Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

OMM (Online Media Monitor)
CRG Mediale Konstruktionen des Klimawandels & ICDC / CRG Media Constructions of Climate Change & ICDC (Integrated Climate Data Center)
Team: Michael Brüggemann, Remon Sadikni, Lars Guenther
Stand 20.6.2022

Online Media Monitor (OMM) – Funktionsweise, Methode und technische Dokumentation
Online Media Monitor (OMM) – Functionality, method and technical documentation

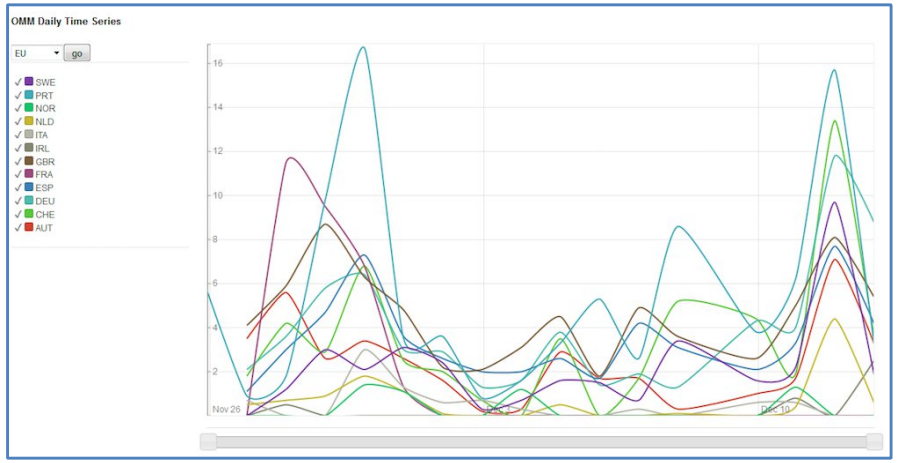
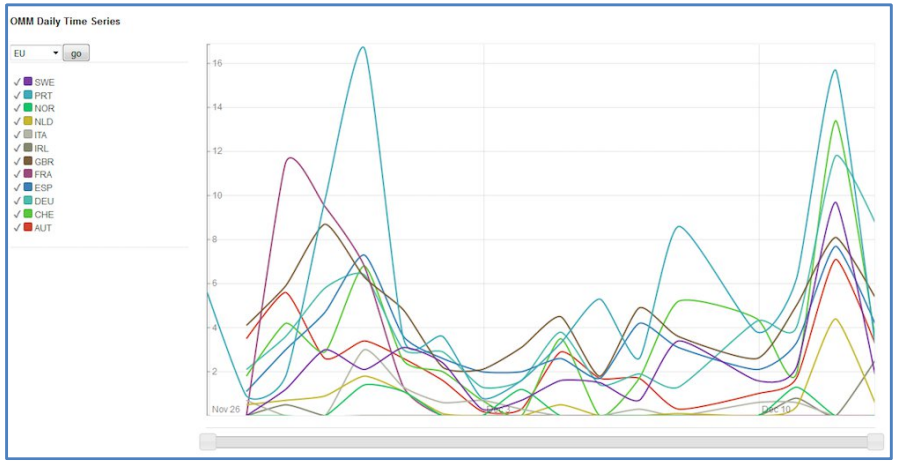
Deutsche Version	*English Version*
Was ist der OMM? Der OMM (Online Media Monitor) ist ein automatisiertes Tool, welches tägliche Auswertungen zur weltweiten Online-Klimaberichterstattung in momentan 18 Ländern erstellt und diese in einer Datenbank speichert. Diese Länder sind: Argentinien, Australien, Brasilien, Deutschland, Großbritannien, Indien, Italien, Kanada, Malaysia, Neuseeland, Norwegen, Polen, Schweden, Schweiz, Singapur, Spanien, Südafrika und die USA. Die Länder wurden nach unterschiedlichen Graden der Betroffenheit durch den Klimawandel (Klimarisikoindex[1]) ausgewählt, als auch bezogen auf eine möglichst hohe Varianz (Bevölkerungszahlen, Internetdistribution etc.) innerhalb der Kontinente.	**What is OMM?** OMM (Online Media Monitor) is an automated tool, which provides daily analyses of worldwide online climate reporting in currently 18 countries. The evaluated countries are: Argentina, Australia, Brazil, Germany, Great Britain, India, Italy, Canada, Malaysia, New Zealand, Norway, Poland, Sweden, Switzerland, Singapore, Spain, South Africa, and the United States. The countries included in the sample were chosen based on varying degrees of the impacts of climate change and a high level of variance (e.g. population, internet distribution rate) on all continents.
Welche Medien beinhaltet der OMM?	**What media are being evaluated?**

[1] Kreft, Sönke/David Eckstein/Lisa Junghans/Candice Kerestan/Ursula Hagen (2015): Climate Risk Index 2015. Germanwatch e.V., URL: www.germanwatch.org/en/cri (letzter Zugriff am 29.11.2015).

