

# GDELT DOC 2.0 API Debuts!

© JUNE 20, 2017

## GDELT API

We are incredibly excited to announce today the debut of the new GDELT 2.0 DOC API, which is our full text search API. A year and a half after the unveiling of our [first full text search API](#) on Christmas Day 2015, our new 2.0 API builds upon all of the lessons we've learned from that first API and all of the requests we've heard from all of you.

Perhaps the two biggest changes are that the API now searches a rolling window of the last 3 months of coverage, rather than just the last 24 hours of the original API and now includes all of the images processed by the Visual Global Knowledge Graph (VGKG), meaning for the first time you can both perform near-term longitudinal analyses and search for images based on the objects and activities they depict! In a nod to the intense demand we've heard from all of you for more seamless integration with web-first workflows and visualizations, the new API also supports JSON and JSONP output formats!

### Search Back to January 2017

One of the most-heard requests from all of you was for our search API to break the 24 hour barrier and enable you to search over much larger time periods. Thus, the new API now searches back to January 1, 2017. You can narrow your search to any time range within the last 3 months meaning you can still search just the last 24 hours if you want, but for those analyses more interested in longitudinal trends, we are very excited to see what you are able to do with this new historical search capability!

### Search The World's News Imagery

As with our new GEO 2.0 API, the DOC API offers seamless searching of the GDELT [Visual Knowledge Graph](#) (VGKG) deep learning global news imagery cataloging. Now you can search for all news images depicting fire or flooding or containing the Red Cross logo or mentioning Donald Trump in the caption and more! To our knowledge this new API represents the first global-scale deep learning-powered image search engine ever created, allowing you to explore the ever-more-critical visual narratives of the world's news coverage.

### Search Across 65 Languages

One of the most powerful aspects of the DOC 2.0 API is that you can search across all 65 machine translated languages supported by GDELT using English keywords/phrases as your search terms. [GDELT's Translingual](#) infrastructure machine translates 100% of all monitored coverage in 65 languages comprising 98.4% of GDELT's daily non-English monitoring volume. To our knowledge this is one of the largest initiatives in the world to mass machine translate global news coverage in realtime. In short, GDELT monitors news coverage from across the world, machine translates all of the coverage it sees in 65 of those languages into English and then allows you to search those machine translations. This allows you to "look across languages" and find all global coverage of your topic regardless of the language it was published in – an absolutely critical element in allowing you to peer deeply into local narratives and perspectives.

### Instant Embeddable Visualizations And JSON

Creating powerful interactive browser-based visualizations takes a lot of effort and so we've done all of the hard work for you and created advanced visualizations for each of the API's output modes that are custom designed to be dropped into your own web pages via iframe embedding. By just inserting an iframe into your page and setting its URL to this API you are able to instantly embed a live-updating advanced visualization that reflects global coverage from across the world in 65

modern statistical and data mining toolkits for further analysis. We also set the CORS ACAO header to the wildcard "\*" and add additional headers to make embedding as seamless as possible.

## QUICK START EXAMPLES

Here are some really simple examples to get you started using the API!

