Embedded Human Computation for Knowledge Extraction and Evaluation (uComp)

D1.3: Domain-Specific Content Repository

30 June 2014
Version: 1.0

Version History

Version	Date	Author	Comments
0.1	02/05/2014	A. Scharl	Initial Draft Document
0.2	20/05/2014	A. Scharl	Major Revision
0.3	02/06/2014	H.-P. Lang	Content Statistics and Input Filter
1.0	30/06/2014	A. Scharl	Final Revision and Layout

Dissemination Level: PU – Public

Table of Contents

Overview

The domain-specific content repository described in this deliverable is a collection of unstructured, domain-relevant content in three target languages (EN, DE, FR). It is complemented by an RDF triple store to manage associated metadata, particularly the information about named entities that were detected in the unstructured corpus (locations, persons, and organizations). An Application Programming Interface (API) allows exporting data in either JSON or XML format, for reusing the gathered social media content in third-party applications.

Data Acquisition Services

To compile the domain-specific content repository outlined in the previous sections, we used the *extensible Web Retrieval Toolkit* (eWRT),[1] a modular open-source Python API to retrieve social data from Web sources such as Delicious, Flickr, Yahoo! and Wikipedia, including various helper classes for effective caching and data management (Weichselbraun et al., 2013). The toolkit is developed by MOD and a group of academic partners. In addition to components for content acquisition and caching, it also provides low-level natural language processing functionalities such as phonetic string similarity measures and methods for string normalization.

The uComp project, as part of D1.2, has also developed the TwitIE open-source plugin for pre-processing unstructured social media content (Bontcheva et al., 2013),[2] which is being distributed as part of GATE (gate.ac.uk). TwitIE provides a tweet tokenizer (identifies individual words, user names, hashtags, URLs, etc), language identification components, part-of-speech tagger, tweet normalizer, and English named entity recognizer.

Unstructured Content

T1.5 has built a comprehensive repository of relevant climate change content, disambiguated and managed by means of a PostgreSQL relational database management system. The repository is accessible via the dashboard of the *Media Watch on Climate Change*[3] hosted by MOD (see Figure 1), which comprises terabytes of unstructured Web content classified by topic and location. uComp has helped to significantly increase the scope of the content repository. Besides adding new English sources, German and French versions of the *Media Watch on Climate Change* have been developed. Publicly released in June 2014, they are available at the following addresses:

- www.ecoresearch.net/climate/de
- www.ecoresearch.net/climate/fr

---

[1] www.weblyzard.com/ewrt
[2] gate.ac.uk/wiki/twitie.html
[3] www.ecoresearch.net/climate