Grundsätzlich versucht der OMM mindestens 2 Medien pro Land abzubilden. In der Regel werden die Medien in der Landessprache analysiert – es sei denn, es gibt ein der Originalsprache etwa äquivalentes Angebot auch für andere Sprachen (i.d.R. Englisch). Die von uns berücksichtigten Sprachen sind: Englisch, Deutsch, Italienisch, Spanisch, Polnisch, Norwegisch und Schwedisch Für die Aufnahme eines Mediums ins Sample müssen zudem vier weitere Bedingungen erfüllt sein: • **Aktualisierung***:* Es müssen tagesaktuell Nachrichten auf der Home-Page veröffentlicht werden (Ausnahme: Samstag/Sonntag); temporäre Ausfälle ergeben sich z.T. auf Seiten der Anbieter durch geopolitische Konflikte (derzeit: arabischer Raum) • **Nationaler Radius der Berichterstattung***:* Die Berichterstattung des Mediums muss überregional (das heißt: national) sein. Dies wird z.B. durch Ressorts, wie „National News“, deutlich. • **Online-Reichweite**: Das Medium muss in weltweiten Reichweitenrankings von Webseiten (hier wird auf kommerzielle Unternehmen wie ALEXA Internet, 4imn oder Newspaper24 zurückgegriffen) möglichst weit oben gelistet sein. Die Reichweite ist für die Auswahl der Medien pro Land letztlich das entscheidende Kriterium. • **Journalistisches Angebot für unterschiedliche Zielgruppen**: Die Nachrichtenseite muss sich zu einer der beiden Gruppen (Qualität- oder Boulevardmedium) zuordnen lassen Ausgeschlossen werden Webseiten, die nur begrenzt Nachrichteninhalte anbieten, sondern primär andere Webdienste (E-Mail, etc.).	In general, OMM tries to cover at least 2 media outlets per country. The OMM is capable of processing various languages including English, German, Spanish, Italian, Polish, Norwegian, and Swedish. Primarily, the media are analysed in the respective official language. However, if there is a comprehensive offer of English content (e.g., in or Singapore), these newspapers are being included in the sample. Furthermore there are four criteria to be met by each medium in order to be part of the sample: • **Topicality:** News has to be published on a daily basis (except for Saturday/Sunday); temporary downtimes may occur due to geopolitical conflicts (e.g. the current situation in the Arabian region). • **Nation-wide reporting:** The respective reporting has to take place on a national level. This is indicated by a certain section, such as “National News”. • **Relevance & Range:** This is the most important criterion for the selection of the media for each country. The medium has to be ranked preferably high in worldwide range position tables. Therefore commercial websites such as ALEXA Internet, 4imn and newspaper24 were considered. • **Journalistic offers that address different target groups:** Each news site has to be classified as belonging to the group of elite or popular reporting (see footnote 3). Furthermore, media that primarily offer other web services (such as email service) are excluded from the sample. News coverage in Germany is therefore represented by SPIEGEL ONLINE, sueddeutsche.de. Whereas the media representing Great Britain in the sample are: The Guardian and The Daily Telegraph. A breakdown by type of media shows that 71% of the 42 media within

