إذا كانت التغريدة مكررة أم لا (انظر لقطة الشاشة).



	tokens	topic	created_date	tweet_geo_country_code	frequency	original	retweets	created_weeknum
85486	crisis	Zika	2020-08-17	BA	1	1	0	202034
85487	crisilenime	plague	2020-08-17	SK	1	1	0	202034
85488	crismenshova	malaria	2020-08-17	ID	4	3	1	202034
85489	crismenshova	malaria	2020-08-17	RO	4	3	1	202034
85490	crisoliveira89	seasonal%20influenza	2020-08-17	RO	3	1	2	202034
85491	crispación	rabies	2020-08-17	AZ	1	1	0	202034
85492	crisper	seasonal%20influenza	2020-08-17	BS	1	1	0	202034
85493	crispr	Ebola	2020-08-17	CN	21	3	18	202034
85494	crispr	Ebola	2020-08-17	SL	6	2	4	202034

اكتشاف الإشارة

يتمثل الهدف الرئيسي من تشغيل حزمة epitweetr في اكتشاف الإشارات في تدفقات البيانات المرصودة، أي الإحصاءات في السلاسل الزمنية المجمعة التي تتجاوز المتوقع لها. ويستخدمepitweetr لاكتشاف الإشارات نسخة موسعة من خوارزمية EARS (نظام الإبلاغ عن الانحراف المبكر) (فريكير وهيغلير ودونفي 2008)، والتي يُشار إليها فيما يلي بخوارزمية EARS(ears الموسعة). وتعد هذه الخوارزمية جزءاً من مراقبة حزمة لغة البرمجيات R (على النحو الذي طرحه سلمون وشوماخر وهوله 2016).

ويستخدم ضمن إعداداته الافتراضية، نافذة متنقلة للأيام السبعة الماضية لحساب الحد الأدنى. وفي حال تجاوز العدد لليوم الحالي هذا الحد، تولد إشارة.

تفاصيل الخوارزمية الكامنة وراء الكشف عن الإشارة

تُطبّق خوارزمية eears على المجاميع الآتية من كتل الأربع وعشرين ساعة السابقة الكتلة الأربع وعشرين ساعة الحالية. ويُحسب متوسط المحرك والانحراف المعياري الجاري على النحو التالي:

$$\overline{y}_0 = \frac{1}{7}\sum_{t=-7}^{-1} y_t \quad \text{و} \quad s_0^2 = \frac{1}{7-1}\sum_{t=-7}^{-1}(y_t - \overline{y}_0)^2,$$

حيث المعادلة $y_t, t = \cdots, -2, -1, 0$ تشير إلى السلاسل الزمنية لبيانات العد المرصودة مع مؤشر الوقت 0الذي يشير إلى الكتلة الحالية. وعلاوة على ذلك، يشير مؤشر الوقت $-1, \ldots, -7$ إلى الكتل السبع السابقة للكتلة الحالية.

وفي ظل وجود الفرضية الصفرية المتمثلة في عدم وجود طفرات، يُفترض أن y_t ستُوزع بصورة متماثلة ومستقلة $N(\mu, \sigma^2)$ مع متوسط غير معروف μ وتباين غير معروفσ^2. وبالتالي فإن الحد الأقصى لفاصلة زمنية تنبؤية لمكون إضافي بسيط بنسبة $100٪ \times (1-\alpha)$ أحادي الجانب y_0استناداً إلى $y_{-7}, \ldots, y_{-1}$ يُكتب على النحو التالي

$$U_0 = \overline{y}_0 + z_{1-a} \times s_0,$$

حيث z_{1-a} هو (α - 1) - دالة التوزيع الكمي للتوزيع الطبيعي القياسي. ويفعّل التنبيه إذا $y_0 > U_0$. إذا استخدمنا α=0.025، فهذا يقابل التحقيق، في حال تجاوز y_0 تقدير المتوسط زائد 1.96 ضعف الانحراف المعياري. ومع ذلك، كما أشار كل من أليفيوس وهوله (2017)، فإن النهج الصحيح هو مقارنة الرصد بالحد الأقصى لفاصل تنبؤي y_0ثنائي الجانب بنسبة 95 ٪، لأن هذا يعكس كلاً من تباين أخذ العينات لعملية رصد جديدة وعدم اليقين الناشئ عن تقدير المَعلمة للمتوسط والتباين.س وبالتالي فإن النموذج الإحصائي المناسب هو حساب الحد الأقصى على النحو

$$U_0 = \overline{y_0} + t_{1-a}(7-1) \times s_0 \times \sqrt{1+\frac{1}{7}}.$$

حيث $t_{1-a}(k-1)$ يشير إلى دالة التوزيع الكمي 1 - α لتوزيع-t مع 1 - k درجة من الحرية.

تخفيض ثقل الإشارات السابقة

في حال أُدرجت الإشارات السابقة دون أن يطرأ تعديل على القيم التاريخية عند حساب متوسط المتحرك والانحراف المعياري لاكتشاف الإشارة، فقد يصبح المتوسط التقديري والانحراف المعياري أكبر مما ينبغي له. وقد يعني هذا أن الإشارات الحالية الهامة لن تُكتشف. ولمعالجة هذه المشكلة، يحدُّ epitweetr من قيمة الإشارات السابقة، فيُعدّل بذلك المتوسط وتقدير الانحراف المعياري لهذه القيم المتطرفة باستخدام نهج مشابه لتلك المستخدمة لدى فارينغتون وآخرون. (1996). والقيم التاريخية التي لم تُحدد على أنها إشارات سابقة تُعطى وزنًا يساوي "1". وبالمثل، تُعطى القيم التاريخية المحددة كإشارات وزنًا أقل من واحد ومن ثم يجري تهيئة ملاءمة جديدة باستخدام هذه الأوزان (s.t. محددة الحجم تُجمع مرة أخرى مع 7 نتائج رصد). يمكنك الاطلاع على مزيد من التفاصيل حول إجراءات الحد من الأثر في الملحق 1 لوثيقة المستخدم هذه.

توقيت الكشف عن الإشارة

تُنفذ آلية اكتشاف الإشارة استناداً إلى عنصر "الأيام"، وهي عبارة عن نوافذ متحركة مدتها 24 ساعة، تتحرك وفقًا لمدى الاكتشاف (انظر أيضاً قسم تطبيق المستخدم التفاعلي (*Shiny app*> صفحة *configuration* > عام). ويُحسب الأساس المتعلق بعنصر "الأيام" من -1 إلى -8 (على أساس أن "اليوم" الحالي يساوي صفر).

وتتولد الإشارات وفقاً لمدى الاكتشاف (انظر قسم تطبيق المستخدم التفاعلي (*Shiny app*> صفحة *configuration* > عام)، عبر إرسال

- تنبيهات عامة في البريد الإلكتروني متبعةً مدى الاكتشاف هذا (على سبيل المثال، إذا امتد مدى الاكتشاف لأربع ساعات، فسُترسل إشارات تنبيه بواسطة البريد الإلكتروني في كل أربع ساعات)
- في الوقت الحقيقي. وتُحذف الإشارات التي نشأت بصورة مسبقة في نوع التنبيهات هذه.

ويمكن تخصيص الأنواع المختلفة لتنبيهات البريد الإلكتروني لكل مستخدم في صفحة التكوين (انظر قسم تطبيق المستخدم التفاعلي (*Shiny app*> صفحة *configuration* > عام).

مَعلمة ألفا: المعدل الإيجابي الكاذب لاكتشاف الإشارة

تتمثل السمة الرئيسية لاكتشاف الإشارة في قدرة الخوارزمية على اكتشاف التهديدات أو الأحداث الحقيقية دون أن تثقل كاهل المحققين بالكثير من الإيجابيات الكاذبة. وبهذه الطريقة، تُحدد مَعلمة ألفا عتبة الفاصل الزمني للكشف. في حال كانت قيمة ألفا مرتفعة، فتتولد المزيد من الإشارات المحتملة إما إذا كانت قيمة ألفا منخفضة، فيكون عدد الإشارات المحتمل ظهورها أقل (ولكن قد ينجم عن ذلك إهدار فرصة تلقي تهديدات أو أحداث المحتملة). وغالباً ما يُحدد إعداد مَعلمة ألفا يكونه تجريبي، وتعتمد أيضاً على الموارد المتوفرة بين أيدي من يحققون في الإشارات ومدى خطورة فقدان فرصة تلقي تهديد أو حدث محتمل.

وهناك مَعلمة ألفا عالمية النطاق، يمكن ضبطها/تغييرها عبر التوجه إلى صفحة `epitweetr` configuration تحت خيار "Signal false positive rate" (انظر قسم تطبيق المستخدم التفاعلي (*Shiny app*> صفحة *configuration* > عام) أضف إلى ذلك، يمكنك تعطيل خيار ألفا الافتراضي من قائمة الموضوعات. وفي حالة رغبت في ذلك، يمكنك أن تربط كل موضوع بمَعلمة ألفا محددة، اعتماداً على الأهمية المقدّرة لموضوع الصحة العامة أو للحدث المتعلق به أو لخطر التهديد المحدق به.

تصحيح بونفيروني

بغية حساب الاختبارات المتعددة كإعداد افتراضي لاكتشاف الإشارة الخاصة بكل بلد، تُقسم قيمة ألفا على عدد البلدان. أما بالنسبة لاكتشاف الإشارات الخاصة بالقارة، فتُقسّم قيمة ألفا على عدد القارات. وهذا ما يسمى بتصحيح بونفيروني وهو نوع من اختبارات المقارنة المتعددة.

ولتعطيل هذه الخاصية، يمكنك إلغاء تحديد خيار "تصحيح بونفيروني" بالتوجه إلى قسم "اكتشاف الإشارة" في صفحة تكوين Shiny app.

استخدام أيام الأسبوع ذاتها كأساس

قد يكون هناك "تأثير يوم من أيام الأسبوع"، حيث يشهد يوم معين من أيام الأسبوع حركة أكبر في نشر التغريدات مقارنة بباقي الأيام (على سبيل المثال، يوم الاثنين). ولتجنب حدوث ذلك، تحدد اختيار حساب الأساس ليس على شكل أيام متتالية، بل على شكل N أيام مضت التي تتوافق مع ذات النافذة الممتدة لـ24 ساعة وN أيام إلى الوراء. وبهذه الطريقة إذا كان N = 7، يُحسب الأساس باستخدام "الأيام" من -7 و-14 و-21 و-28 و-35 و-42 و-49 و-56 (إذا كان "اليوم" الحالي يساوي صفر).

هذا الخيار متوفر في صفحة التكوين Default same weekday baseline "Shiny app".

إرسال تنبيهات عبر البريد الإلكتروني

يُرسل `epitweetr` تلقائياً إشعارات عبر البريد الإلكتروني للتنبيه بقائمة الإشارات المكتشفة وفقاً لمدى الاكتشاف وقائمة المشتركين. وبسبب الوقت الذي تستهلكه عمليات جمع التغريدات وتحديد موقعها الجغرافي وتجميعها، تفقد التنبيهات عبر البريد الإلكتروني أحدث التغريدات التي لم تدخل بعد في هذه العمليات. يُقدر الفاصل الزمني بين التغريدات والتنبيهات بأقل من (2 * detect_span + (collect_span)) والتي تعادل 3 ساعة و30 دقيقة باستخدام القيم الافتراضية.

وتتضمن إشعارات التنبيه عبر البريد الإلكتروني المعلومات التالية حول الإشارات لكل موضوع:

- التاريخ والساعة التي وقع فيها اكتشاف الإشارة
- الموقع (المواقع) الجغرافية التي أكتشفت فيها الإشارة
- الكلمات الأكثر تواتراً (قائمة بأهم الكلمات) في التغريدات
- عدد التغريدات والحد الأدنى
- نسبة التغريدات التي غردها أهم المستخدمين
- معلومات عن الإعدادات، من قبيل: هل أُستخدم تصحيح بونفيروني، وهل أُستخدم أساس يوم الأسبوع، وهل أُدرجت التغريدات المكررة، إلخ.

وتتوفر هذه المعلومات أيضاً في صفحة التنبيهات في Shiny app.

يستطيع المشتركين تلقي إشعارات تنبيه في الزمن الحقيقي (أي بمجرد الانتهاء من حلقة الكشف) أو التنبيهات المجدولة (على سبيل المثال مرة أو مرتين في اليوم واحد). كما يمكنك تغيير قائمة المشتركين في صفحة التكوين عن طريق تنزيل جدول بيانات بصيغة Excel. يحتوي هذا الملف على المتغيرات التالية:

- "User": اسم المشترك (على سبيل المثال، جين دو).
- "Email": البريد الإلكتروني الخاص بالمشترك (مثل.(jane.doe@email.com
- Topics"": قائمة الموضوعات التي سيتلقى المشترك إشعارات تنبيه مجدولة بشأنها. يجب أن تتطابق الأسماء المستخدمة مع عمود "Topic" في قائمة الموضوعات.
- "Excluded": الموضوع الذي لن يتلقى المشتركون إشعارات تنبيه مجدولة بشأنه.
- "Real time Topics": قائمة بالموضوعات التي سيتلقى المشترك إشعارات تنبيه بشأنها في الوقت الحقيقي.
- "Regions": قائمة المناطق التي سيتلقى المشترك إشعارات تنبيه مجدولة.
- "Real time Regions": قائمة المناطق التي سيتلقى المشترك إشعارات تنبيه في الوقت الحقيقي لها.
- "Alert Slots": وهي فتحات حلقة الكشف التي سيتلقى المشترك بعدها إشعارات تنبيه مجدولة. يمكنك الحصول على الفتحات المتاحة من "Launch slots" في قسم "General" من صفحة configuration. وفي حال عدم تضمين أي قيمة، فسيتلقى المشترك إشعارات تنبيه في الوقت الحقيقي لجميع الموضوعات والمناطق، حتى إذا كانت هناك مواضيع أو مناطق محددة بصورة آنية في جدول بيانات Excel.