- **Volume timeline of coverage mentioning "islamic state" over the last 3 months.** This is one of the most basic kinds of visualizations and tracks the percent of all online coverage worldwide in all 65 languages that GDELT monitored over the last 3 months that contained the phrase "islamic state" anywhere in the article. This particular example uses the special "timelinevolinfo" mode that produces a popup as you mouse over the timeline that shows you the top 10 most relevant articles at each timestep.
  - URL: <https://api.gdeltproject.org/api/v2/doc/doc?query=%22islamic%20state%22&mode=timelinevolinfo&TIMELINESMOOTH=5>
- **Tone chart of "donald trump" coverage over last 3 months.** This is a powerful visualization that displays a histogram of how many articles mentioning the phrase "donald trump" over the last 3 months fell into each tone bin, from extremely negative to extremely positive. Mousing over each bar will show the top 10 most relevant articles that fell into that tone bin. This lets you rapidly assess the narratives in the middle neutral area and those taking an extreme positive and extreme negative stance, showing you the stories and views framing an issue.
  - URL: <https://api.gdeltproject.org/api/v2/doc/doc?query=%22donald%20trump%22&mode=tonechart>
- **Collage of images captioned "donald trump" over last 3 months.** This displays a list of 150 images that appeared somewhere on the web with caption or context information suggesting they either depicted or in some way related to Donald Trump or his presidency. Under each image is the number of times Google Images has seen that image on the open web and up to six example pages where the image has appeared before.
  - URL: <https://api.gdeltproject.org/api/v2/doc/doc?query=imagewebtag:%22donald%20trump%22&mode=imagecollageinfo&maxrecords=150>
- **Collage of images captioned "drone" over last 3 months.** Identical to the Donald Trump example above, this example shows global imagery with captions or contextual information suggesting they depict or relate to drones and UAVs.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(imagewebtag:%22drone%22%20OR%20imagewebtag:%22uav%22\)&mode=imagecollageinfo&maxrecords=150](https://api.gdeltproject.org/api/v2/doc/doc?query=(imagewebtag:%22drone%22%20OR%20imagewebtag:%22uav%22)&mode=imagecollageinfo&maxrecords=150)
- **Collage of images depicting the Red Cross over the last 3 months.** This displays all images that either had captions or contextual information suggesting they depict or relate to the Red Cross or OCR of the image revealed it contained the phrase "red cross" somewhere in the image or the image file contained embedded metadata with the phrase "red cross" or the Red Cross logo was observed somewhere in the image.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(imagewebtag:%22red%20cross%22%20OR%20imageocrmeta:%22red%20cross%22\)&mode=imagecollageinfo&maxrecords=150](https://api.gdeltproject.org/api/v2/doc/doc?query=(imagewebtag:%22red%20cross%22%20OR%20imageocrmeta:%22red%20cross%22)&mode=imagecollageinfo&maxrecords=150)
- **Collage of images capturing disasters from the previous week.** While the image examples above look to the textual captions and context of each image, this example uses Google's deep learning neural networks to look at the visual contents of each image and determine what it depicts. Here the first 150 images are returned that depict flooding, rubble, earthquake damage or fire that were found in articles published in the last 7 days.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(imagetag:%22flood%22%20OR%20imagetag:%22rubble%22%20OR%20imagetag:%22earthquake%22%20OR%20imagetag:%22fire%22\)&mode=imagecollageinfo&maxrecords=150&timespan=1week](https://api.gdeltproject.org/api/v2/doc/doc?query=(imagetag:%22flood%22%20OR%20imagetag:%22rubble%22%20OR%20imagetag:%22earthquake%22%20OR%20imagetag:%22fire%22)&mode=imagecollageinfo&maxrecords=150&timespan=1week)
- **List of articles containing images of disasters.** This is the exact same query as above, but instead of returning an image collage, it switches to Article List mode to return a list of articles that contained such images.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(imagetag:%22flood%22%20OR%20imagetag:%22rubble%22%20OR%20imagetag:%22earthquake%22%20OR%20imagetag:%22fire%22\)&mode=artlist&maxrecords=100&timespan=1week](https://api.gdeltproject.org/api/v2/doc/doc?query=(imagetag:%22flood%22%20OR%20imagetag:%22rubble%22%20OR%20imagetag:%22earthquake%22%20OR%20imagetag:%22fire%22)&mode=artlist&maxrecords=100&timespan=1week)
- **List of articles mentioning the Islamic State.** Here the first 100 articles mentioning the phrase "islamic state" or "isis" or "daesh" published in the last week are returned.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(%22islamic%20state%22%20OR%20isis%20OR%20daesh\)&mode=artlist&maxrecords=100&timespan=1week](https://api.gdeltproject.org/api/v2/doc/doc?query=(%22islamic%20state%22%20OR%20isis%20OR%20daesh)&mode=artlist&maxrecords=100&timespan=1week)
- **List of articles mentioning wildlife crime.** Here the first 100 articles mentioning the phrase "wildlife crime" or "poaching" or "illegal fishing" or "wildlife trade" published in the last week are returned. This allows you to look across global coverage on the topic across all 65 languages.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(%22wildlife%20crime%22%20OR%20poaching%20OR%20illegal%20fishing%22%20OR%20%22wildlife%20trade%22\)&mode=artlist&maxrecords=100&timespan=1week](https://api.gdeltproject.org/api/v2/doc/doc?query=(%22wildlife%20crime%22%20OR%20poaching%20OR%20illegal%20fishing%22%20OR%20%22wildlife%20trade%22)&mode=artlist&maxrecords=100&timespan=1week)

article as one item and, as a separate item, the URL of the mobile/AMP edition of the page, if available. This demonstrates how to use the API as a data source for web archiving.

- URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(%22climate%20change%22%20OR%20%22global%20warming%22\)&mode=artlist&maxrecords=200&timespan=1h&sort=datedesc&format=rssarchive](https://api.gdeltproject.org/api/v2/doc/doc?query=(%22climate%20change%22%20OR%20%22global%20warming%22)&mode=artlist&maxrecords=200&timespan=1h&sort=datedesc&format=rssarchive)
- **Gallery display of articles about climate change.** Here the most relevant 50 articles about climate change from the last three months are displayed in a "high design" magazine-style "gallery" format. This allows you to create visually striking summaries of coverage of a topic.
  - URL: <https://api.gdeltproject.org/api/v2/doc/doc?query=%22climate%20change%22&mode=artgallery>
- **Gallery display of images about pollution and litter.** Here the most relevant 50 global news images identified by Google's Cloud Vision API as either containing imagery of "pollution OR smog OR litter" or mentioning one of those three topics in its caption on the page or elsewhere on the web is displayed in a "high design" magazine-style "gallery" format. This allows you to create visually striking visual summaries of the visual news narrative around a topic.
  - URL: [https://api.gdeltproject.org/api/v2/doc/doc?query=\(imagetag:%22pollution%22%20OR%20imagetag:%22smog%22%20OR%20imagetag:%22litter%22%20OR%20imagewebtag:%22smog%22%20OR%20imagewebtag:%22pollution%22%20OR%20imagewebtag:%22litter%22\)&mode=imagegallery](https://api.gdeltproject.org/api/v2/doc/doc?query=(imagetag:%22pollution%22%20OR%20imagetag:%22smog%22%20OR%20imagetag:%22litter%22%20OR%20imagewebtag:%22smog%22%20OR%20imagewebtag:%22pollution%22%20OR%20imagewebtag:%22litter%22)&mode=imagegallery)

## FULL DOCUMENTATION

The GDELT GEO 2.0 API is accessed via a simple URL with the following parameters. Under each parameter is the list of operators that can be used as the value of that parameter.

- **QUERY.** This contains your search query and supports keyword and keyphrase searches, OR statements and a variety of advanced operators. NOTE – all of the operators below must be used as part of the value of the QUERY field, separated by spaces, and cannot be used as URL parameters on their own.
  - `""`. Anything found inside of quote marks is treated as an exact phrase search. Thus, you can search for "Donald Trump" to find all matches of his name.
    - `"donald trump"`
  - **(a OR b)**. You can specify a list of keywords to be boolean OR'd together by enclosing them in parentheses and placing the capitalized word "OR" between each keyword or phrase. Boolean OR blocks cannot be nested at this time. For example, to search for mentions of Clinton, Sanders or Trump, you would use "(clinton OR sanders OR trump)".
    - `(clinton OR sanders OR trump)`
  - `-`. You can place a minus sign in front of any operator, word or phrase to exclude it. For example `-sourcelang:spanish` would exclude Spanish language results from your search.
    - `-sourcelang:spanish`
  - **Domain.** Returns all coverage from the specified domain. Follow by a colon and the domain name of interest. Search for `domain:cnn.com` to return all coverage from CNN.
    - `domain:cnn.com`
  - **Domains.** This is identical to the main "Domain" operator above, but requires an exact match, allowing searching for common short domains like `un.org`. For example, when searching for `domain:un.org` many other domains that end in `un.org` are returned like `catholicsun.org`. Using this option you can restrict to a precise match, allowing you to return only articles from the `un.org` domain.
    - `domainis:un.org`
  - **ImageFaceTone.** Searches the average "tone" of human facial emotions in each image. Only human faces that appear large enough in the image to accurately gauge their facial emotion are considered, so large crowd photos where it is difficult to see the emotion of peoples' faces may not be scored accurately. The tone score of an average photograph typically ranges from +2 to -2. To search for photos where visible people appear to be sad, search `imagefacetone<-1.5`. Only available in any of the "image" modes.
    - `imagefacetone<-1.5`
  - **ImageNumFaces.** This searches the total number of foreground human faces in the image. Typically only unobstructed human faces facing toward the camera and in the foreground of the image are counted – large crowd scenes will not be counted properly. Use this to identify images depicting a certain number of people in the foreground of the photo. You can search for `<` less than, `>` more than or `=` – searching `imagenumfaces=3` will identify images with three human faces, while `imagenumfaces>5` will return images with more than 5 human faces. Only available in any of the "image" modes.
    - `imagenumfaces>3`