The official launch of the German and French versions was accompanied by a press workshop held at MOD, which attracted considerable media interest and resulted in a number of articles in major Austrian news outlets (ORF Science, Kurier), as well as a radio interview (ORF OE1).
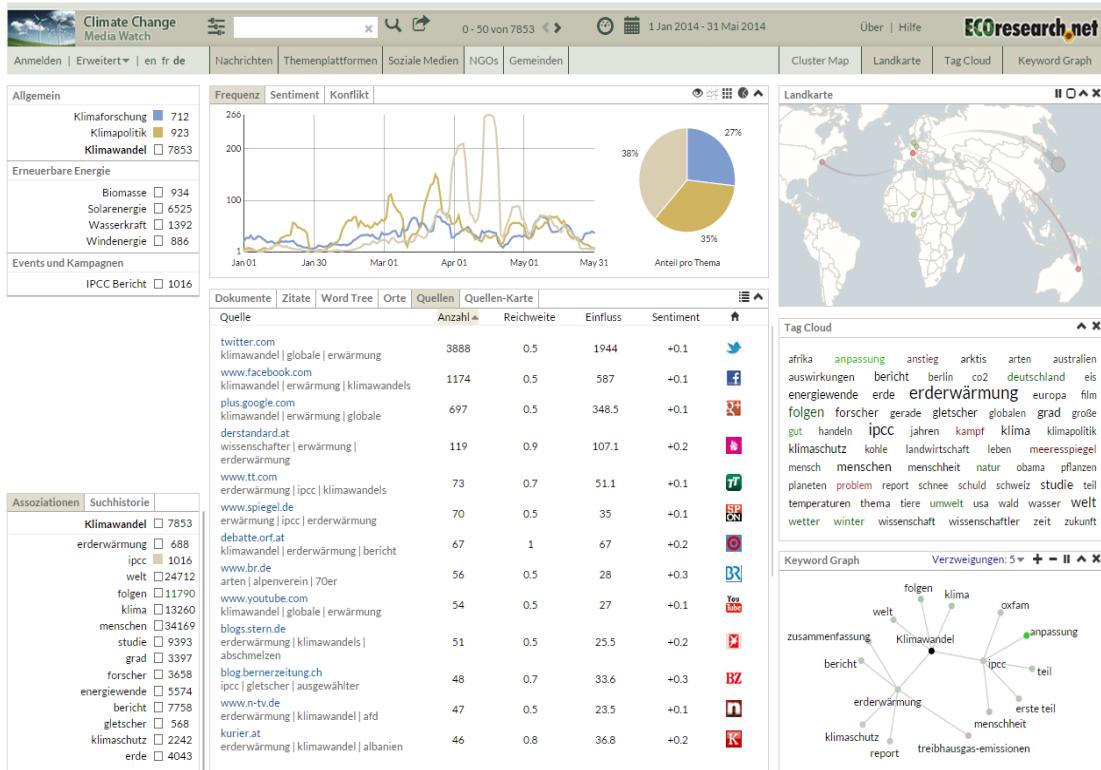

Figure 1. Media Watch on Climate Change screenshot with German news and social media sources reporting about "klimawandel" (= climate change) between January and May 2014

Input Filter Definition

In a joint effort among all consortium members, detailed sets of regular expressions were developed to serve as the input filter in the three target languages. The peculiarities of the APIs of social media platforms required to distinguish:

- *Unique Single Words.* Social media sources will be searched for these words and will be added to the Media Watch on Climate Change *without* any further relevance checks.

- *Ambiguous Single Words.* Social media sources will be searched for these more general terms, but will then be subjected to the white list of the appendix (see below). This two-step process ensures high recall (capturing a large number of documents in phase 1) and high precision (filtering out irrelevant documents using the white list in phase 2).

Only articles and postings matching the following list of regular expressions are being added to the Media Watch on Climate Change. The white list regular expressions are applied to most documents (news, Web sites, as well as social media content stemming from the search for "ambiguous single words"). The only exceptions are

unique single words from social media, if the option to add all documents from a specific source is chosen. Please refer to the appendix for the list of regular expressions to define the input filter.

Information Sources

Between January and June 2014, the data acquisition services for the Media Watch on Climate Challenge gathered more than **4.75 million climate-related documents** (English: more than 4 million; German: 500,000; French: 285,000).

The **German and French content streams** comprise the following sources (to provide the NGO tab, we crawl the multilingual ch.wwf.ngo and climate.unfccc samples and use language detection algorithms to assign mirrored documents to the correct monolingual corpus):

- **GE News Media:** CH (244), AT (172), DE (63).
- **GE Social Media:** Twitter, Facebook, Google+ and YouTube postings.
- **GE NGOs:** ch.wwf.ngo (197); climate.unfccc (1250).
- **GE Municipalities:** CH (2318), AT (616), DE (753).
- **FR News Media:** CH (244), FR(67)
- **FR Social Media:** Twitter, Facebook, Google+ and YouTube postings.
- **FR NGOs:** fr.ngos.national (9), fr.ngos.regional (20), climate.unfccc (1250)
- **FR Companies:** fr.cac40 (39), fr.cac.next.20 (20),

This complements the sources of the **English portal**, which has also been extended and improved in terms of disambiguation and content cleanup:

- **EN News Media:** 163 (US, CA, UK, ZA, AU, NZ).
- **EN Social Media:** Twitter, Facebook, Google+ and YouTube postings.
- **EN NGOs:** ch.wwf.ngo (197); climate.unfccc (1250); us.npo (32)
- **EN Fortune:** Top 1000 U.S. companies in terms of revenue.