وعند تضمين أكثر من موضوع و/أو منطقة في قائمة المشتركين، ينبغي فصلها بفاصلة منقوطة (؛) ومن دون وضع مسافات (على سبيل المثال الإيبولا؛ الأمراض المعدية؛ حمّى الضنك). كما يجب أن تتطابق الأسماء مع عمود "Topic" في قائمة الموضوعات ومع عمود "Name" في قائمة البلد/المنطقة من صفحة التكوين.

B	C	D	E	F	G	H	I
User	**Email**	**Topics**	**Excluded Topics**	**Real time Topics**	**Regions**	**Real time Regions**	**Alert Slots**
Jane Doe	jane.doe@email.com			infectious diseases;zoonoses		Southern Europe;EU+EEA	8;20

تركيبة المجلد

تُخزّن حزمة `epitweetr` التغريدات والتغريدات المجمعة والتهيئة التي تمت في المجلد المعنون "dat folder" الذي يتوجب عليك تعيينه عندما تشرع بتشغيل التطبيق.

يحتوي مجلد **data** على 3 ملفات جسون:

- properties.json، الذي ينتج من معلومات الخصائص العامة لتطبيق Shiny app
- يُدار ملف topic.json بواسطة حلقة البحث: فهو يتعقب خطط جمع التغريدات ومدى التقدم الحاصل في أداء المهام.
- يُدار ملف tasks.json بواسطة حلقة الكشف: فهي تحتفظ بمعلومات المهام التي تنفذها هذه العملية وحالاتها المختلفة.

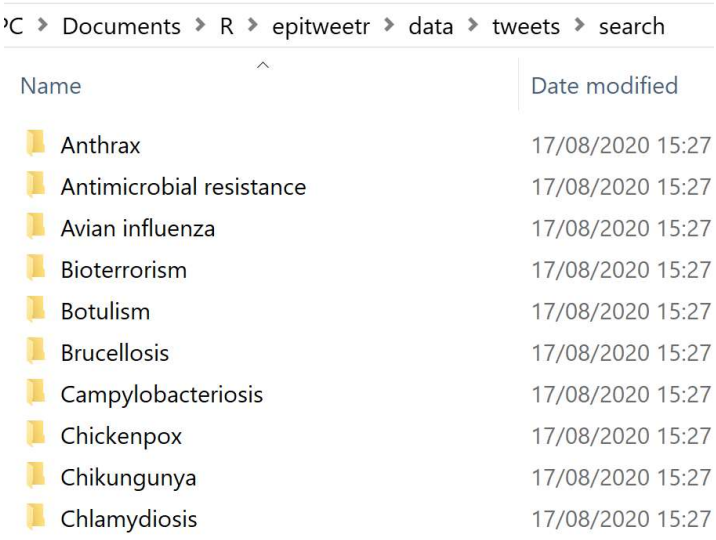
كما يحتوي المجلد على المجلدات الفرعية التالية:

- "geo"، يخزن بيانات GeoNames على هيئة ملفات نصية ومفهرسة
- "hadoop"، يخزن تبعيات سبارك الخاصة بأنظمة تشغيل Windows
- "jars"، يخزن مجموعات تبعيات جافا المطلوبة في عمليات تحديد المواقع الجغرافية والتجميع
- "languages"، يخزن فهارس ونماذج ملفات النص السريع التي تُستخدم لأداء تحديد الموقع الجغرافي في نص التغريدة
- "stats"، يخزن ملفات جسون وإحصاءات التقارير المستخدمة لتحسين العملية الإجمالية عن طريق ربط ملفات التغريدات بالتواريخ المنشورة للتغريدات
- "alerts" يخزن ملفات جسون المتعلقة بالتنبيهات المكتشفة بواسطة حلقة الاكتشاف.
- "tweets" و"series" نورد أدناه مزيداً حولهما.

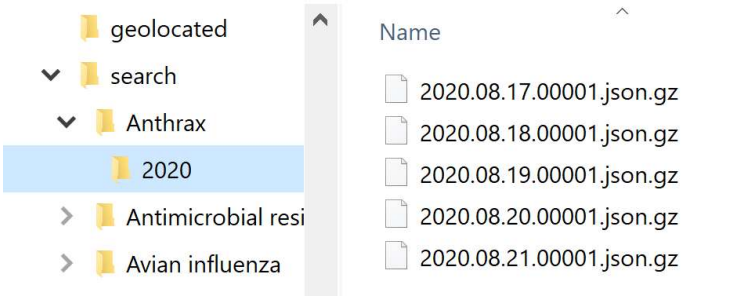
Data folder > tweets

في مجلد البيانات، يحتوي المجلد الفرعي "tweets" على مجلدين فرعيين آخرين وهما: **search** و**geolocated**

يحتوي المجلد *search folder* على مجلدات فرعية لكل موضوع مدرج في قائمة الموضوعات:



وداخل كل موضوع من هذه الموضوعات، نجد عام (على سبيل المثال 2020) ثم ملف جسون مضغوط يحتوي على التغريدات لكل يوم من كل عام. وتنحصر التواريخ بالتاريخ الذي جُمعت بها التغريدات (وليس بتاريخ نشرها). وقد نجد أكثر من ملف واحد لليوم الواحد إذا تجاوز الملف 100 ميغا بايت.



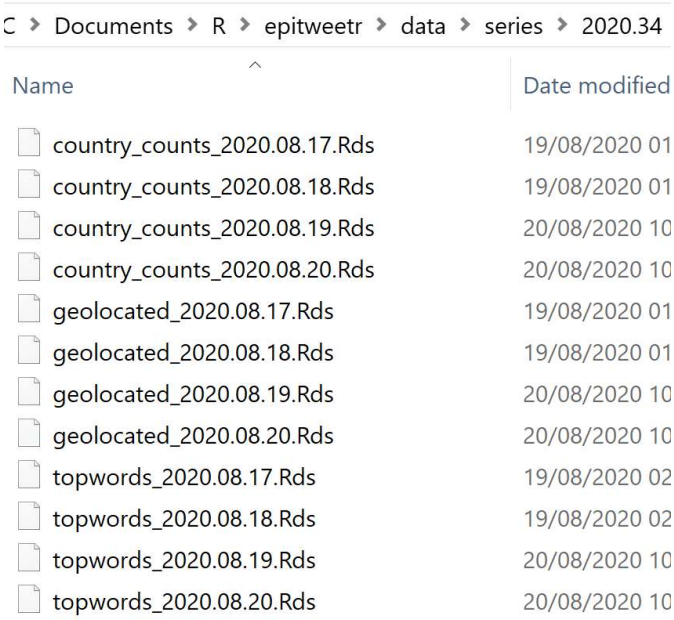
يحتوي المجلد *geolocated folder* على ملفات جسون مضغوطة مع معلومات تحديد المواقع الجغرافية التي أنتجتها خوارزمية تحديد الموقع الجغرافي.

Data folder > series

في مجلد *series*، يُخزّن `epitweetr` البيانات المجمّعة للتغريدات التي حُدد موقعها الجغرافي بالإضافة إلى أهم الكلمات.

يوجد مجلد لكل أسبوع محدد بالتقويم الأسبوعي لتاريخ المجموعة يحتوي على ملفات Rds (ملف بصيغة R أصلية) عن كل يوم وسلسلة:

- يحتوي geolocated_YYY.MM.DD.Rds على عدد التغريدات اليومية مع أدق مستوى ممكن للموقع
- يحتوي topwords_YYY.MM.DD.Rds على أهم الكلمات المتواترة في التغريدات اليومية على مستوى البلد
- يحتوي country_counts_YYY.MM.DD.Rds على عدد التغريدات بالساعة على مستوى البلد



هذا هو إجمالي المعلومات الواردة في قسم "كيف يعمل؟ البنية العامة الكامنة وراء تصميم epitweetr> آلية التجميع".

تطبيق المستخدم التفاعلي (Shiny app)

يمكنك الشروع بتشغيل تطبيق المستخدم التفاعلي (Shiny app) عبر حزمة epitweetr من جلسة R عن طريق الكتابة في وحدة التحكم R (استبدلْ "data_dir" بدليل البيانات المطلوب):

```
epitweetr_app("data_dir")
```

ويمكنك بدلاً من ذلك تشغيل التطبيق كما يلي: في أحد الملفات التنفيذية bat أو sh، ضع المحتوى التالي، (مع استبدال "data_dir" بدليل البيانات المتوقع)

R –vanilla -e epitweetr::epitweetr_app('data_dir')

يحتوي تطبيق المستخدم التفاعلي epitweetr على خمس صفحات:

- dashboard، تُمكّن المستخدم من رؤية التغريدات واستكشافها
- صفحة configuration، تُمكّن المستخدم من تغيير الإعدادات والتحقق من حالة العمليات الأساسية
- صفحة alerts، تُمكّن المستخدم من استعراض إشعارات التنبيه الحالية والمعلومات المرتبطة بها
- صفحة geotag evaluation، تُمكّن المستخدم من تقييم خوارزمية تحديد المواقع الجغرافية في حقول تغريدة مختلفة لاختيار الحد الأدنى لتحديد المواقع الجغرافية يدوياً
- صفحة troubleshoot، تفسح المجال لإجراء عمليات فحص تلقائية وتعطي المستخدم تلميحات حول استخدام epitweetr بجميع وظائفه

Dashboard: واجهة المستخدم التفاعلية للبيانات المرئية

dashboard هي المكان الذي يمكنك فيه استكشاف البيانات المصوّرة للتغريدات بشكل تفاعلي. ويتضمن خط رسم بياني(خط الاتجاه) مع إشعارات تنبيه وخريطة وأكثر الكلمات تواتراً في التغريدات التي تناولت موضوع معين. لاحظ أنه في المرة الأولى التي يُحدد فيها فترة ما، عليك الانتظار 1-2 دقيقة حتى تلمس النواتج. وبصورة مماثلة، بالنسبة لأي اختيار جديد (كإضافة منطقة أو تغيير الموضوع، إلخ)، تبدأ حزمة epitweetr في قراءة البيانات المقابلة؛ لذلك إذا حُددت عدة اختيارات، فقد تحتاج إلى الانتظار لمدة 1-2 دقيقة حتى يظهر تحديد الاختيار الأخير في dashboard. وعندما تقرأ epitweetr بيانات جديدة، تكون المخرجات أقل كثافة الأمر الذي يمنحك إشارة إلى أن البيانات الجديدة هي قيد القراءة والتخطيط.

ولاستكشاف البيانات بصورة تفاعلية، يمكنك أن تختار من بين عدة عوامل تصفية، من قبيل الموضوعات والبلدان والمناطق والمدة الزمنية والوحدة الزمنية ومؤشر الثقة والأيام في الأساس.

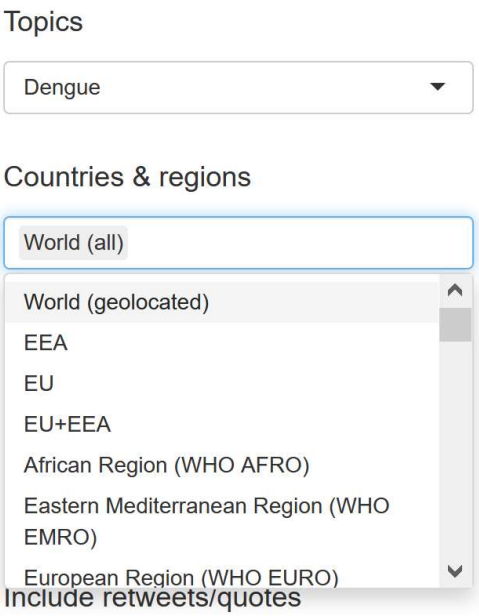
لاحظ أنه مهما كانت الخيارات/الإعدادات التي تحددها على لوحة التحكم، فلن تخلف أي تأثير يذكر على آلية كشف التنبيه. تُحدد جميع إعدادات الكشف عن التنبيهات في صفحة التكوين الخاصة بتطبيق Shiny app.