the image. To search for images of a specific event, such as "mobile congress" you would use this field, since that information would most likely either be found in signage in the background of the image, provided in the EXIF metadata in the image or listed in the caption under the image. The search parameter for this field must always be enclosed in quote marks, even when searching for a single word like "imageocrmeta:"zika"". Only available in any of the "image" modes.

- *imageocrmeta:"zika"*

- **ImageTag.** Every image processed by GDELT is assigned one or more topical tags from a universe of more than 10,000 objects and activities recognized by Google's algorithms. This is the primary and most accurate way of searching global news imagery monitored by GDELT, as these tags represent the ground truth of what is actually depicted in the image itself, whereas other fields like "imageocrmeta" and "imagewebtag" reflect metadata and caption information provided by others about the image. Always remember that these tags are assigned 100% by computer and thus you will always find some error in the results. You can find a list of all tags appearing in at least 100 images over the past year ([Image Tag Lookup](#)) – in addition the two special tags "safesearchviolence" and "safesearchmedical" can also be used. Searching for "imagetag:"safesearchviolence"" will return violent images, for example. Values must be enclosed in quote marks. Only available in any of the "image" modes.

- *imagetag:"safesearchviolence"*

- **ImageWebCount.** Every image processed by GDELT is run through the equivalent of a reverse Google Images search that searches the web to see if the image has ever appeared anywhere else on the web that Google has seen. Up to the first 200 web pages where the image has been seen are returned. This operator allows you to screen for popular versus novel images – searching for "imagewebcount<10" will search for relatively novel images while "imagewebcount>100" will return images that appear widely online. Note that this records only the number of pages that Google has seen the image on, not the number of sites, meaning that if, for example, CNN uses a single image widely in its reporting of a breaking news event and publishes many articles on the event with the same image, this count will be high for that image, even though it is a novel image. Only available in any of the "image" modes.

- *imagewebcount<10*

- **ImageWebTag.** Every image processed by GDELT is run through the equivalent of a reverse Google Images search that searches the web to see if the image has ever appeared anywhere else on the web that Google has seen. The system then takes every one of those appearances from across the web and looks at all of the textual captions appearing beside the image and compiles a list of the major topics used to describe the image across the web. This offers tremendous descriptive advantage in that you are essentially "crowdsourcing" the key topics of the image by looking at how it has been described across the web. Values must be enclosed in quote marks. Only available in any of the "image" modes. You can access a list of all tags appearing in at least 100 images ([Image WebTag Lookup](#)).

- *imagewebtag:"drone"*

- **Near.** Allows you to specify a set of keywords that must appear within a given number of words of each other. To use this operator, you specify the word "near", followed by the maximum distance all of the words can appear apart in a given document and still be considered a match, a colon, and then the list of words in quote marks. Phrase matching is not supported at this time, so the list of words is treated as a list of individual words that must all appear together within the given proximity. Note that if the words appear in a document in a different order than specified in the "near" operator, each ordering difference increments the word distance counted by the "near" operator. (Thus, near10:"donald trump" will return documents where "trump" appears within 10 words after "donald", but will also return documents in which "donald" appears within 9 words after "trump".) The distance measure is not precise and can count punctuation and other tokens as "words" as well. It is also important to remember that proximity in a document does not necessarily imply two words are connected semantically each other.