In the second quarter of 2014, we have also started gathering climate change related content of the Web sites of the European Environment Agency (EEA),[4] the British Library[5], the World Bank's Connect4Climate initiative,[6] and the 13 U.S. government agencies participating in the Global Change Research Program.[7] This complements a continuously extended archive compiled by crawling the Web sites of the National Oceanic and Atmospheric Administration (NOAA)[8] and its various partner organizations.

---

[4] www.eea.europa.eu
[5] www.bl.uk
[6] www.connect4climate.org
[7] www.globalchange.gov
[8] www.climate.gov

Structured Content

The initial version of the structured repository contains datasets from multiple sources: DBpedia[9], Geonames[10], and Freebase[11]. Some of these datasets are simply dumps of the triples, while others have required a more elaborate process including slicing, cleaning and fixing steps.

- For **Geonames**, it was enough to use the dumps generated by Python scripts, as it contained only geographical entities (Places), and hardly any embedded links. There was no need to remove any links from this dataset. Approximate number of triples used: 155,5 million.

- Each **DBpedia** dataset contains all the triples related to a single entity type in a single language (sliced by language and entity type). We used RDFSlice[12] and a series of bash scripts to create the datasets and upload them to a Sesame triple store. We removed all the triples that contained bad encodings, as these many of these links were typically related to links to articles from the web that we were not interested in. We determined the broken links by trying to upload several filecd s, and then we added the bad cases in a bash script that automatically removed all the similar links from all the files. All DBpedia datasets were based on the recent 3.9 version, which also uses a type inference mechanism to check data types (Paulheim and Bizer, 2013). We have created separate datasets for the following languages and entity types:

    - **English** (people: 50 million triples, organizations: 20 million triples, places: 53.5 million triples),

    - **French** (people: 8,5 million triples, organizations: 2 million triples, places: 11 million triples),

    - **German** (people: 12 million triples, organizations: 2.5 million triples, places: 12.5 million triples).

- For **Freebase**, the BaseKB[13] build was provided in a cleaned and preprocessed state. Approximate number of triples used: 337 million.

We removed malformed URLs, bad formatting or encoding, and some of the internal links that were not needed (cleaning and fixing steps). While it can be argued that these pre-processing steps might cause some loss of contextual information, we have not removed any disambiguation links or references to other DBpedia-supported languages.

---

[9] wiki.dbpedia.org/Downloads2014
[10] www.geonames.org/
[11] www.freebase.com/
[12] www.aksw.org/Projects/RDFSlice.html
[13] www.basekb.com

Conclusion

The domain-specific and multi-lingual content repository outlined in this deliverable provides an abundant source of raw data for uComp's factual and affective knowledge extraction services (Scharl et al. 2014; Weichselbraun et al. 2014). The launch of the German and French versions of the Media Watch on Climate Change has been accompanied by a press workshop held at MOD, which attracted significant media attention.

Scalable data acquisition and pre-processing methods have been deployed and will be continually refined and aligned with semantic methods to integrate, annotate and classify the gathered content. In close collaboration with the DecarboNet project (www.decarbonet.eu) and its semantic enrichment algorithms, we will use extra-linguistic knowledge to enrich the repository with domain-specific climate change information, and provide additional mechanisms to query and analyze the repository.

References

Bontcheva, K., Derczynski, L., et al. (2013): An Open-Source Information Extraction Pipeline for Microblog Text. *Int'l Conference on Recent Advances in Natural Language Processing* (RANLP-2013). Hissar, Bulgaria. 83-90.

Paulheim, H. and Bizer, C. (2013). "Type Inference on Noisy RDF data," The Semantic Web - *12th International Semantic Web Conference (ISWC-2013),* Proceedings, Part I. Sydney, NSW, Australia: Springer. 510–525.