Filters

Topics

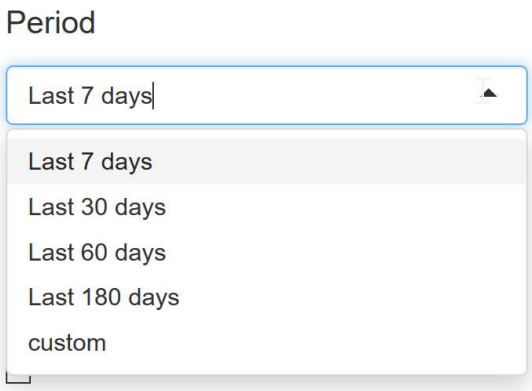
يمكنك تحديد عنصر واحد من القائمة المنسدلة للموضوعات، والتي مُلِئت بما تم تحديده في الموضوعات على صفحة التكوين. ويمكنك أيضاً الشروع في الكتابة في حقل النص وتعيين الموضوعات من القائمة المنسدلة التي تمت تصفيتها.

Countries & regions



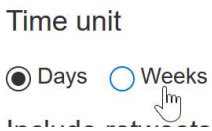
إذا حددت **World (all)**، فستُعرض جميع التغريدات بغض النظر عن موقعها الجغرافي. يمكنك تحديد بلد بحد ذاته، وتحديد المناطق والمناطق الفرعية، وتحديد عدة عناصر في الوقت ذاته. كما يمكنك أيضاً البدء في الكتابة في حقل النص وتحديد العنصر الجغرافي من القائمة المنسدلة.

Period



يمكنك الاختيار من الأيام السبعة الماضية (التعيين الافتراضي) أو من الأيام 30 أو 60 أو 180 الماضية. ويمكنك أيضاً تحديد "custom" وعندها يظهر خيار التقويم لتعيين مدة التقصي والبحث. وبذلك تُحدد الفترة الزمنية لإدراجها في البيانات المرئية. عند اختيارك فترة مخصصة، يرجى التأكد من أن التاريخ الأول يقع قبل يوم واحد على الأقل من التاريخ الثاني.

Time unit

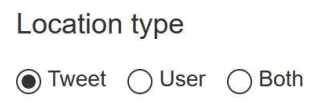


يمكنك عرض الجدول الزمني لعدد التغريدات مقترناً بوحدات زمنية على شكل أسابيع أو أيام. والتعيين الافتراضي هو الأيام.

Include Retweets/quotes

بحسب التعيين الافتراضي، لا تُدرج التغريدات المكررة في البيانات المرئية. إذا عينت خيار " include retweets/quotes"، فستعرض البيانات المرئية نتائج التغريدات والتغريدات المكررة/المقتبسة. وخلافاً لذلك، تعرض البيانات المرئية التغريدات فقط (دون التغريدات المكررة/المقتبسة).

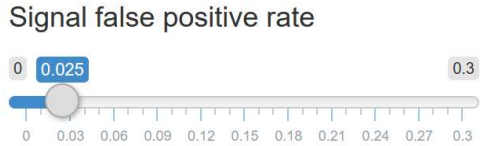
Location type



تُحدد مواقع التغريدات جغرافياً في المناطق والمناطق دون الإقليمية والبلدان. ويشير "نوع الموقع" إلى ما يتوجب استخدامه لتحديد الموقع الجغرافي:

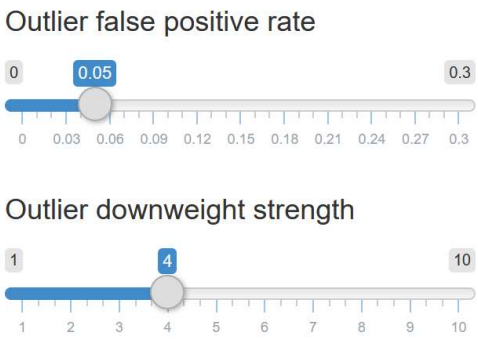
- Tweet: ويشمل المعلومات الجغرافية الواردة في نص التغريدة أو، إن لم تكن متوفرة، المعلومات الجغرافية الواردة في التغريدات المكررة/ النص المقتبس، إن وجدت.
- User: وهي معلومات جغرافية تم الحصول عليها من موقع المستخدم. وبحسب ترتيب أولوياتها، هذا هو موقع المستخدم عند نشر التغريدة، أو موقع واجهة برمجة تطبيقات المستخدم أو "البلد" المخصص في الملف الشخصي العام إذا لم يكن أي منها متاحاً.
- Both: تكون المعلومات الجغرافية المستخدمة للتغريدة، وبحسب ترتيب أولوياتها، الموقع الوراد في نص التغريدة، في حال لم يكن موقع المستخدم متاحاً.

Signal detection false positive rate



استخدم شريط التمرير لتستعرض الاختلافات الواقعة في الإشارات التي أُستدرجت عند تغيير مَعلمة ألفا للمعدل الإيجابي الكاذب. ولاحظ أن هذا لن يغير المعدل الإيجابي الكاذب لإشعارات التنبيه المرسلة عبر البريد الإلكتروني. فهذه مجرد أداة لأغراض المستخدم لاستكشاف هذه المَعلمة. والتعيين الافتراضي هو 0.025. ومن شأن تعيين معدل إيجابي كاذب أعلى أن يزيد من حساسية الإشارات المكتشفة وربما يزيد من عددها والعكس صحيح.

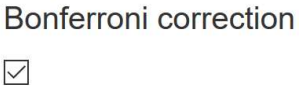
Outlier false positive rate وoutlier downweight strength



outlier false positive rate يرتبط بالمعدل الإيجابي الكاذب لتحديد ماهية القيمة المتطرفة عند تخفيض ترجيح القيمة المتطرفة/الإشارات السابقة. وكلما انخفضت القيمة، قلّ عدد القيم المتطرفة السابقة المحتمل إدراجها. ومن المحتمل أن تتضمن القيمة الأعلى مزيداً من القيم المتطرفة السابقة.

وتحدد outlier downweight strength المقدار الذي سيُخفّض عليه ثقل القيمة المتطرفة. وبالمقابل كلما زادت القيمة ارتفع مقدار الحد من الأثر. لمزيد من المعلومات، انظر الملحق1.

تصحيح بونفيروني



يُعيَّن Bonferroni correction افتراضياً. ويعمل بدوره على تفسير الكشف عن الإشارات الإيجابية الكاذبة عبر اختبارات مقارنة متعددة. فبالنسبة للكشف عن إشارة مخصصة لبلد بعينها، تُقسّم مَعلمة ألفا على عدد البلدان. أما بالنسبة لاكتشاف الإشارات الخاصة بالقارة، فتُقسّم قيمة ألفا على عدد القارات.

وإذا لم تكن ترغب في استخدام هذا التصحيح، فيمكنك إلغاء تعيينه.

Days in baseline

عدد الأيام افتراضياً المُعين في الأساس هو 7 أيام. ويمكن للمستخدم استكشاف تأثير متمثل بوجود أيام مختلفة في الأساس. ويخدم هذا فقط البيانات المرئية، أما التغييرات المدخلة على إشعارات التنبيه المرسلة عبر البريد الإلكتروني فيجب إجراؤها في صفحة التكوين.

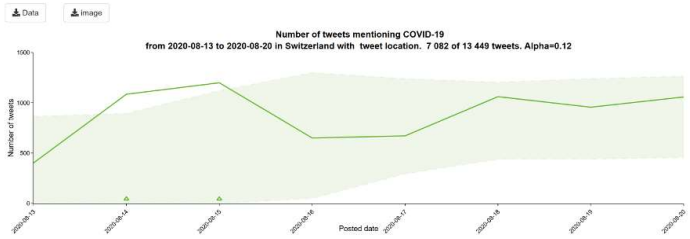
Same weekday baseline

قد يكون هناك "تأثير يوم من أيام الأسبوع"، حيث يشهد يوم معين من أيام الأسبوع حركة أكبر في نشر التغريدات مقارنة بباقي الأيام (على سبيل المثال، يوم الاثنين). يمكنك أيضاً أن تختار حساب الأساس ليس على شكل أيام متتالية، ولكن في

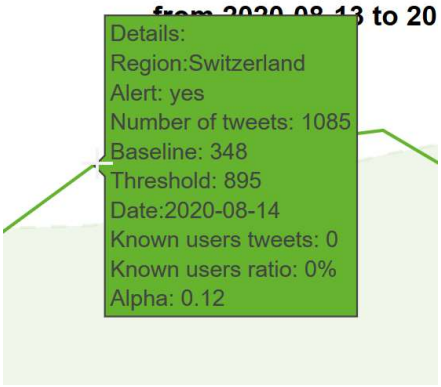
الأيام N الماضية التي تتوافق مع النافذة ذاتها الممتدة ل 24 ساعة N يوم إلى الوراء. وبهذه الطريقة إذا كان N = 7، يُحسب الأساس باستخدام "الأيام" من -7 و-14 و-21 و-28 و-35 و-42 و-49 و-56 (إذا كان "اليوم" الحالي يساوي صفر).

المخطط الزمني

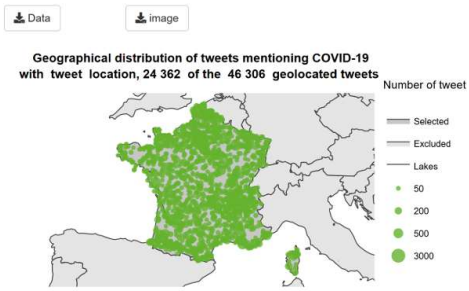
الرسم البياني للمخطط الزمني هو عبارة عن سلسلة زمنية، تمنحك إمكانية رؤية عدد التغريدات التي تتناول موضوعًا معينًا والوحدة الجغرافية وفترة البحث والتقصي. ويشار إلى الإشارات بوصفها مثلثات على الرسم البياني، مع مَعلمة ألفا وأيام الأساس كما هو محدد في المرشحات. ويشار إلى المنطقة الواقعة دون العتبة باللون الأخضر المظلل. لاحظ أن الإشارات مرتبطة باختيار ألفا والأيام في الأساس في المرشحات على لوحة التحكم، بدلاً من ما يجري استخدامه في إشعارات التنبيه المرسلة بالبريد الإلكتروني. وبهذه الطريقة يمكنك تقصي التأثير الناتج عن تغيير هذه المَعلمات وتكييف إعدادات رسائل التنبيه الإلكترونية إذا لزم الأمر.



وإذا حركت مؤشر الفأرة فوق الرسم البياني، ستظهر لك معلومات إضافية عن الدولة والتاريخ وعدد التغريدات وعدد التغريدات من قائمة المستخدمين المعروفين، ونسبة المستخدمين المعروفين وما يقابلها من مستخدمين غير معروفين، وما إذا كان عدد التغريدات مرتبطاً بإشارة والحد الأدنى ومَعلمة ألفا.

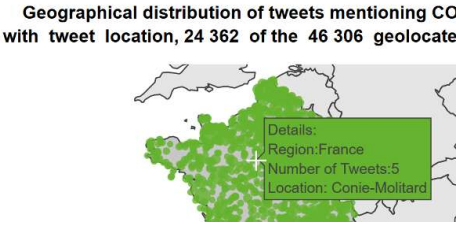


الخريطة



تعرض خريطة الرموز المتناسبة للتغريدات حسب البلد والموضوع على مدى فترة البحث والتقصي. وكلما كبرت الدائرة، تعاظم عدد التغريدات.

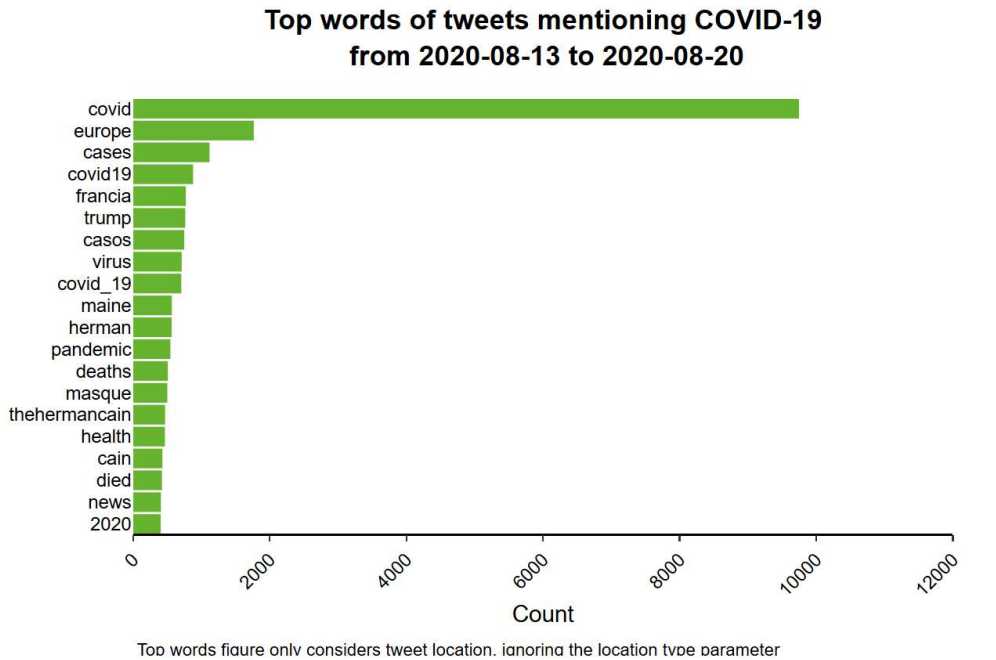
تستند المعلومات الجغرافية للخريطة إلى اختيار المرشحات: البلد/المنطقة/المنطقة دون الإقليمية ونوع الموقع (تغريدة، مستخدم أو كلاهما).