- *near20:"trump putin"*

- **Repeat.** Allows you to specify that a given word must appear at least a certain number of times in a document to be considered a match. To use this operator, you specify the word "repeat", followed by the number of times the word should appear, followed by the word itself in quote marks. Only a single word is permitted using this operator, it does not support phrase searches at this time. By limiting results to articles that mention a word multiple times, you can filter to just those articles more likely to actually be about your keyword, rather than merely casually mentioning it. Note that the "repeat" operator only requires that a document mention the keyword AT LEAST the requested number of times – a document will match even if it mentions the keyword many more times than the requested number.

- *repeat3:"trump"*

- **SourceCountry.** Searches for articles published in outlets located in a particular country. This allows you to narrow your scope to the press of a single country. For countries with spaces in their names, type the full name without the spaces (like "sourcecountry:unitedarabemirates" or "sourcecountry:saudi Arabia"). You can also use their 2-character FIPS country code ([Country Lookup](#)).

- *sourcecountry:france*

- **SourceLang.** Searches for articles originally published in the given language. The GEO API currently only allows you to search the English translations of ☐ all coverage, but you can specify that you want to limit your search to articles published in a particular language. Using this operator by itself you can

languages are supported ([Languages Lookup](#)).

- *sourcelang:spanish*

- **Theme.** Searches for any of the GDELT Global Knowledge Graph (GKG) Themes. GKG Themes offer a more powerful way of searching for complex topics, since they can include hundreds or even thousands of different phrases or names under a single heading. To search for coverage of terrorism, use "theme:terror". You can find a list of all themes that have appeared in at least 100 articles over the past two years ([GKG Theme Lookup](#)).

- *theme:TERROR*

- **Tone.** Allows you to filter for only articles above or below a particular tone score (ie more positive or more negative than a certain threshold). To use, specify either a greater than or less than sign and a positive or negative number (either an integer or floating point number). To find fairly positive articles, search for "tone>5" or to search for fairly negative articles, search for "tone<-5".

- *tone<-5*

- **ToneAbs.** The same as "Tone" but ignores the positive/negative sign and lets you simply search for high emotion or low emotion articles, regardless of whether they were happy or sad in tone. Thus, search for "toneabs<1" for fairly neutral articles or search for "toneabs>10" for fairly emotional articles.

- *toneabs>10*

- **MODE.** This specifies the specific output you would like from the API, ranging from timelines to word clouds to article lists.

- **ArtList.** This is the most basic output mode and generates a simple list of news articles that matched the query. In HTML mode articles are displayed in a table with its social sharing image (if available) to its left, the article title, its source country, language and publication date all shown. RSS output format is only available in this mode.

- **ArtGallery.** This displays the same information as the "ArtList" mode, but does so using a "high design" visual layout suitable for creating magazine-style collages of matching coverage. Only articles containing a social sharing image are included.

- **ImageCollage.** This displays all matching images that have been processed by the GDELT Visual Global Knowledge Graph (VGKG), which runs each image through Google's Cloud Vision API deep learning image cataloging. If your query does not contain any image-related search terms, this mode will return a list of all VGKG-processed images that were contained in the body of matching articles, while if your search included image terms, only matching images will be shown. Thus, this mode is most relevant when used with the various image-related query terms. Each image is provided with a link to the article containing it. Note that the document extraction system used by GDELT may on occasion make mistakes and associate an image with a news article in which it appeared only as an inset or unrelated footer, though this is usually rare. This mode is most useful for understanding the visual portrayal of your search.

- **ImageCollageInfo.** This yields identical output as the ImageCollage option, but adds four additional pieces of information to each image: 1) the number of times (up to 200) it has been seen before on the open web (via a reverse Google Images search), 2) a list of up to 6 of those web pages elsewhere on the web where the image was found in the past, 3) the date the photograph was captured via in the image's internal metadata (EXIF/etc), and 4) a warning if the image's embedded date metadata suggests the photograph was taken more than 72 hours prior to it appearing in the given article. Using this information you can rapidly triage which of the returned images are heavily-used images and which are novel images that have never been found anywhere on the web before by Google's crawlers. (You can also use the "imagewebcount" query term above to restrict your search to just images which have appeared a certain number of times.) Only a relatively small percent of news images contain an embedded capture timestamp that documents the date and time the image was taken or created and it is not always accurate, but where available this can offer a powerful indicator that a given image may be older than it appears and for applications that rely on filtering for only novel images (such as crisis mapping image cataloging), this can be used as a signal to perform further verification on an image.