Scharl, A., Kamolov, R., Fischl, D., Rafelsberger, W. and Jones, A. (2014). Visualizing Contextual Information in Aggregated Web Content Repositories. *9th Latin American Web Congress* (LA-WEB 2014). J.M. Almeida et al. Ouro Preto, Brazil: IEEE Press: 114-118.

Weichselbraun, A., Gindl, S. and Scharl, A. (2014). "Enriching Semantic Knowledge Bases for Opinion Mining in Big Data Applications", *Knowledge-Based Systems,* 69: 78-86.

Weichselbraun, A., Scharl, A. and Lang, H.-P. (2013). Knowledge Capture from Multiple Online Sources with the Extensible Web Retrieval Toolkit (eWRT). *7th International Conference on Knowledge Capture (K-CAP 2013).* Banff, Canada: ACM: 129-132.

Appendix - Input Filter

English List of Regular Expressions

#AR5, #COP18, #UNFCCC, #apcc, #earthhour, #ipcc, ((dissipation|wast(e|ing|age) of)?energy(dissipation| source)?, ((switch|turn) (off|on)|shut down), (@|#)?earth hour, (@|#)?earthhour, (Solar|milankovitch)(-|)cycles?, (ant)?arctic(a|al)?(change)?, (awareness raising|sensiti(s|z)(ing|ation)), (city |municipality)?(-|)?rating, (climate|climate change(s)?)?), (climate|climate change) (variability|feedback|extremes?|anomaly|anomalies|variability| predictability|oscillation|impact| observations?|monitor|models?|assessment|science|research|predictions?|projections?|forecasts?|policy|adaption|resilient|vulnerability|mitigation|scenarios?), (company|enterprise|distributor, (crop)?(grow(ing)?|cultivate), (endangered|threatened) species, (energy|fuel)(-|)efficien(t|cy|cies), (environmental protection|action for the environment), (environmental)?(-|)?criteria, (kyoto|montreal) protocol, (natural|renewable|alternative|green) ((re)?sources?|energ(y|ies)|power), (natural|renewable|alternative|green) (re)?sources?, (natural|renewable|alternative|green) energ(y|ies), (natural|renewable|alternative|green) power, (non-?)?renewables?, (pure |mono) (culture|cropping|plantation), (road |street |public)?lighting(s)?, (street|road) lighting, 1990(-|)levels, @ClimateCoLab, @NASA_EO, @NOAAClimate, @UN_ClimateTalks, @YaleClimateComm, Annex(-|)I(-|)(country|countries|party|parties), Anthropogenic, Biofuel, Biomass, CDI, COP ?(14|15|16|17|18|19|20|21), CRT, Clean Development Mechanism, Climate Data Initiative, Climate Resilience Toolkit, ClimateCoLab, EH14, EH2014, EartHour2014, EarthHour14, EarthHour2014, Europe(-|)2020, Footprint, Holdren, Market Transformation Initiative, NASA, NASA_EO, NOAA, NOAAClimate, Non(-|)Annex(-|)I, Podesta, REDD, RSPO, Rio\+20, UNCSD, UNFCCC, UN_ClimateTalks, YaleClimateComm, acidif[^]+, actions?, aerosol, agenda|agreements?|summits?, al gore, albedo, alternative energy, anthropogenic, apcc, arctic change, atmosphere, awarenes raising, bio(-|)fuel, bio(-|)fuel(s)?, biodiversit(y|ies), biodiversity, biofuels, cap and trade, carbon, carbon (offset|footprint|tax|sequestration|emission| reduction|trade|capture|storage), carbon (offset|trading|reduction|emissions?), carbon cycles?|carbon budget|carbon sources?|carbon sink|biomass burning|carbon tax(es)?|fuels tax(es)?, carbon dioxide, carbon dioxides?|co2|chlorofluorocarbon|methane, carbon finance, certification system, certified(palm(-|)oil)?, city challenge, city challenge(s)?, clean-technolog(y|ies)?|clean technolog(y|ies)?, climate, climate (variability|feedback|extremes?|anomaly|anomalies|variability|predictability| oscillation|impact| observations?|monitor|models?|assessment|science|research|predictions?|projections?|forecasts?|policy|adaption|resilient|vulnerability|mitigation|scenarios?), climate change, climate changes?, climate forcing|climate feedback|climate extreme|climate anomaly|climate predictability|climate observations|climate monitoring|climate assessments?, climate impact|climate research|climatology| climate data reanalysis|climate models?|climate services?, climate killer, climate systems?|earth systems?|climate cycles?, climate talks?, climatecatastrophe, climatechange, climatedebate, climatekiller, climatelie, climatepolitics, climateresearch, climatolog[^]+, climatology, co2, co2(-|)(equivalent|eq), coal|soot|mountaintop removal mining, conservation, consume behaviour, copenhagen(-|)summit, coral bleaching, coral bleaching|coral reef death, crop growing, crowdfounding, crowdsourcing, cultivate, cultivation, decarbonet, deforestation, desertif[^]+, desertification, earth hour, earth hour(blue)?, earth summit, earthhour, earthhour.org, eco(system(s)?|log(y|ies|ical|ists?)), ecolog(y|ies|ical|ists?), ecological, ecologically, ecology, ecoresearch, ecosystem, ecosystems, ecosystems?, el ni(n|ñ)o, el niño|enso|la niña|north atlantic oscillations?|pacific decadal oscillations?|madden-julian oscillations?|arctic oscillations?|antarctic oscillations?, emission, emission (caps?|reductions?), emission scenarios, emission(s)? (caps?|reductions?), emission(s)? trading, emissions, emissions?, endangered species, energ(y|ies)?, energy, energy conservation|hybrid cars?|energy star, energy cycles?|energy budget|energy sources?|energy sink|radiant energy|albedo|shorwave radiation|longwave radiation|thermal emissions?, environment, environmental, environmental damage, environmental(ist(s)?)?, environmental(ists?)?, environmentalist, environmentally, environments?, epa|ipcc|noaa|intergovernmental panel on climate change|national