وعندما تُحرك مؤشر الفأرة فوق الخريطة، تعرض الخريطة معلومات حول عدد التغريدات ومسميات الوحدات الجغرافية الكامنة وراء الدوائر على الخريطة.

وعند اختيارك دولة واحدة، تعرض الرموز التوزيع الجغرافي للتغريدات على المستوى المحلي. وعند اختيارك دولتين أو أكثر أو أي كيان جغرافي آخر (مثل المناطق أو القارات)، تعرض الرموز التوزيع الجغرافي للتغريدات على المستوى الوطني. لاحظ أنه إذا احتوت تغريدة على وسم جغرافي على مستوى الدولة (مثل فرنسا)، فلن تُعرض عند اختيار تلك الدولة فقط نظراً لعدم توفر سمات جغرافية دون وطنية.

أكثر الكلمات تواتراً في التغريدات



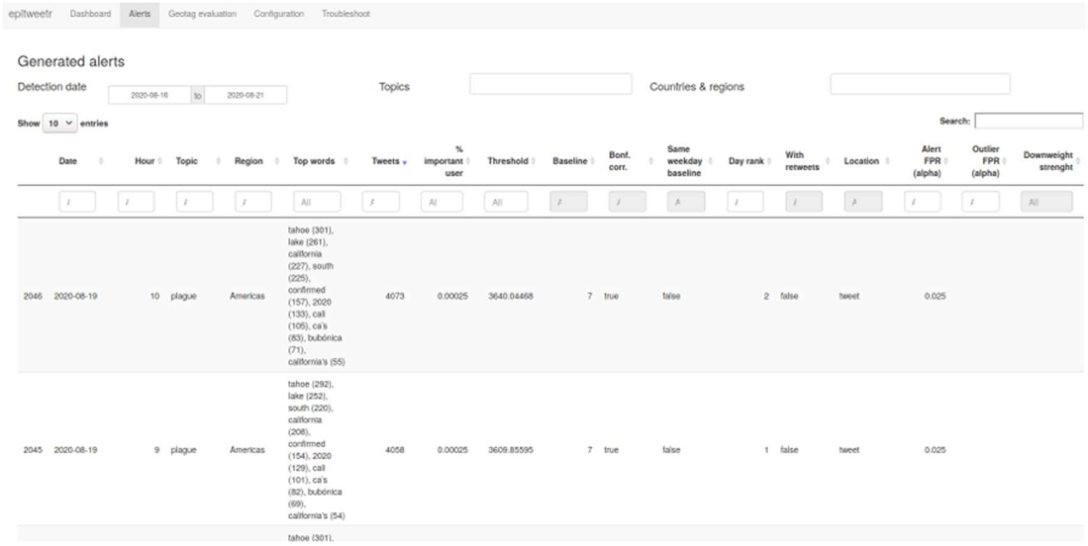
تخضع التغريدات للتحليل على الصعيد العالمي لنخرج بقائمة بأكثر 500 كلمة تواتراً لكل من اللغة واليوم والموضوع في مجموعات. ثم تُصنّف هذه الكلمات البالغ عددها 500 كلمة حسب البلد.

ويعرض هذا الرسم البياني بعد ذلك قائمة بأهم كلمات التغريدات حسب الموضوع طوال مدة دراسة الوحدات الجغرافية المعينة، ووفقاً لعوامل التصفية المطبّقة على التغريدات/التغريدات المكررة.

لاحظ كيف ترتبط الكلمات الأكثر تواتراً الواردة في التغريدات دائماً بـ "موقع التغريدات" ولا تتأثر بعامل التصفية المطبّق لاختيار الموقع (المستخدم أو موقع التغريدة)، على عكس البيانات المرئية الأخرى.

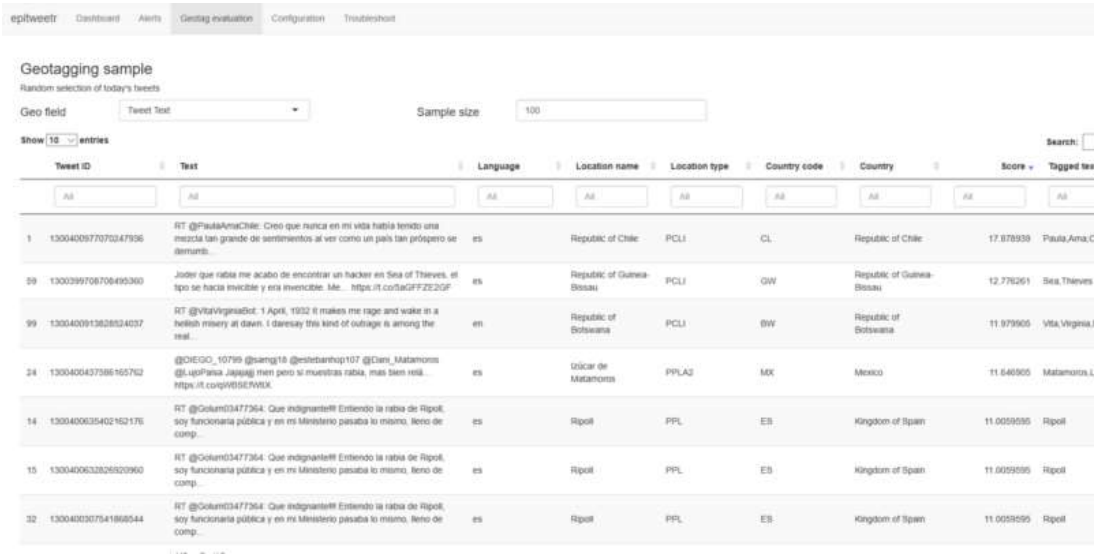
صفحة alerts

تقدم صفحة alerts ملخصاً عن الإشارات المكتشفة على مدى فترة الدراسة المحددة. ويتضمن هذا الملخص تاريخ الإشارة وساعة ظهورها وموضوعها ووحدتها الجغرافية، إلى جانب أهم الكلمات التي تتضمنها، وعدد التغريدات، وعدد تغريدات المستخدمين المهمين والعتبة. كما تُطلعك على عدة إعدادات كانت قد حُددت في صفحة التكوين، وأُستخدمت في آلية الكشف عن الإشارات. وهذا هو أيضاً الناتج الذي تتلقاه في إشعارات التنبيه المرسلة عبر البريد الإلكتروني.



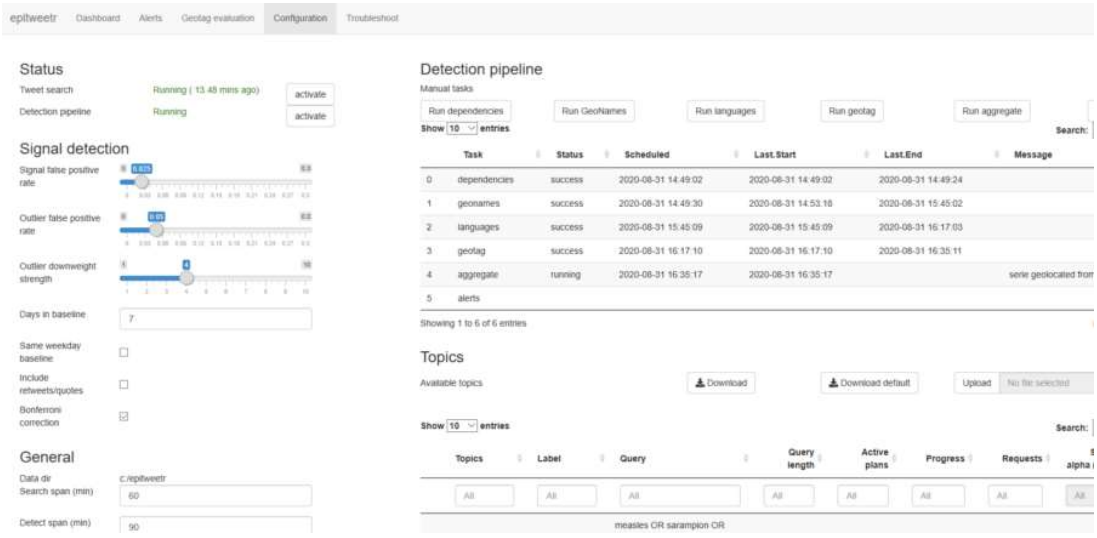
صفحة geotag evaluation

تدعم هذه الصفحة المستخدِم في وضع عتبة لعامل تحديد الموقع الجغرافي في صفحة التكوين. ويمكن للمستخدم اختيار مجال التغريدات المراد اختبارها وعدد التغريدات التي يود إلحاقها بعينة الاختبار. وتعتبر هذه الصفحة صفحة بيانات مرئية فقط ولا تسمح بإجراء أي تغييرات في الموقع الجغرافي الذي تعقبته حزمة epitweetr.



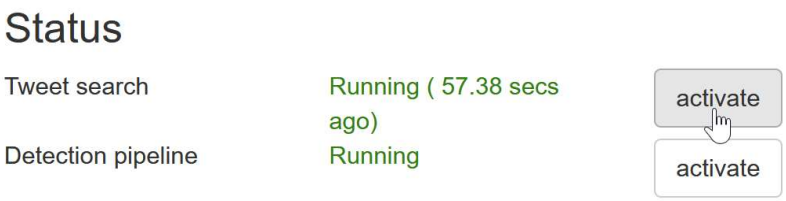
صفحة configuration

تتيح لك هذه الصفحة تغيير إعدادات الأداة، والتحقق من حالة العمليات/الأنابيب المختلفة للأداة كما يمكنك إضافة موضوعات وحذفها وتعديلها بالإضافة للطلبات المرتبطة بها وتعديل اللغات المعنية بتحديد الموقع الجغرافي وقائمة "المستخدمين المهمين" والمشتركين في إشعارات التنبيه عبر البريد الإلكتروني. تذكّر أن تنقر على زر "Update Properties" في نهاية قسم ""General، عند تجري تعديلاً على أي خيار في قسم "اكتشاف الإشارة" أو "General". وتورد الأقسام التالية شرحاً لصفحة configuration بمزيد من التفصيل.



Status

يتيح لك قسم status إمكانية إجراء تقييماً سريعاً لآخر نقطة زمنية و/أو حالة عمليات جمع التغريدات (Tweet Search) وتحديد الموقع الجغرافي والتجميع واكتشاف الإشارات (Detection pipeline).



ففي قسم status، يمكنك معرفة ما إذا كان كل من خط أنابيب البحث وخط أنابيب الكشف مفعّل أم لا. وإذا كنت تعمل على نظام تشغيل Windows، فانقر على خيار "activate" لتسجيل هذه العمليات كمهام مجدولة ولتشغيلها يدوياً من مجدول مهام نظام تشغيل Windows.

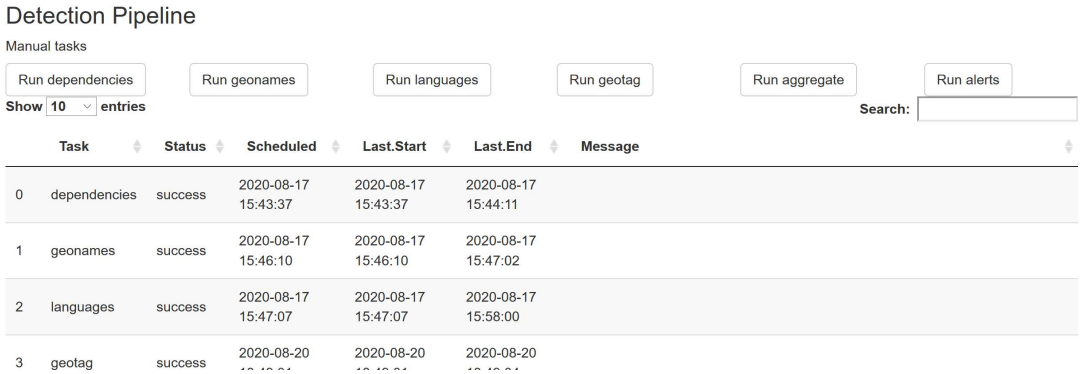
أنابيب خط الكشف

في المهام اليدوية في ""Detection pipeline، يتعين عليك عند شروعك لأول مرة في استخدام epitweetr، أن تُشغّل مهام التبعيات والمسميات الجغرافية واللغات يدوياً بالنقر على خيارات"Run dependencies" و"Run geonames" و""Run languages، وتعيد الكرّة فقط عند تنزيل إصدارات جديدة. وتتصل المسميات الجغرافية واللغات بالمواقع الجغرافية ونماذج اللغة المستخدمة عبر epitweetr. فإذا كنت ترغب في تحديثها (وهو أمر ليس عليك أن تجريه بصورة دورية، بل على نحو سنوي أو أبعد من ذلك)، انقر على زر "Run".

ويمكن تفعيل خيارات "Run geotag" و"Run aggregate" و"Run alerts" للشروع بهذه المهام في حال ظهور أي خطأ أو مشكلة. كما يمكنك التحقق من حالتها في جدول "Detection Pipeline".