Für Deutschland beispielsweise sind somit folgende Nachrichtenmedien Teil des Samples: SPIEGEL ONLINE und sueddeutsche.de . Für Großbritannien The Guardian und der Daily Telegraph Die Verteilung der Online-Newsmedien in unserem Sample (42 Titel) nach Medientyp zeigen, dass 71% der Titel Online-Ableger von Zeitungen sind. Nur 14% sind „online only"-Angebote, 11% Fernsehsender, die restlichen Sonstige (Agenturen, Magazine, etc.). Haben die Nachrichtenseiten eine **Paywall**, so kann diese umgangen werden, sofern es sich um eine *metered paywall* (z.B. New York Times) handelt. *Hard paywalls* können nicht umgangen werden. Hier zieht der OMM alle frei verfügbaren Artikel. Letztendlich konnten nur die Medien aufgenommen werden, die RSS Feeds anbieten und in Reliablitätstests im Vergleich zur Googlesuche mindestens 80% der Klimawandelartikel gezogen haben.	the sample are online versions of printed newspapers, while only 14% are "online only" providers – 11% are web presences of news channels and the rest are agencies, magazines etc. Some News sites have **paywalls** that can be bypassed (*metered paywalls*, see e.g. New York Times). However, there was no technical possibility to overcome *hard paywalls*. In this case, only the freely available content is included in the analysis. In the end, media outlets could only be included that offer RSS feeds and show at least 80% of climate change articles compared to google search results.
Wie wird ein Artikel definiert? Der OMM greift nur Textdaten ab. Das heißt: Artikel, die aus Videos, Bildern oder anderen Formaten bestehen, werden nicht analysiert. Sollten Bilder oder Videos zusätzliche Elemente eines Artikels sein, werden diese Zusatzmaterialien nicht erhoben. Ein Artikel muss mindestens eine Überschrift, ein Datum sowie einen Textkorpus haben, um vom OMM aufgenommen zu werden. Zudem muss der Artikel auf der Stammseite des Online-Nachrichtenportals erschienen sein (z.B. SPIEGEL ONLINE, keine Analyse von Artikeln, die von Spiegel Online auf das Manager-Magazin bspw. verlinken). Weitere	**What does OMM consider an article?** OMM crawls through the news sites analyzing only the textual contents. Thus, news that mainly consist of videos, pictures, slideshows or other formats are not considered articles by the tool. To be taken into the account, the content has to meet the minimum requirements of an article, which are: a headline, publishing date and a text body. Also, the article must be published on the news sites' main page : links that lead to other news sites are not analyzed(e.g. articles of Spiegel online linking to the Manager-Magazin). Articles that do not show the standard layout of the original website are not analyzed by the OMM. If there are several subpages for one article,

technisch bedingte Ausnahmen gibt es für Artikel, welche nicht das Standardlayout der Original-Nachrichtenseite verwenden. Diese werden nicht vom OMM erfasst. Artikel, die sich über mehrere Seiten erstrecken, werden derzeit nur partiell erfasst.	only the first page is considered.
Wie werden die Artikel von den Webseiten abgegriffen? Der OMM bedient sich RSS Feeds, um potenzielle Artikelseiten zu erfassen. Diese werden in regelmäßigen, der Updatefrequenz angepassten Abständen eingelesen. Um Artikel auf den erhaltenen Webseiten zu erfassen, werden diese durch den Python lxml.html Parser in ein Objektmodell zerlegt, aus dem die jeweiligen relevanten Inhalte selektiert werden. Diese Selektion findet durch seitenspezifische XPath Ausdrücke statt, die Inhalte anhand ihrer Position innerhalb der HTML-Struktur einer Seite auswählen. Eine Unterseite gilt als Artikelseite, wenn strukturell alle Elemente eines Artikels dort gefunden werden, unabhängig vom Inhalt dieser Elemente. Einige wenige Medien, die keine RSS Feeds zur Verfügung stellen, werden alle 3 Monate heruntergeladen und damit zeitversetzt aktualisiert.	**How are the articles collected from the websites?** The OMM uses RSS feeds to detect potential articles. They are read in regular intervals. To detect items on the resulting web pages, they are parsed by the Python module lxml.html into an object model that provides the relevant content. This selection is done by site-specific XPath expressions, which select the content based on their position within the HTML structure of a page. A page is considered an article, if all the elements of an article can be found, regardless of the content of these elements. A few media outlets that don't provide RSS feeds are downloaded every 3 months and updated in a delayed mode.
Welche Suchbegriffe nutzt der OMM zur Klassifizierung der Klimaberichterstattung? Diese Suchbegriffe (siehe Anhang) sind individuell für jedes Land festgelegt und mit Hilfe von Muttersprachlern entwickelt worden. Hier gibt es sprachspezifische Unterschiede. Während im Englischen mit den Suchworten *climate change* (bzw. climat* chang*) und *global warming* (bzw. global warm*) der überwiegende Teil der Berichterstattung zum Klimawandel abgedeckt wird, müssen z.B. im Deutschen und Niederländischen die Suchstrings feiner definiert werden. Die jeweiligen Suchstrings pro Land sind innerhalb des Datensatzes	**Which keywords are being used in order to classify the climate reporting?** Since there are language-specific differences between the countries, slightly different keywords for every language (see appendix) have been elaborated with the help of native speakers. In English, the major part of the climate reporting is covered by the keywords *climate change* (or climat* chang*) and *global warming* (or global warm*), whereas the German and Dutch keywords have to be defined more precisely. If one of the keywords is found anywhere in the textual content of the article, this causes the articles entry in the data base. The respective