- **ImageGallery.** This displays most of the same information as the "ImageCollageInfo" mode (though it does not include the embedded date warning), but does so using a "high design" visual layout suitable for creating magazine-style collages of matching coverage.

- **ImageCollageShare.** Instead of returning VGKG-processed images, this mode returns a list of the social sharing images found in the matching news articles. Social sharing images are those specified by an article to be shown as its image when shared via social media sites like Facebook and Twitter. Not all articles include social sharing images and the images may sometimes only be the logo of the news outlet or not representative of the article contents, but in general they offer a reasonable visual summary of the core focus of the article and especially how it will appear when shared across social media platforms.

- **TimelineVol.** This is the most basic timeline mode and returns the volume of news coverage that matched your query by day/hour/15 minutes over the search period. Since the total number of news articles published globally varies so much through the course of a day and through the weekend and holiday periods, the API does not return a raw count of matched articles, but instead divides the number of matching articles by the total number of all articles monitored by GDELT in each time step. Thus, the timeline reports volume as a percentage of all global coverage monitored by GDELT. For time spans of less than 72 hours, the timeline uses a time step of 15 minutes to provide maximum temporal resolution, while for time spans from 72 hours to one week it uses an hourly resolution and for time spans of greater than a week it uses a daily resolution. In HTML mode the timeline is displayed as an interactive browser-based visualization.

- **TimelineVolRaw.** This is identical to the standard TimelineVol mode, but instead of reporting results as a percent of all online coverage monitored by GDELT, it returns the actual number of distinct articles that matched your query. In CSV and JSON output modes, an additional "norm" field is returned