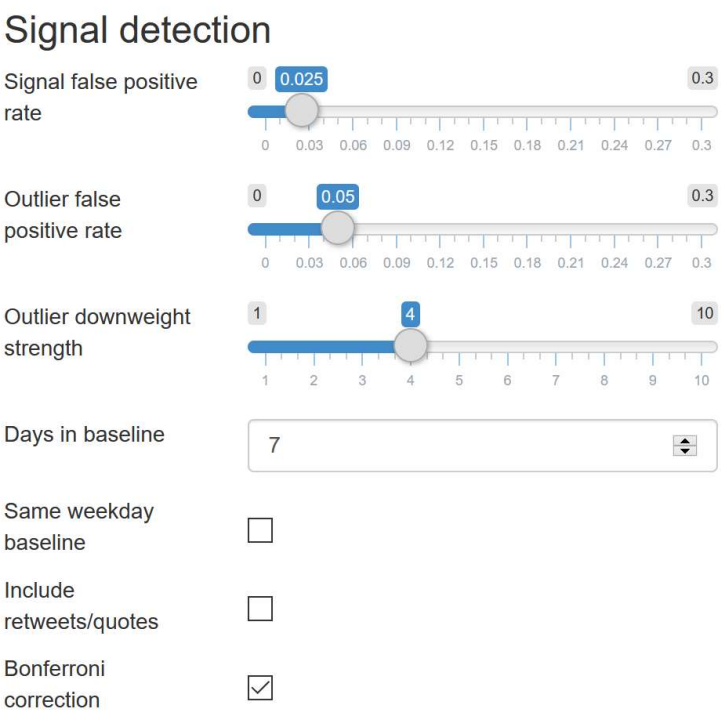
ويزودك خط أنابيب الكشف بالمزيد من المعلومات حول حالة عمليات `epitweetr`. ويفيدك هذا في استكشاف أي مشكلة قد تظهر ولتراقب تقدم العمليات. كما يحتوي على المهام الخمس التي تعمل في الأنظمة الخلفية. فمهام المسميات الجغرافية واللغات هي التي تتولى تنزيل النسخ المحلية منها وتحديثها. ولن تصبح هذه المهام قيد التشغيل ما لم تُضيف لغة أو تُحدّث عامل GeoNames. وبشكل عام ستجد تواريخ البدء والانتهاء أقدم بكثير من تواريخ تحديد الوسم الجغرافي والتجميع وإشعارات التنبيه.

ويجب أن تكون تواريخ العلامات الجغرافية والتجميع وإشعارات التنبيه أحدث إذا كانت خطوط أنابيب البحث والكشف في حالة نشطة وقيد التشغيل. وقد جُدولت مهامها وفقاً لعامل مدى الكشف. يمكن أن تتضمن status حالات running أو scheduled أو pending أو failed أو aborted (في حال فشلت أكثر من ثلاث مرات).



اكتشاف الإشارة

في قسم detection section الموجود في صفحة configuration، يمكنك تعيين مَعلمة ألفا على signal false positive، الأمر الذي سيزيد (إذا كانت أكبر) من مدة الكشف (وينجم عنه اكتشاف المزيد من الإشارات)، أو تُخفّض من (إذا كانت أصغر) الفاصل الزمني للكشف(وعليه اكتشاف أقل للإشارات).



outlier false positive rate يرتبط بالمعدل الإيجابي الكاذب لتحديد ماهية القيمة المتطرفة عند تخفيض ترجيح القيمة المتطرفة/الإشارات السابقة. وكلما انخفضت القيمة، قلّ عدد القيم المتطرفة السابقة المحتمل إدراجها. ومن المحتمل أن تتضمن القيمة الأعلى مزيداً من القيم المتطرفة السابقة.

وتحدد outlier downweight strength المقدار الذي سيُخفّض عليه ثقل القيمة المتطرفة. وبالمقابل كلما زادت القيمة ارتفع مقدار الحد من الأثر. لمزيد من المعلومات، انظر الملحق1.

تُعين حزمةepitweetr الحد الأدنى لتحديد ما إذا كان العدد الحالي للتغريدات في إطار 24 ساعة يتجاوز ما هو متوقع (راجع قسم "كيف يعمل؟ البنية العامة الكامنة وراء تصميم *epitweetr*< كشف الإشارات"). وتعتمد في تعيينها للحد الأدنى على القيمة الافتراضية للأيام السبعة الماضية. ويمكنك تغيير عدد الأيام من حقل "الأيام الافتراضية في الأساس".

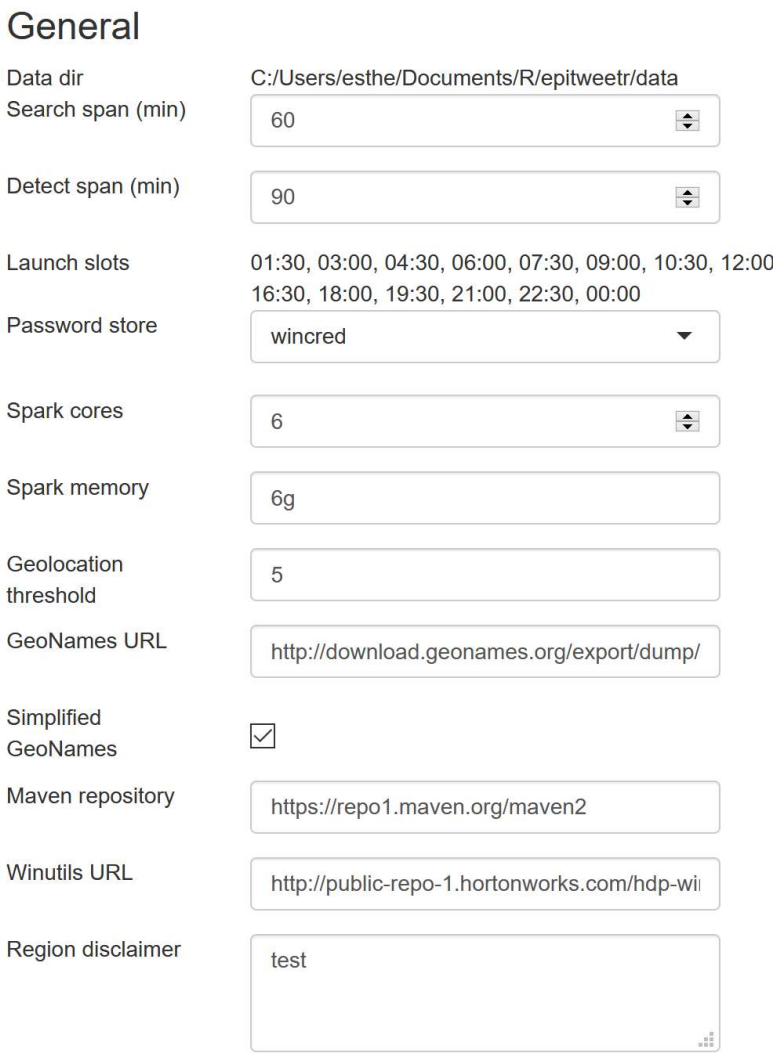
كما يمكنك تغيير الإعداد الافتراضي لاستخدام الأيام السبعة الماضية لحساب الأساس إلى الأيام السبعة السابقة ذاتها من الأسبوع، وذلك لتجنب عامل "تأثير يوم من أيام الأسبوع" (على سبيل المثال، قد تظهر حركة أوسع من التغريدات حول موضوع ما في يوم اثنين، الأمر الذي قد يؤثر على اكتشاف الإشارة).

ويمكنك أيضاً تحديد ما إذا كان الكشف عن الإشارة ينجح فقط من خلال نص تغريدة، أو يتضمن التغريدة المكررة/المقتبسة (ضع علامة في خانة اختيار"Default with retweets/quotes").

وتأخذ الخانة الأخيرة المفعلّة "Default with Bonferroni correction" في اعتبارها اختبارات المقارنة المتعددة، والتي يمكن أن تؤدي إلى نتائج إيجابية كاذبة. وفي حال حددت هذه الخانة، فستُقسَّم مَعلمة ألفا على عدد المواقع الجغرافية التي يُنفذ فيها الكشف عن الإشارة بغية كشف الإشارة. فعلى سبيل المثال، على مستوى الدولة، تُقسّم مَعلمة ألفا على إجمالي عدد البلدان. وعلى مستوى القارة، تُقسّم معلمة ألفا على إجمالي عدد القارات.

وعند إجرائك لأي تغيير في قسم "اكتشاف الإشارة"، لا تنسَ النقر على زر "Update Properties" في نهاية قسم "General".

General



- في **Data directory**، يمكنك الاطلاع على الدليل الذي تستخدمه حزمة `epitweetr` لتخزين التغريدة والبيانات المرتبطة التي جُمعت. وهو الدليل ذاته الذي تستخدمه لوحة التحكم لتحصل على مجموعات البيانات بغية عرض البيانات المرئية. وعليك تعيين هذا المجلد عند البدء في تشغيل `epitweetr` أو تعيين متغير البيئة على "EPI_HOME".

- يتعلق **Search span** بمدة تنفيذ خطة البحث. فالوقت الافتراضي معين عند 60 دقيقة. وتتحكم هذه القيمة في حجم نافذة البحث الخاصة بالتغريدات. فإذا خفضت من هذه القيمة، فستتلقى تغريدات في وقت أسرع، ولكنك قد "تهدر" الطلبات المتعلقة بالموضوعات التي نُشر بشأنها عدد قليل جداً من التغريدات. أما إذا عملت على زيادة القيمة، فستمضي وقتاً أطول للحصول على التغريدات، لكن في المقابل ستستدرج المزيد من طلبات التغريدات الشائعة الأمر الذي سيرفع من فرص شموليتها. ويمكنك أن ترى عندما تفشل في جمع التغريدات على صفحة تكوين تطبيق Shiny جراء وجود أكثر من خطة واحدة نشطة بشأن بعض الموضوعات.

- يتعلق **Detect span** بعدد المرات التي يُنفذ فيها خط أنابيب الكشف عملياته (تحديد الوسم الجغرافي والتجميع وإشعارات التنبيه عن الكشف). فالوقت الافتراضي معين عند 90 دقيقة. وتُرسل إشعارات التنبيه عبر البريد

الإلكتروني في نهاية حلقة الكشف. تُعالج هذه القيمة على أساس أنها حد أدنى، وقد تستهلك حلقة الكشف مزيداً من الوقت لإنهاء العملية بالنظر إلى حجم التغريدات ومواصفات نظام التشغيل الذي تعمل عليه.

- تعمد **Launch slots** إلى مباعدة مواعيد عمليات خط أنابيب الكشف وفقاً لـ ""Detect span، حيث تبدأ العملية الأولى عند منتصف الليل. ويمكن استخدام هذه القيم في ملف المشتركين في صفحة التكوين.

- ولتتجنب تخزين بيانات اعتماد تويتر في ملفات عادية، يستخدم `epitweetr` خاصية تخزين كلمات مرور لدخول النظام، والتي تُحفظ في **Password store**. كما يمكنك اختيار الآلية التي تناسب البيئة التي يعمل فيها`epitweetr`، وذلك بالاعتماد على نظام التشغيل الذي تعمل عليه. للاطلاع على مزيد من المعلومات حول آلية التنفيذ، تفضل بزيارة الوصلة *https://CRAN.R-project.org/package=keyring*
  - wincred: (يتوافق مع نظام تشغيل Windows فقط) ويستخدم مدير بيانات اعتماد نظام تشغيل Windows.
  - macos: (يتوافق مع نظام تشغيل Mac فقط) ويستخدم خدمات سلسلة مفاتيح نظام تشغيل Mac
  - file: يستخدم الملفات المشفّرة المحمية بكلمة مرور
  - secret service: (يتوافق مع نظام Linux فقط) ويستخدم خدمة Linux السرية
  - environment: يستخدم متغيرات البيئة (يلزمك تعيين إعداد إضافي، انظر *https://CRAN.R-project.org/package=keyring*)

- **Spark cores** و**spark memory**: يمكنك أيضاً تخصيص ذاكرة حزمة `epitweetr` عبر تحديد وحدة المعالجة المركزية (Spark cores) وذاكرة الوصول العشوائي (Spark Memory) في قسم "general". فالتعيين الافتراضي هو 6 أنوية و6 غيغابايت لذاكرة الوصول العشوائي. وسيعتمد هذا على سعة وحدة المعالجة المركزية CPU وذاكرة الوصول العشوائي RAM الموجودة في جهازك والتي ينبغي لها أن تكون مساوية لها أو أقل منها.

- **Geolocation threshold:** أثناء عملية تحديد الموقع الجغرافي، تخضع مجموعات الكلمات للمعالجة وتُحدد التطابقات المحتملة مع المواقع الحالية وتُمنح درجة. وكلما سجلت درجات أعلى، ارتفعت احتمالية صواب الموقع الجغرافي المعين. ويجري تعيين الحد الأدنى في حزمة `epitweetr`، والتي بموجبها لا يمكن اعتبار أي مطابقات جيدة بصورة كافية لتحديد الموقع الجغرافي. ويتدرج المقياس من 1 إلى 10، وتُعين القيمة الافتراضية على درجة 5.