einsehbar. Taucht einer der Suchbegriffe im Textkorpus des Artikels auf, so wird der entsprechende Artikel in der OMM-Datenbank gespeichert. Die momentan funktionierenden Medien und ein Error Log finden sich auf der Webseite https://climatematters.blogs.uni-hamburg.de/omm/omm-logfile/.	keyword that caused the entry can be found linked to the according article. The currently working media outlets and an error log can be found on the website https://climatematters.blogs.uni-hamburg.de/omm/omm-logfile/.
Welche Analysen gibt der OMM täglich aus? Der OMM liefert täglich aktualisierte Daten zur weltweiten Klimaberichterstattung in Online-Nachrichtenmedien: **Zeitreihendaten**:  *Abbildung 1: Zeitreihe für die Berichterstattung in Europa während des COP21* Tägliche Zeitreihen geben sowohl den *relativen*, als auch den *absoluten* Anteil der Klimaberichterstattung an der Gesamtberichterstattung eines Mediums/eines Landes an. Hierbei gibt es Übersichten je nach Kontinent.	**What analyses are provided day-to-day?** The OMM provides daily updated data regarding the worldwide climate reporting: **Time series data:**  *Figure 2: Time series (Europe) for the coverage during COP21* Daily time series show the relative as well as the absolute number of climate reporting in proportion to the overall reporting of a medium/country.

Kartendaten Die Darstellung anhand einer Weltkarte ermöglicht einen besseren Vergleich zwischen den Ländern. Farbliche Abstufungen zeigen dabei die landespezifischen Unterschiede der *relativen* Klimaberichterstattung an der Gesamtberichterstattung pro Land des gestrigen Tages.	**Map data** Visualization of the data on a world map allows a comparison between the countries surveyed. Country-specific differences in yesterdays' relative parts of climate reporting are displayed by color graduation.
Welche Daten werden von Twitter erhoben? Der OMM durchsucht auch Tweets, die den Klimawandel zum Thema haben. Es werden folgende Tweets in der Online Auswertung berücksichtigt: - Mindestens einer dieser Suchbegriffe muss enthalten sein: #climatechange OR "climate change" OR "global warming" OR Klimawandel - Mindestens 5 Retweets Bis zum 18.1.2021 musste mindestens ein Link im Tweet enthalten sein, danach wurde dieses Kriterium (auch rückwirkend) fallengelassen. Dabei ist zu beachten, dass die Twitter API im „compat" mode, den wir nutzen, den Text des Tweets nach 140 Zeichen abschneidet und in diesem Fall dann ein Link zu Twitter auf den vollständigen Tweet gesetzt wird. Dieser wird von uns dann als erster Link gewertet und beeinflusst sowohl die Anzahl der Tweets in der Zeitreihe als auch das Domain-Ranking. Seit dem 22.7.2021 wurde der Suchstring erweitert: #climatechange OR "climate change" OR "global warming" OR Klimawandel **OR "climate crisis" OR "climate emergency" OR "Klimakrise"** Um 3 Uhr nachts (UTC) werden die Tweets des Vortags gezogen, um auch späten Tweets noch Retweets zu ermöglichen.	**Which data from Twitter are considered?** The OMM searches Tweets that are related to climate change. The following tweets are considered in the publicly available evaluation: - At least one of these search strings has to be present: #climatechange OR "climate change" OR "global warming" OR Klimawandel - At least 5 retweets Until Jan 18, 2021, there had to be additionally at least one link in the tweet. After this date, this criterion has been removed. The Twitter API in the "compat mode" (that we are using) truncates the text of the tweet after 140 characters and in this case a link to Twitter is placed in the text directing to the complete tweet. These artificial links are evaluated as the first tweet link and have an impact on the time series and the Domain ranking. From July 22, 2021 on, we extended the search string: #climatechange OR "climate change" OR "global warming" OR Klimawandel **OR "climate crisis" OR "climate emergency" OR "Klimakrise"** OMM downloads yesterdays tweets at 3 am (UTC) to provide the possibility of retweets.