- **TimelineVolInfo.** This is identical to the main TimelineVol mode, but for each time step it displays the top 10 most relevant articles that were published during that time interval. Thus, if you see a sudden spike in coverage of your topic, you can instantly see what was driving that coverage. In HTML mode a popup is displayed over the timeline as you mouse over it and you can click on any of the articles to view them, while in JSON and CSV mode the article list is output as part of the file.
- **TimelineTone.** Similar to the main TimelineVol mode, but instead of coverage volume it displays the average "tone" of all matching coverage, from extremely negative to extremely positive.
- **TimelineLang.** Similar to the TimelineVol mode, but instead of showing total coverage volume, it breaks coverage volume down by language so you can see which languages are focusing the most on a topic. Note that the GDELT APIs currently only search the 65 machine translated languages supported by GDELT, so stories trending in unsupported languages will not be displayed in this graph, but will likely be captured by GDELT as they are cross-covered in other languages. With the launch of GDELT3 later this summer, the resolution and utility of this graph will increase dramatically.
- **TimelineSourceCountry.** Similar to the TimelineVol mode, but instead of showing total coverage volume, it breaks coverage volume down by source country so you can see which countries are focusing the most on a topic. Note that GDELT attempts to monitor as much media as possible in each country, but smaller countries with less developed media systems will necessarily be less represented than larger countries with massive local press output. With the launch of GDELT3 later this summer, the resolution and utility of this graph will increase dramatically.
- **ToneChart.** This is an extremely powerful visualization that creates an emotional histogram showing the tonal distribution of coverage of your query. All coverage matching your query over the search time period is tallied up and binned by tone, from -100 (extremely negative) to +100 (extremely positive). (Though typically the actual range will be from -20 to 20 or less). Articles in the -1 to +1 bin tend to be more neutral or factually-focused, while those on either extreme tend to be emotionally-laden diatribes. Typically most sentiment dashboards display a single number representing the average of all coverage matching the query ala "The average tone of Donald Trump coverage in the last week is -7". Such displays are not very informative since its unclear what precisely "-7" means in terms of tone and whether that means that most coverage clustered around -7 or whether it means there were a lot of extremely negative and extremely positive coverage that averaged out to -7, but no actual coverage around that tonal range. By displaying tone as a histogram you are able to see the full distributional curve, including whether most coverage clusters around a particular range, whether it has an exponential or bell curve, etc. In HTML mode you can mouse over each bar to see a popup with the top 10 most relevant articles in that tone range and click on any of the headlines to view them.
- **WordCloudImageTags.** This is identical to the WordCloudEnglish mode, but instead of the article text words, this mode takes all of the VGKG-processed images found in the matching articles (or which matched any image query operators) and constructs a histogram of the top topics assigned by Google's deep learning neural network algorithms as part of the Google Cloud Vision API.
- **WordCloudImageWebTags.** This is identical to the WordCloudImageTags mode, but instead of using the tags assigned by Google's deep learning algorithms, it uses the Google knowledge graph topical taxonomy tags assigned by the Google Cloud Vision API's Web Annotations engine. This engine performs a reverse Google Images search on each image to locate all instances where it has been seen on the open web, examines the captions of all of those instances of the image and compiles a list of topical tags that capture the contents of those captions. In this way this field offers a far more powerful and higher resolution understanding of the primary topics and activities depicted in the image, including context that is not visible in the image, but relies on the captions assigned by others, whereas the WordCloudImageTags field displays the output of deep learning algorithms considering the visual contents of the image.
- **FORMAT.** This controls what file format the results are displayed in. Not all formats are available for all modes. To assist with website embedding, the CORS ACAO header for all output of the API is set to the wildcard "\*", permitting universal embedding.
  - **HTML.** This is the default mode and returns a browser-based visualization or display. Some displays, such as word clouds, are static images, some, like the timeline modes, result in interactive clickable visualizations, and some result in simple HTML lists of images or articles. The specific output varies by mode, but all are intended to be displayed directly in the browser in a user-friendly intuitive display and are designed to be easily embedded in any page via an iframe.
  - **CSV.** This returns the requested data in comma-delimited (CSV) format. The specific set of columns varies based on the requested output mode. Note that since some modes return multilingual content, the CSV is encoded as UTF8 and includes the UTF8 BOM to work around Microsoft Excel limitations handling UTF8 CSV files.
  - **RSS.** This output format is only available in ArticleList mode and returns the list of matching article URLs and titles in RSS 2.0 format. This makes it possible to display the results using any standard RSS reader. It also makes it seamless for web archives to create tailored archival feeds to preserve news coverage on certain topics or meeting certain criteria.
  - **RSSArchive.** This special format is also only available in ArticleList mode and extends the standard RSS output by including both the main article URL and its alternative mobile or AMP version, if available, as a separate item. If no mobile versions of the search result articles are available, the output of this format will be identical to the standard RSS output format. For any article in the search results that had an alternative mobile or AMP edition, a second item will appear in the RSS feed for the mobile/AMP version. If both an AMP version and a mobile version of the page is available, only the AMP version will be returned. This format is intended for use by web archives to create tailored feeds that preserve both the desktop and mobile versions of matching coverage given that mobile versions are often different than their desktop counterparts. By consuming this feed as a data source, web archives can automatically ensure they are capturing both desktop and mobile experiences of matching content.
  - **JSON.** This returns the requested data in UTF8 encoded JSON. The specific fields varies by output mode.