- **Geonames URL:** تجد عنوان URL المستخدم لتنزيل قاعدة بيانات GeoNames (المستخدمة لإنشاء المواقع) في قسم general. وفي حال تغيير عنوان URL هذا، يمكنك إجراء التعديل هنا.

- **Simplified geonames:** نظراً لأن GeoNames يتسم بكونه ملفاً ضخماً في حجمه، تُستخدم نسخة مبسطة منه بشكل افتراضي، متضمناً ذلك المواقع الجغرافية الموجودة وفقط تلك التي تُعرف بالكثافة السكانية. ويمكنك إلغاء تحديد هذا الخيار إذا كنت ترغب في استخدام قاعدة بيانات GeoNames بأكملها.

- **Maven repository:** وهو عنوان URL لمستودع مافين الذي سيُستخدم لتنزيل تبعيات JAR لحلقة الكشف، ويستهدف إطاري الحوسبة سبارك ولوسين بصورة أساسية.

- **Winutils URL:** هو عنوان URL يُستخدم لتنزيل winutils.exe. وهو عبارة عن متوافقة ثنائية لنظام Windows وتعد ضرورية لتشغيل سبارك داخل نظام تشغيل Windows. وإذا كنت لا ترغب في استخدام هذا الإصدار، فيمكنك تشكيل إصدار بنفسك عبر تنزيل Hadoop 2.8.4 أو إصدار أعلى منه وتجميعه على جهاز يعمل بنظام تشغيل Windows.

- **Region disclaimer:** إذا أردت إضافة عنصر إخلاء المسؤولية إلى الخريطة التي تستخدمها، فسيَضاف إلى عنصر تصدير صورة الخريطة في لوحة التحكم، وكذلك لعنصر تصدير الصورة بصيغة PDF في لوحة التحكم.

مصادقة تويتر

أنت أمام خيارين لتحقيق أغراض المصادقة على تجميع التغريدات، إما باستخدام Twitter account (باستخدام حزمة rtweet) أو باستخدام Twitter developer application. ويمكنك تحديد الخيار الذي ستعتمده في قسم مصادقة تويتر. راجع قسم "كيف يعمل؟ البنية العامة الكامنة وراء *epitweetr*> مجموعة التغريدات> مصادقة تويتر " لمزيد من التفاصيل حول كيفية إجراء مصادقة تويتر.

Email authentication (SMTP)

يتعين عليك في هذا القسم تخصيص تفاصيل تصديق البريد الإلكتروني (بروتوكول نقل البريد البسيط) للبريد الإلكتروني المعني بإرسال إشعارات التنبيه.

وعند التحقق من **Unsafe certificates**، فسيستخدم `epitweetr` خادم بروتوكول نقل البريد البسيط الخاص بك حتى في حال أرسل الخادم شهادة غير صالحة.

وعند إجرائك لأي تغيير في قسم " "general، لا تنسّ أن تنقر على زر " Update Properties".

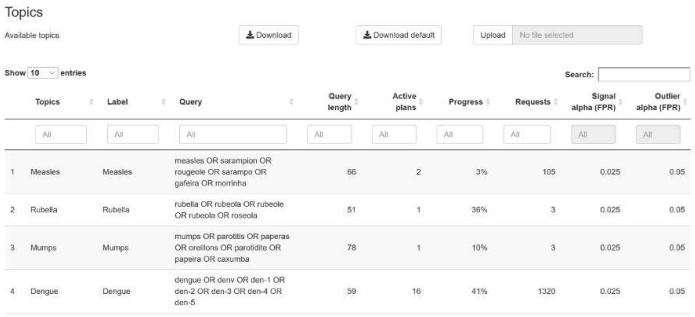
Topics

تحدد الموضوعات ما ينبغي لحزمة `epitweetr` أن تجمعه من تغريدات. وتنفذ ذلك عبر جدول بيانات بصيغة Excel يحتوي على الموضوعات والطلبات المرتبطة بها التي يستخدمها `epitweetr` للاستعلام عن واجهة برمجة التطبيقات الخاصة بموقع تويتر.

ويتكون الاستعلام من كلمات أساسية وعوامل تشغيل تُستخدم لمطابقة سمات التغريدات. راجع قسم "كيف يعمل؟ البنية العامة الكامنة وراء *epitweetr*> مجموعة التغريدات>موضوعات التغريدات المستهدفة وآلية الاستعلام" للاطلاع على مزيد من التفاصيل حول آلية الاستعلام

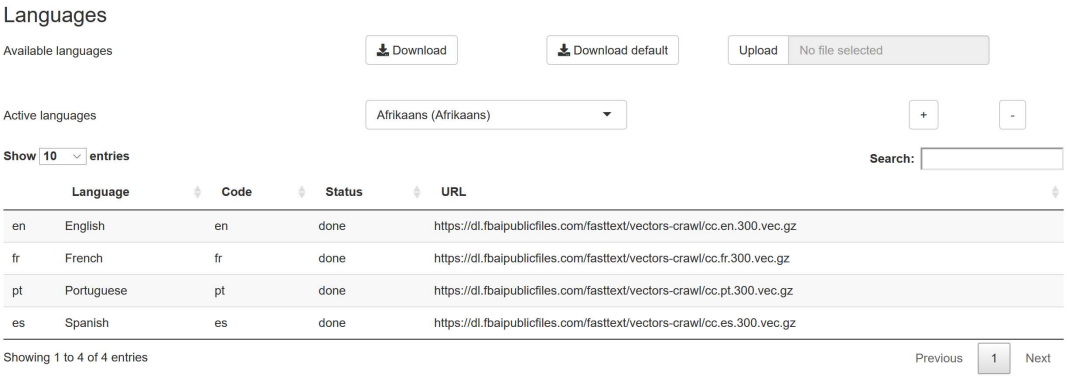
تشمل حزمة `epitweetr` قائمة موضوعات افتراضية بالصيغة التي استخدمها فريق المركز الأوروبي للوقاية من الأمراض ومكافحتها المعني باستخبارات الأوبئة في تاريخ إصدار جيل الحزمة (1 سبتمبر 2020). يمكنك تنزيل قائمة الموضوعات هذه وتحميل موضوعاتك الخاصة في قسم "الموضوعات المتاحة" في صفحة التكوين. راجع قسم "كيف يعمل؟ البنية العامة الكامنة وراء تصميم *epitweetr*، مجموعة التغريدات> موضوعات التغريدات المستهدفة وآلية الاستعلام" للاطلاع على مزيدٍ من التفاصيل حول كيفية تنظيم هيكلية قائمة الموضوعات.

وفي قسم الموضوعات في صفحة التكوين، يمكنك استعراض الموضوع والاستعلام المرتبط به وطول الاستعلام وعدد خطط البحث النشطة المرتبطة بالاستعلام. وفي حال كان هناك أكثر من خطة بحث مفعّلة، فهذا يعني أن `epitweetr` قد فشل في جمع كل التغريدات الممكن جمعها في الجلسة الأخيرة. أضف إلى ذلك، يمكنك رصد مدى التقدم وعدد الطلبات من خطة البحث الأخيرة.



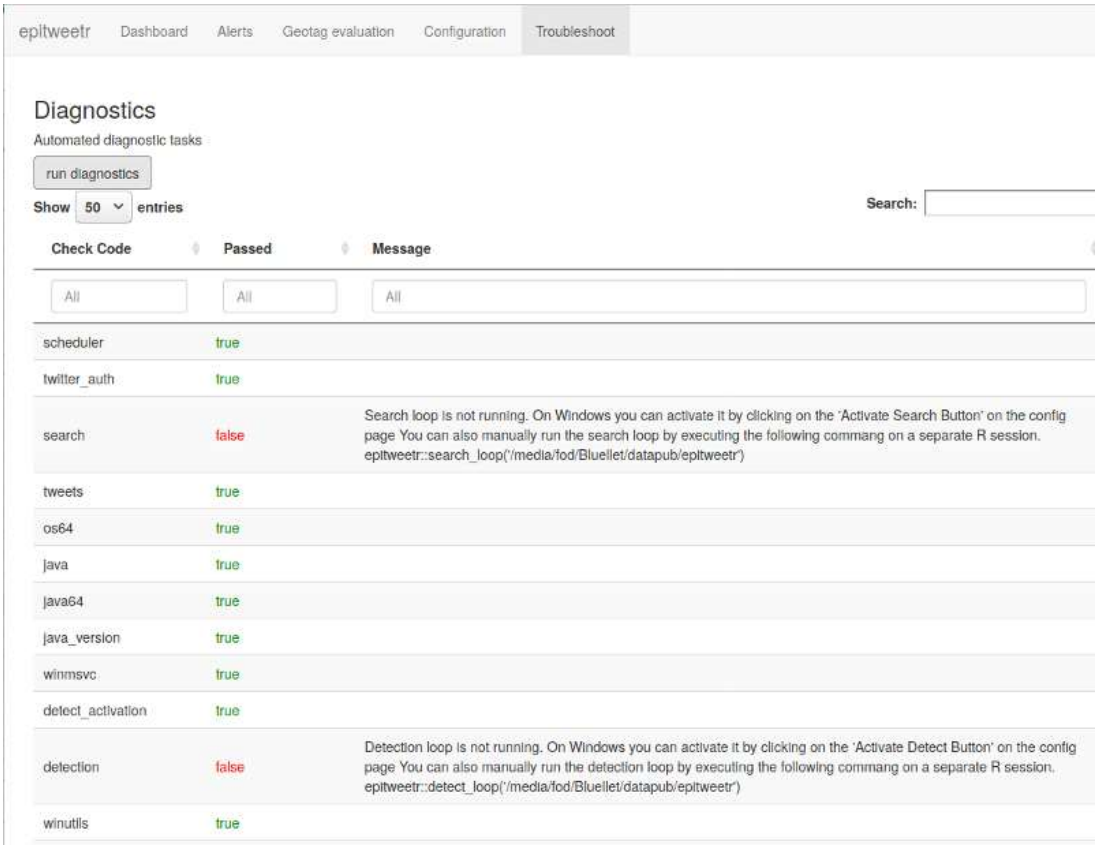
Languages

يفسح لك قسم اللغات إمكانية تحديد نماذج اللغة المستخدمة لتحديد ما يرد في النصوص أثناء عملية تحديد الموقع الجغرافي. واللغات الافتراضية المعينة هي الفرنسية والإنكليزية والبرتغالية والإسبانية. ويمكنك تنزيل نماذج اللغات وتحميلها من قسم "اللغات المتاحة" وإضافة وحذف اللغات التي ترغب باستخدامها عبر حزمة `epitweetr` من قسم "اللغات النشطة". نرجو أن تضع في اعتبارك التكلفة الحاسوبية عند إضافة عدد لا طائل منه من نماذج اللغات، وأن تراعي سعة جهازك في ذلك.



صفحة troubleshoot

تضع هذه الصفحة بين يدي حزمة `epitweetr`قائمة بخطوات تُحقق فحصاً تلقائياً وتلميحات يمكن الاستفادة منها لاستخداماته أثناء أداء وظائفه. انقر على خيار "Run diagnostics" لاستعراض قائمة بخطوات الفحص، سواء اجتاز عملية التحقق ("true") أم لم يجتازه ("false")، وقائمة بتلميحات في حالة عدم اجتيازه الفحص. تفضل بالاطلاع على مزيد من المعلومات المفصلّة في المرفق II من وثيقة المستخدم.



خاصية تنزيل المخرجات من واجهة المستخدم التفاعلية (Shiny app)

تتوفر خاصية تنزيل كافة البيانات المرئية من لوحة تحكم تطبيق Shiny app بصيغة صورة، بمجرد نقرك على زر"image button". ملف png هو ملف تنسيق رسومي لشبكات المحمول وهو صيغة ملف متعدد الاستخدامات للصور التي لا تحتاج إلى دقة عالية جدًا (مثل رسومات الطباعة الاحترافية).

لاحظ أن تنسيق png غير مدعوم في متصفح Internet Explorer (ولكن يمكنك تنزيل ملف svg عوضاً عنه).

ويمكنك أيضاً تنزيل بيانات لكل المرئيات بالنقر على زر البيانات. وسيمنحك هذا الخيار ملف بصيغة csv يحتوي على البيانات الأساسية التي يمكنك استخدامها لإجراء مزيد من التحليل أو لإنشاء رسوم بيانية خاصة بك.

وعوضاً عن ذلك، يمكنك استخدام النقر على زر PDF أو Md أسفل المرشحات لتنزيل ملف بصيغة PDF أو ملف HTML من لوحة التحكم. لاحظ أنك بحاجة لتثبيت أدوات MiKTeX أو TinyTeX لتستلم هذه الملفات بيسر وسهولة.

المرفق 1: الحد من ثقل الإشارات السابقة

مقدمة

نتناول في هذا الملحق طريقة الحد من الأثر طُورت كجزء من خوارزمية `ears` المستخدمة في حزمة `epitweetr` والتي عرّجنا عليها في أقسام سابقة.