oceanic and atmospheric administration, famine|famine early warning systems?|crop failure|vegetation index anomaly, flexibility mechanisms?, floods?|environmental health|heat wave|heat stress|urban heat island|temperature anomaly|heavy rain|heavy rain event|heavy downpour|torrential rain|monsoon, fossil, fossil fuel(s)?, fossil fuels?, fossile, fracking, freak weather, garbage, geotherm(al|ic), geothermal, geothermic, glaciers?|snow cover|permafrost, global cooling, global deal|dangers?|rainfall|weather| temperatures?|scientists?, global warming, global warming lie, globali(z|s)ation, greenhouse, greenhouse (emissions?|effects?), greenhouse effect, greenhouse gas(es)?, greenhouse?, greenpeace, hydroelectricity, hydropower, i will if you will challenge, ice sheet|greenland|antarctica?|iceberg|ice age|interglacial|warm period|ice melt|glacial retreat|ice thinning|melting day anomaly, impacts?, inconvenient truth, ipcc, iwiyw, kyoto protocol, la ni(n|ñ)a, land cover change, land degradation, little ice age, monocropping, monoculture, natural gas, natural gas|oil|petroleum, nature, non-renewable, nuclear power, ocean acidification, oil spill(s), orangutan, ozone (hole|layer|depletion), palm kernel oil, palm oil, palm oil purchase, palm(-|)?oil, palm(-|)oil(-|)?(acquisition|purchase|supply|purchasing), palmoil, photovoltaic, photovoltaics, plantation, plantation(s)?, plantations, polar ice caps, pollutant(s)?, pollutants?, pollute, pollution, pollution|impact|cost|threat| disasters?|vulnerability|devastation, public lighting, pure culture, recycl(e|ing|ed|es), recyclability, recycle, recycling, renewable, rio+20, road lighting, round(-|)?table, roundtable, roundtable on sustainable palmoil, rspo, sea level rise, sea level|subsidence|storm surge|snow cover|snow melt|snow cover extent|shore protection, seasonal shift|spring arrival|fall arrival|growing season|winter arrival|winter ending, sensitising, sensitizing, solar, solar cycles?|sunspot|solar irradiance|solar maximum|solar minimum|sunlight|shortwave radiation, solar panel|solar cell|photovoltaic|solar energy, street lighting, sustainab(le|ility), sustainability, sustainable, sustainable development, sustainable development|sustainable practices|urban growth|urban planning|recycling|compact fluorescent|resources? management, switch light(s)? (off|on), temperature anomaly, threatened species, toxic waste, tropical, tropics, tropospheric ozone|carbon monoxides|nitrous oxides?, unfccc, volcanic eruption|volcanic plume|sulfur|sulfate|outgassing|agung|el chichon|pinatubo, waste (treatment(s)?|management|prevention|separation), waste treatments?, water cycles?|water budget|water sources?|water sink|water vapor|evaporation|convection|evapotranspiration| precipitation, wind farm|wind turbine|hydropower|geothermal energy|biofuels?|synthetic fuels?|ethanol, world wide fund for nature, wwf int., wwf(int(.|ernational))?