- **JSONFeed.** This output format is only available in ArticleList mode and returns the list of matching article URLs and titles in JSONFeed 1.0 format.
- **TIMESPAN.** By default the DOC API searches the last 3 months of coverage monitored by GDELT. You can narrow this range by using this option to specify the number of months, weeks, days, hours or minutes (minimum of 15 minutes). The API then only searches documents published within the specified timespan backwards from the present time. If you would instead like to specify the precise start/end time of the search instead of an offset from the present time, you should use the STARTDATETIME/ENDDATETIME parameters.
  - **Minutes.** Specify a number followed by "min" to provide the timespan in minutes.
  - **Hours.** Specify a number followed by "h" or "hours" to provide the timespan in hours.
  - **Days.** Specify a number followed by "d" or "days" to provide the timespan in days.
  - **Weeks.** Specify a number followed by "w" or "weeks" to provide the timespan in weeks.
  - **Months.** Specify a number followed by "m" or "months" to provide the timespan in months.
- **STARTDATETIME/ENDDATETIME.** These parameters allow you to specify the precise start and end date/times to search, instead of using an offset like with TIMESPAN.
  - **STARTDATETIME.** Specify the precise date/time in YYYYMMDDHHMMSS format to begin the search – only articles published after this date/time stamp will be considered. It must be within the last 3 months. If you do not specify an ENDDATETIME, the API will search from STARTDATETIME through the present date/time.
  - **ENDDATETIME.** Specify the precise date/time in YYYYMMDDHHMMSS format to end the search – only articles published before this date/time stamp will be considered. It must be within the last 3 months. If you do not specify a STARTDATETIME, the API will search from 3 months ago through the specified ENDDATETIME.
- **MAXRECORDS.** This option only applies to the ArticleList and various ImageCollage modes, it is ignored in all other modes. To conserve system resources, in Article List and the ImageCollage modes, the API only returns up 75 results by default, but this can be increased up to 250 results if desired by using this URL parameter.
- **TIMELINESMOOTH.** This option is only available in the various Timeline modes and performs moving window smoothing over the specified number of time steps, up to a maximum of 30. Due to GDELT's high temporal resolution, timeline displays can sometimes capture too much of the chaotic noisy information environment that is the global news landscape, resulting in jagged displays. Use this option to enable moving average smoothing up to 30 days. Note that since this is a moving window average, peaks will be shifted to the right, up to several days or weeks at the heaviest smoothing levels.
- **TRANS.** Only available in ArticleList mode with HTML output, this embeds a machine translation widget in the results page to seamlessly machine translate all of the article titles into your requested language. Currently only the Google Translate Widget is supported. This means that if your primary language is French, all article titles in your search results across all 65 core languages that GDELT supports will be transparently translated in your browser instantly by Google Translate into French.
  - **GoogTrans.** Set to "googtrans" to embed the Google Translate Widget, which is the only translation widget presently supported.
- **SORT.** By default results are sorted by relevance to your query. Sometimes you may wish to sort by date or tone instead.
  - **DateDesc.** Sorts results by publication date, displaying the most recent articles first.
  - **DateAsc.** Sorts results by publication date, displaying the oldest articles first.
  - **ToneDesc.** Sorts results by tone, displays the most positive articles first.
  - **ToneAsc.** Sorts results by tone, displays the most negative articles first.
  - **HybridRel.** This is the default new relevance sorting mode for all searches of content published after 12:01AM September 16, 2018. It uses a combination of the textual relevance of the article and other signals, including the "popularity" of the outlet to rank highly relevant content from well known outlets at top, rather than ranking content exclusively based on its textual relevance, which tends to surface obscure coverage. We will be constantly refining the underlying scoring models over time to yield the best possible results and once we have a final model that performs well in all scenarios we will retroactively apply it to our entire backfile and make it available for all searches. This mode is not currently available for image searches, only textual article searches.
- **TIMEZOOM.** This option is only available for timeline modes in HTML format output and enables interactive zooming of the timeline using the browser-based visualization. Set to "yes" to enable and set to "no" or do not include the parameter, to disable. By default, the browser-based timeline display allows interactive examination and export of the timeline data, but does not allow the user to rezoom the display to a more narrow time span. If enabled, the user can click-drag horizontally in the graph to select a specific time period. If the visualization is being displayed directly by itself (it is the "parent" page), it will automatically refresh the page to display the revised time span. If the visualization is being embedded in another page via iframe, it will use postMessage to send the new timespan to the parent page with parameters "startdate" and "enddate" in the format needed by the STARTDATETIME and ENDDATETIME API parameters. The parent page can then use these parameters to rewrite the URLs of any API visualizations embedded in the page and reload each of them. This allows the creation of dashboard-like displays that contain multiple DOC API visualizations where the user can zoom the timeline graph at the top and have all of the other displays automatically refresh to narrow their coverage to that revised time frame.



Select Month▼