لنفترض أن الدالة $\mathbf{y}$ تشير إلى متجه القيم التاريخية البالغ طوله n. وجزء من حساب فاصل التنبؤ في وقت 0 هو حساب المتوسط والانحراف المعياري لهذه القيم التاريخية، أي

$$\overline{y}_0 = \frac{1}{n}\sum_{t=-n}^{-1} y_t \quad \text{و} \quad s_0^2 = \frac{1}{n-1}\sum_{t=-n}^{-1}(y_t - \overline{y}_0)^2$$

ثم تُحسب عتبة فاصل التنبؤ $(1-\alpha)\times 100\%$ أحادي الجانب للرصد y_0 بحسب $y_t \overset{\text{iid}}{\sim} N(\mu, \sigma^2), t = -n, \ldots, 0$ كالتالي

$$U_0 = \overline{y}_0 + t_{1-\alpha}(n-1) \times s_0 \times \sqrt{1 + \frac{1}{n}},$$

حيث يشير $t_{1-\alpha}(n-1)$ إلى $1-\alpha$ نقاط التجزئة لمقدار التوزيع t مع $n-1$ درجة من الحرية. يتوافق حساب عتبة هذه الخوارزمية مع الحساب الإحصائي السليم للعتبة (كما طرحه أليفيوس وهوله 2017).

وتتطلب الخوارزمية أعلاه امتداداً هو معالجة الإشارات السابقة بالقيم التاريخية. وقد عولجت هذه المسألة في إطار معادلة انحدار كواساسي-بواسون الخاص بطرح فارينغتون وآخرون لعام (1996) من خلال إجراء ملاءمة GLM أولاً ثم إعادة تركيب GLM بأوزان تستند إلى معادلة انحدار أنسكومب. نتبع هنا المبدأ العام ذاته، لكننا نكيّفه مع الاستجابة الغوسية المستخدمة في خوارزمية EARS والرواسب المقابلة من النموذج الخطي.

خوارزمية EARS كنموذج خطي

نلاحظ أولاً أنه التقدير أعلاه لـ μ و σ^2 وخلال $\overline{y}_0$ و s_0^2 في وقت 0 يمكن تضمينه داخل نموذج الانحدار الخطي، أي بالنسبة $i = 1, \ldots, n$ نضع النموذج

$$y_i = \mu + \epsilon_i, \quad \text{حيث} \quad \epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

لاحظ أننا، ولتحقيق غرض التوافق مع العرض القياسي في نظرية النموذج الخطي، قمنا بفهرسة قيم s.t. yy_{-n} يتوافق مع y_1 و y_{-1} يتوافق مع y_n. من حيث المصفوفة، دع الدالة $\mathbf{y} = (y_1, \ldots, y_n)'$ ولأغراض نموذج اعتراض فقط، تكون مصفوفة التصميم عبارة عن $\mathbf{X} = (1, \ldots, 1)'$ ودالتها $k = 1$. وبالتالي من منظور نهج المربعات الصغرى المعتادة القياسية:

$$\hat{\mu} \quad = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

التي تتوافق مع $\overline{y}_0$. علاوة على ذلك، دع الرواسب الخام تُعرّف على أنها $e_i = y_i - \hat{\mu}$ من أجل $i = 1, \ldots, n$ وتُدل عليها $\mathbf{e} = (e_1, \ldots, e_n)'$ وهو المتجه المقابل للرواسب. ثم

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{P}\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{y}$$

حيث $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ هي ما يسمى بمصفوفة القبعة المعروفة بالنمذجة الخطية. وباستخدام هذا الترميز يمكننا كتابة التقدير σ^2 كما في طرح شاترجي وهادي (1988):

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{\mathbf{y}'(\mathbf{I}-\mathbf{P})\mathbf{y}}{n-k} = \frac{1}{n-1}\sum_{t=-7}^{1}(y_t - \hat{\mu})^2,$$

والذي يتوافق مع العبارة الحسابية المستخدمة أعلاه لـ s_0^2.

تخفيض الثقل

نجري الآن حساب ما يسمى في الإحصائيات المتبقي الطالب الخارجي (7عر 1988)

$$r_i^* = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-p_{ii}}}, \quad i = 1, \dots, n,$$

حيث p_{ii} العنصر القطري i'th لمصفوفة القبعة **P** من النموذج الخطي المقابل المستخدم أعلاه. أضف إلى ذلك،

$$\hat{\sigma}^2_{(i)} = \frac{\mathbf{y}_{(i)}'(\mathbf{I} - \mathbf{P}_{(i)})\mathbf{y}_{(i)}}{n-k-1}$$

هو تقدير التباين الذي نتج من الانحدار الخطي، حيث تُزال الملاحظة الأولى i'th. تنص نظرية النمذجة الخطية (شاترجي وهادي 1988) الآن على ذلك

$$r_i^* \overset{\text{متطابق}}{\sim} t(n-k-1).$$

لاحظ أن القيم المتبقية موزعة بشكل متماثل فقط، لأنها ليست مستقلة (انظر القسم 4.2.1. من طرح شاترجي وهادي(1988) لمزيدٍ من التفاصيل). بيد أنه يسمح لنا الشكل التوزيعي أعلاه بتقييم كل قيمة تاريخية، إذا كان من الممكن اعتبارها قيمة خارجية. لهذا الغرض، حدد $r_{\text{الحد الأدنى}}$كنقاط تجزئة$\alpha_{\text{المتطرفة}} - 1$ لتوزيع t مع $n-k-1$درجة من الحرية. القيمة التاريخية هي قيمة متطرفة (أحد التفسيرات المحتملة لذلك هو أنها تنشأ نتيجة زيادة حقيقية في التغريدات، على سبيل المثال حالة تفشي الفيروس)، إذا$r_{\text{الحد الأدنى}} < r_i^*$. سنستخدمها لصياغة مخططات ترجيح للقيم التاريخية:

تخفيض ثقل -القيم المتطرفة:

$$w_i^{(\text{dw})} = \begin{cases} 1 & \text{if } r_i^* < r_{\text{الحد الأدنى}} \\ \left(\frac{r_{\text{الحد الأدنى}}}{(r_i^*)}\right)^k & \text{وإلا} \end{cases}$$
$$= \min\left\{1. \left(\frac{r_{\text{الحد الأدنى}}}{r_i^*}\right)^k\right\},$$

حيث تكون مَعلمة الاضمحلال $k > 0$ تساوي كمية معروفة. في الطرح الأصلي لفارينغتون وآخرون عام (1996) استخدمت خوارزمية$k = 2$. أضف إلى ذلك، أُستخدمت قيمة حدية قدرها 1. في وقت لاحق في طرح نفيلي وآخرون لعام (2013)، ومع ذلك، جرت التوصية بقيمة حدية قدرها 2.58. ملاحظة: كلا القيمتين مخصصتان لرواسب أنسكومب المعيارية، والتي تتبع التوزيع الطبيعي القياسي. إذا أخذنا الكميات المقابلة لتوزيع t مع 6 درجات من الحرية، فستخرج القيم كالتالي 1.09 و3.72. لاحظ أيضاً أن المصطلح $(r_{\text{الحد الأدنى}}/r_i^*)^k$ هو تعديل طفيف لما طرحه فارينغتون وآخرون لعام (1996)، والذين يستخدمون $1/(r_i^*)^2$ عوضاً عنه. تتمثل ميزة مقترحنا في أنه يضمن معالجة سلسة للقيم الحدية إذا لم تساوي العتبة 1. وقد يكون من المفيد التفكير في قوة أعلى من 2 لضمان الحد من أثر القيم الإجمالية المتطرفة. كما أن القيمة الافتراضية الحالية لمعامل الاضمحلال فيepitweetr تساوي 4.

وأخيراً كما طرح فارينغتون وآخرون في عام (1996)، قمنا ضبطنا صيغة الأوزان بحيث ينتج عنها مجموع وقدره n

$$w_i^* = n \times \frac{w_i}{\sum_{i=1}^n w_i}$$

ثم أعدنا تركيب النموذج الخطي بهذه الأوزان. ولتحقيق هذا الغرض، حدد مصفوفة الوزن على أنها= **W** $\text{diag}(w_1^*, \dots, w_n^*)$. يمكننا استخدام نهج المربعات الصغرى المرجح لاحاً لإيجاد

$$\hat{\mu}_W \quad = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y} = \frac{1}{n}\sum_{i=1}^{n} w_i^* \, y_i,$$

حيث تكون علامة التساوي الثانية لأن $(\mathbf{X}'\mathbf{W}\mathbf{X}) = \sum_{i=1}^{n} w_i = n$ و $\mathbf{X}'\mathbf{W}\mathbf{y} = \sum_{i=1}^{n} w_i^* \, y_i$. أضف إلى ذلك،

$$s_W^2 = \frac{\mathbf{y}'(\mathbf{I} - \mathbf{P}_W)\mathbf{y}}{n-k} = \frac{\sum_{i=1}^{n} w_i^* \, (y_i - \mu_W)^2}{n-1},$$

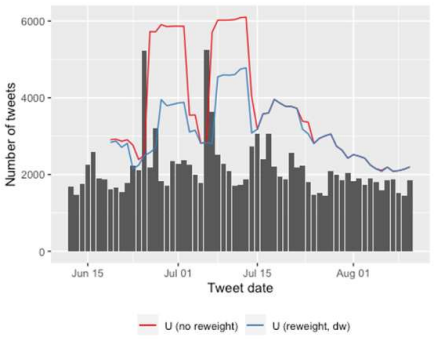
حيث $P_W = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}\mathbf{W}$ هي مصفوفة القبعة للمربعات الصغرى الموزونة.

وهكذا فإن الإجراء الذي تم حد أثره يعمل مع μ_W وs_W^2 عوضاً عن $\overline{y}_0$ وs_0^2، على التوالي، عند حساب الحد الأعلى U_0 باستخدام الصيغة المذكورة أعلاه.

مثال على طريقة الحد من الأثر باستخدام بيانات الإيبولا

يوضح الشكل 5 أدناه الحد الأعلى لعتبة اكتشاف الإشارة فيما يتعلق بالبيانات التي جمعتها حزمة `epitweetr` عن فيروس إيبولا، سواء الأصلية (باللون الأحمر) التي لم يُحدّ من أثرها أو الحد الأعلى الذي حدّ من أثرها (باللون الأزرق) بعد وضع الإشارات السابقة من القيم التاريخية في الحسبان. لاحظ أن العتبة العليا تكتشف ثلاث إشارات إضافية، مقارنةً بالعتبة الأصلية.

الشكل 5: الحد الأعلى مع تقليل وزن خلاصة بيانات الإيبولا وبدونها



المرفق II استكشاف الأخطاء وإصلاحها ونصائح تقنية

يحتوي هذا الملحق على بعض النصائح والحلول الشائعة للأخطاء التي يقع فيها أو المشكلات التي قد يتعرض لها مستخدمو `epitweetr`، متضمناً ذلك شرحاً لعمليات التحقق تجدها في صفحة استكشاف الأخطاء وإصلاحها.

صفحة troubleshoot

بعد تفعيلك لعمليات التشخيص في صفحة استكشاف الأخطاء وإصلاحها، تعرض لك الصفحة عمليات الفحص والحالات المتعلقة بالجوانب التالية:

- **scheduler**: حزمة `Rtaskschedule R` قد ثُبتت. ينطبق هذا الأمر فقط على الأجهزة التي تعمل بنظام Windows
- **twitter_auth**: يضاف رمز تويتر المميز بعد إنهاء عملية المصادقة إما باستخدام حساب تويتر أو تطبيق مطوّر تويتر
- **search_running**: مهمة البحث قيد التشغيل
- **tweets**: جُمعت التغريدات

- R: **os64** هي 64بت
- **java**: تم تثبيت جافا وهو متاح أمام حزمة epitweetr
- **java64**: تم تثبيت جافا 64بت وأصبح متاحاً أمام حزمة epitweetr
- **java_version**: تم تثبيت إصدار جافا وأصبح متوافقاً مع `epitweetr`
- **winmsvc**: تم تثبيت حزمة مايكروسوفت فيجوال C++ 2010 SP1 Redistributable. ينطبق هذا الأمر فقط على الأجهزة التي تعمل بنظام Windows
- **detect_activation**: فُعّلت حلقة الكشف
- **detection_running**: عملية الكشف قيد التشغيل
- **winutils**: winutils تم تثبيت. وإن كان مغلوطاً، فيمكن تحميله عن طريق تشغيل مهمة تحديث التبعيات. ينطبق هذا الأمر فقط على الأجهزة التي تعمل بنظام Windows
- **java_deps**: تم تثبيت تبعيات جافا
- **move_from_temp**: يمكن لحزمة `epitweetr` أن تنقل الملفات تلقائياً من المجلد المؤقت إلى دليل البيانات
- **geonames**: تمّ تحميل وفهرسة قاعدة بيانات Geonames.org
- **languages**: تمّ تحميل وفهرسة متجهات اللغات
- **geotag**: تمّ تشغيل مهمة الوسم الجغرافي بنجاح
- **aggregate**: تمّ تشغيل مهمة التشغيل بنجاح
- **alerts**: أُنشئت التنبيهات
- **pandoc**: تم تثبيت pandoc وهو متاح أمام حزمة epitweetr. يعتبر ضرورياً لإنشاء ملفات بصيغة PDF
- **tex**: تمّ تثبيت توزيع النص وأصبح متاحاً أمام epitweetr. يعتبر ضرورياً لإنشاء ملفات بصيغة PDF

إدارة حلقات البحث والكشف (Windows)

بعد تنشيط أنابيب البحث والكشف من صفحة التكوين الخاصة بـنظام تشغيل `epitweetr` ((Windows، تنشأ مهمتان في مجدول المهام تفضي إلى تحفيز نافذتين طرفيتين. لاحظ أنه عند تسجيل دخول/ إيقاف تشغيل الكمبيوتر أو إغلاق النوافذ الطرفية، تتوقف مسارات البحث والكشف عن عملها.