Auch für die Twitterauswertung gibt es ein Error Log, das beschreibt, wann die Software nicht funktioniert hat und die Ergebnisse daher nicht vollständig sind: https://climatematters.blogs.uni-hamburg.de/omm/omm-twitter-logfile/ Zusätzlich zur Anzahl der Tweets pro Tag seit dem 1.3.2017, werden aus diesen Daten 2 weitere Produkte generiert. Aus den ersten Links jedes Tweets werden die Domains extrahiert und als Ranking über verschiedene Zeiträume aggregiert. Zusätzlich wird der am meisten „geretweetete“ Tweet von gestern gezeigt.	There is also an error log for OMM Twitter that describes when the software did not work properly: https://climatematters.blogs.uni-hamburg.de/omm/omm-twitter-logfile/ In addition to the tweet count per day from March 2017 on, two more products are generated by these data sets. The domains of the first links of every tweet are extracted and a ranking over these domains is generated, aggregated over different periods. Additionally, yesterday’s most retweeted tweet is shown.
Wie kann ich den OMM für eigene Forschungszwecke nutzen? CliSAP (bzw. CEN)-Mitglieder melden bitte ihr Interesse an über michael.brueggemann@uni-hamburg.de Verantwortlich für den OMM ist Prof. Dr. Michael Brüggemann, technischer Support erfolgt über Remon Sadikni.	**How can I use OMM for my research project?** CliSAP (and CEN) members please contact us by writing an email to michael.brueggemann@uni-hamburg.de Responsible for the OMM is Prof. Dr. Michael Brüggemann, technical support is provided by Remon Sadikni.
Disclaimer Ergebnisse und Grafiken des OMM dürfen zu nicht kommerziellen Zwecken frei genutzt werden unter der Angabe der Quelle. Die korrekte Zitation des OMM ist: **Brüggemann, Michael / Sadikni, Remon (2020): Online Media Monitor on Climate Change (OMM): Analysis of Global Tweets and Online Media Coverage. Universität Hamburg. URL: www.climatematters.de** (Zugriffsdatum) Bitte mit der folgenden Danksagung:	**Disclaimer** Results and plots of OMM may be used for non-commercial purposes without restrictions by providing this citation: **Brüggemann, Michael / Sadikni, Remon (2020): Online Media Monitor on Climate Change (OMM): Analysis of Global Tweets and Online Media Coverage. Universität Hamburg. URL: www.climatematters.de** (date of access) Please, with the following acknowledgments:

Die Daten wurden von ICDC, CEN, Universität Hamburg automatisiert erhoben, statistisch ausgewertet und herausgegeben.	These data were gathered, provided and evaluated by the ICDC, CEN, University of Hamburg.

Appendix

The article search strings as regular expressions, the bold ones were added July 22, 2021:

```
english:
global warm OR
climat.{1,2} chang OR
climate crisis OR
climate emergency

german:
Klimawandel OR
Erderwärmung OR
global.{1,2} Erwärmung OR
Klimaerwärmung OR
Klimaänderung OR
Treibhauseffe OR
Klimakrise

norwegian:
klimaendring OR
miljøkris OR
global oppvarming OR
miljøendring OR
```

drivhuseffekt OR
klimakris

spanish:
cambi.{1,3} clim OR
calentamient.{1,3} global OR
efect.{1,3} invernader OR
crisis climática OR
emergencia climática

portuguese:
mudan.{1,4} d.{1,2} clim OR
mudan.{1,4} clim OR
altera.{1,4} clim OR
aqueciment.{1,3} global OR
efeito estufa OR
efeito d.{1,2} estufa OR
crise climática OR
emergência climática

swedish:
klimatförändring OR
global.{1,4} uppvärmning OR
klimatkris

italian:
cambiamento climatico OR
cambiamenti climatici OR
cambiamento del clima OR
riscaldamento climatico OR
surriscaldamento climatico OR
riscaldamento globale OR
riscaldamento del pianeta OR
caos climatico OR

```
catastrofe climatica OR
disastro climatico OR
effett.{1,4} serra

dutch:
opwarming van de aarde OR
klimaat.{1,6}verandering OR
opwarming van het klimaat OR
verandering.{1,4} van het klimaat OR
verandering.{1,4} in het klimaat OR
broeikaseffect

french:
chang.{1,5} clima OR
r.{1,2}chauf.{1,6} glob OR
effet de serre

polish:
globaln.{1,4} ociepleni OR
zmian.{1,4} klimatu OR
katastrof.{1,4} klimatu OR
efekt.{1,4} cieplarnian

russian:
Изменени.{1,3} климат OR
глобальн.{1,3} потеплен OR
потеплен.{1,3} климат OR
климатическ.{1,3} катастроф OR
парников.{1,3} эффект

turkish:
iklim Değişikliği OR
küresel Isınma
```