لكن عند إعادة تفعيل هذه المهام من صفحة تكوين حزمة`epitweetr`، يكتب النظام فوق هذه المهام التي أُنشئت في مجدول المهام. وكإجراء بديل لذلك، بعد نجاحك بإتمام عملية تنشيط هذه المهام من حزمة `epitweetr`، تستطيع إدارتها بسهولة من مجدول المهام. كما يمكنك إيقاف عملها عن طريق إنهاء مهامها وتعطيلها في مجدول المهام، وإعادة تشغيلها عن طريق تفعيلها وتشغيلها في المجدول ذاته.

كما يتيح لك مجدول المهام، توطيد آلية عمل مهام البحث والكشف بحيث تجعلها "تعمل سواء سجّل المستخدم دخوله أم لا" لتتجنب التوقف الاعتباطي لعملها عند تسجيل الخروج أو إعادة تشغيل الكمبيوتر.

إدارة حلقات البحث والكشف (Linux وMac)

تتطلب أنظمة Linux أو Mac تفعيل عمل أنابيب البحث والكشف بصورة يدوية، وإذا تم تسجيل دخول/ إيقاف تشغيل الكمبيوتر أو إغلاق النوافذ الطرفية، عندها تتوقف حلقات البحث والكشف عن عملها. تذكّر أن تتبع الخطوات الواردة في قسم إعداد جمع التغريدات وحلقة كشف إشارات التنبيه لتشغيل هذه المهام مرة أخرى.

تشغيل خط أنابيب البحث والكشف

Cannot execute task #####: the task is already running""

تُنشئ حلقة الاكتشاف والبحث ملفين يحتويان على معرّفات العملية الخاصة بهما الموجودة في مجلد بيانات `epitwitter`: search.PID وdetect.PID. ويظهر هذا الخطأ في حال وضع `epitweetr` يده على عملية R أخرى تعمل في الوقت ذاته مستخدمة ذات المعرّف. ولتصلح هذا الخطأ، تحققْ أولاً ما إذا كانت حلقة البحث/الكشف تعمل حقاً في جلسة R أخرى أم

لا. في هذه الحالة، لا تحاول بدء المهمة لأن epitweetr يدعم فقط مثيلاً واحداً للمهمة ذاتها ليعمل في مجلد البيانات ذاته. لكن إن لم تكن عملية التشغيل مرتبطة بالمهمة، فيمكنك حذف ملف PID يدوياً ومحاولة بدء تشغيله مرة أخرى.

خطأ عند محاولة تجميع الملفات

نعزو هذا الخطأ سببين: - لا توجد مساحة كافية على قرص الملفات المؤقتة. ينتج عن التجميع ملفات مؤقتة قبل أن تُحفظ في مجلد epitweetr المقابل لها. في هذه هي الحالة، يرجى تبديل متغير البيئة في حسابك لـ TMP وTEMP ونقله إلى موقع آخر تتواجد فيه مساحة أكبر. - إذا ظهر الخطأ عند إنشاء ملف معين، فقد يكون هناك ملف سلسلة تالف متعلق بذاك التاريخ. احذف "country_counts" و"geolocated" و"topwords" المتعلقين بذاك التاريخ وأعد تشغيل المهمة يدوياً بالنقر على الزر المقابل في صفحة التكوين.

بدّلْ مستخدم مصادقة تويتر عند استخدام حساب تويتر

1. أنهِ حلقة/مهمة البحث وعطّل عملها في مجدول المهام لدى نظام (Windows) أو أغلق نافذة R / الطرفية مع حلقة/مهمة البحث في حال كنت تعمل مستخدماً نظامي (Linux وMac)
2. ابحث عن ملف يسمى ".rtweet_token" في الملفات المخفية. عادة ما يُحفظ في مجلد المستندات.
3. احذف هذا الملف.
4. ثم انقر على "Update properties" في صفحة configuration الخاصة بـ epitweetr.
5. فعّل خاصية عمل حلقة/مهمة البحث في مجدول المهام لدى نظام (Windows) أو فعّلْ الأمر في نافذة R / طرفية جديدة مع حلقة/مهمة البحث في حال كنت تعمل مستخدماً نظامي (Linux وMac). يمكنك الإطلاع على مزيد من التفاصيل في قسم "إعداد مجموعة التغريدات وحلقة اكتشاف إشعارات التنبيه"

تنزيل GeoNames و/ أو اللغات

"The specified size exceeds the maximum representable size. Error: Could not create the Java Virtual Machine"

إذا ظهر هذا الخطأ عند تشغيل GeoNames، فهذا يعني أن الجهاز يحتوي على Java 32 بت. يتعين عليك تثبيت جافا 64بت. وأن تجعله في متناول يد حزمة epitwitter إما عن طريق تعيين متغير البيئة "JAVA_HOME" أو عن طريق تعيين ملف جافا الثنائي الصحيح على نظام PATH.

"Max number of retried reached failed while processing languages. Error in get_geolocated_period(dataset): To aggregate, or calculate alerts geolocation must have been successfully executed, but no geolocation files were found"

في حال ظهر أمامك هذا الخطأ في صفحة التكوين، فهذا يعني فشل حزمة epitweetr في تحديد الموقع الجغرافي للتغريدات المجمّعة. نوصيك بشدة بالتوجه لإعادة تشغيل مهمة GeoNames واللغات فربما لم تُنزّل في المرة الأولى بصورة صحيحة. وعند تفعيلك هذه المهمة، تأكدْ من أن الجهاز ليس في وضعية عدم تسجيل دخول/إيقاف تشغيل أو في وضعية السكون.

تُظهر "Launch slots" في صفحة التكوين غير متوفر بدلاً من الفترات الزمنية

في المرة الأولى لتثبيت epitweetr وتشغيله، يتعين عليك تفعيل مهمة تحديد الموقع الجغرافي لأنابيب الكشف مرة واحدة على الأقل لمشاهدة الفترات الزمنية في "Launch slots" في صفحة التكوين.

تنزيل ملف بصيغة PDF الخاص بلوحة التحكم

Error in: LaTeX failed to compile C:\Users\name~1\...\file######.tex.""

يظهر هذا الخطأ في نظام تشغيل Windows عند النقر على "PDF" في لوحة التحكم دون حفظك لملف PDF. والسبب هو الطول الشديد الذي يتسم به المسار الواصل إلى متغيرات بيئة TEMP وTMP بالنسبة للمستخدم، فيعمد نظام تشغيل Windows إلى تقصير المسار وعليه يفشل epitweetr في العثور على المسار الجديد. يرجى اتباع الخطوات التالية لإصلاح هذا الخطأ:

1. افتح "متغير البيئة لحسابك"
2. قم بتغيير مسار TEMP وTMP إلى مسار أقصر (على سبيل المثال "C: \ Temp"). استخدم المسار ذاته لكل متغير من متغيري البيئة.
3. سجلْ الخروج ثم سجلْ الدخول
4. يمكنك الآن تنزيل ملف PDF وحفظه من لوحة التحكم

Error: pandoc document conversion failed with error 6""

1. حمّل البرنامج النصي التالي من الوصلة التالية (https://raw.githubusercontent.com/jgm/pandoc/master/macos/uninstall-pandoc.pl)
2. ألغ تثبيت pandoc (https://pandoc.org/installing.html) بوضعك قيد perl uninstall-pandoc.pl التشغيل.

مجاميع مختلفة في مخرجات لوحة التحكم

عند حسابك لعدد إجمالي التغريدات في لوحة التحكم لدى Shiny app أو في البيانات القابلة للتحميل، قد تحصل على نتائج متباينة في العدد الإجمالي للتغريدات بين المخرجات الثلاثة. ونعزو هذا للأسباب التالية:

1. World (all) مقابل World (geolocated)
   - الخيار الافتراضي المُعيّن للمناطق في World (all)، ما يعني أنه يجري أيضاً إدراج التغريدات غير المحددة جغرافياً في خط الاتجاه، ولكن يمكن تصور التغريدات المحددة جغرافياً فقط في الخرائط وجدول الكلمات الأكثر تواتراً، وبالتالي قد يختلف إجمالي عدد التغريدات بين هذه المخرجات عند اختيار World (all) أو تعيين الافتراضي الفارغ.
2. تحليل خاص بكل بلد
   - إذا حددت بلداً واحداً فقط في المرشحات، فسيعرض خط الاتجاه جميع التغريدات الخاصة بهذا البلد فقط، لكن الخريطة ستعرض التغريدات على المستوى دون الوطني في الخريطة. وقد يكون الموقع الجغرافي قد تم تحديده لبعض تغريدات بلد معين، ولكن دون إعطاء مزيداً من البيانات دون الوطنية فيم بعد تصبح هذه التغريدات مرئية في إجمالي خط الاتجاه، ولكن ليس في الفقاعات دون الوطنية في الخريطة.
3. المفردات الأكثر تواتراً
   - على نقيض ما يجري مع المخرجات الأخرى في لوحة التحكم، يعتمد عدد المفردات الأكثر تواتراً دائماً على موقع التغريدة بغض النظر عن عامل التصفية (بسبب سعة الذاكرة). لذا في حال تحديد موقع المستخدم أو كلا الموقعين في مرشّح الموقع، فقد يكتسب هذا الرقم مخرج إجمالي مختلف عن المخرجات الأخرى.

تلقي إشارات التنبيه في الوقت الحقيقي فقط

يتعلق هذا بالمستخدمين الذين حددوا موضوعات و/أو مناطق لتلقي التنبيهات ذات الصلة في الوقت الحقيقي أو حددوا موضوعات و/أو مناطق لتلقي إشعارات التنبيه ذات الصلة في فترة زمنية محددة. في هذه الحالة، إذا استلمت إشعارات التنبيه في الوقت الحقيقي فقط لجميع الموضوعات والمناطق، فقد يرجع ذلك إلى عدم تضمينك فترات زمنية في ملف المشتركين من صفحة التكوين. تُستخدم هذه الفترات الزمنية لإشعارات التنبيه المجدولة وإذا لم تدرج فترات زمنية في الملف، فسترسل إشعارات التنبيه من جميع الموضوعات والمناطق كتنبيهات في الوقت الحقيقي.

عدم استلم إشارات التنبيه عبر البريد الإلكتروني

إذا لم تستلم إشارات التنبيه عبر البريد الإلكتروني وظهر لديك خطأ في `epitweetr` يشير إلى رفض تسجيل الدخول، فهذا يعني أن `epitweetr` لم يتمكن من تسجيل الدخول إلى حساب البريد الإلكتروني المُعين في صفحة التكوين. نعزو أسباب ذلك إلى:

- الخادم أو المنفذ المضمن في صفحة التكوين غير صحيحين
- حظر `epitweetr` من محاولة تسجيل الدخول إلى حساب البريد الإلكتروني من قبل الخادم. من المتوقع حدوث هذه الحالة مع بعض حسابات البريد الإلكتروني الخاصة بالمؤسسات. في هذه الحالة، يرجى الاتصال بقسم تقانة المعلومات في مؤسستك
- وإذا كنت تستخدم حساب جي ميل، يتعين عليك تفعيل إعداد السماح بالتطبيقات الأقل أماناً من إعدادات حسابك

المراجع

Allévius, Benjamin, and Michael Höhle. 2017." " Prospective Detection of Outbreaks. *arXiv:1711.08960 [Stat]*, November. https://arxiv.org/abs/1711.08960.

Chatterjee, Samprit, and Ali S. Hadi. 1988. *Sensitivity Analysis in Linear Regression*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Farrington, C. P., N. J. Andrews, A. D. Beale, and M. A. Catchpole. 1996." A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease." *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159 (3): 547. https://doi.org/10.2307/2983331.

Fricker, Ronald D., Benjamin L. Hegler, and David A. Dunfee. 2008." Comparing Syndromic Surveillance Detection Methods: EARS' Versus a CUSUM-Based Methodology." *Statistics in Medicine* 27 (17): 3407–29. https://doi.org/10.1002/sim.3197.

Noufaily, Angela, Doyo Enki, Paddy Farrington, Paul Garthwaite, Nick Andrews, and Andre Charlett. 2013." An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems." *Online Journal of Public Health Informatics* 5 (1): e148. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3692796/.

Salmon, Maëlle, Dirk Schumacher, and Michael Höhle. 2016." Monitoring Count Time Series in *R* : Aberration Detection in Public Health Surveillance." *Journal of Statistical Software* 70 (10). https://doi.org/10.18637/jss.v070.i10.