French List of Regular Expressions

#AR5, (((N|n)on(\s\t\n)+)?renouvelable(s)?), ((é|e)mission(s)?(\s|\n|\t)+à(\s|\t|\n)+)
?effet(\s|\t\n)+de(\s|\t|\n)+serre, (L|l)\, (R|r)essource(s)?(\s|\t|\n)+(naturelle(s)?|((non(\s\t\n)+)
?renouvelable(s)?)), (e|é)cosystème, (e|é)sp(è|e)ce(\s|\n|\t)+(menac(é|e)e|en(\s|\t|\n)+danger|en
(\s|\t|\n)+voie(\s|\t|\n)+(de(\s|\t|\n)+disparition|d,
(gaz(\s|\n|\t)?à(\s|\n\\t)+effet(\s|\n|\t)de(\s|\n|\t)?serre|GES,
(protocole(\s|\t|\n)+de(\s|\t|\n)+Kyoto|(P|p)rotocole(\s|\t|\n)+de(\s|\t|\n)+Montr(é|e)al) protocol,
(é|e)colog(ie(s)?|ique(s)?|iste(s)?|isme), (é|e)cosystème(s)?,
(é|e)fficacité|(E|e)fficience)(\s\n\t)+(é|e)nerg(é|e)tique, (é|e)mission(s), (é|e)mission(s)?(\s|\t|\n)+(d,
(é|e)mission(s)?(\s|\t|\n)+d, (é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(chlorofluorocarbure(s)?|CFC),
(é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(dioxyde((\s\t\n)+de(\s\t\n)+carbonne)?|CO2|gaz(\s\t\n)+carbonique)
, (é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(fréon|HCFC-22),
(é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(gaz(\s|\n|\t)?à(\s|\n\\t)+effet(\s|\n|\t)de(\s|\n|\t)?serre|EGES,
(é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(méthane|CH4),
(é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(protoxyde(\s|\t|\n)+d,
(é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+(tétrafluorométhane|CF4),
(é|e)mission(s)?(\s|\t|\n)+de(\s|\n|\t)+vapeur(\s|\t|\n)+d,
(é|e)nergie(s)?(\s|\t|\n)+(naturelle(s)?|((non(\s\t\n)+)?renouvelable(s)?)|alternative(s)?|verte(s)?),
@APEnvironnement, @Actuenviro, @AreneIDF, @CDURABLE, @Ecologie_FR, @PlaneteUrgence,
@Sciencesenviro, @ademe, @amisdelaterre, @ecoactivites, @ecoloinfo, @greenpeacefr,
@humaterra, @naturavox, @reforestaction, @reporterdespoir, @sdnfr, CCNUCC, Eolien, apcc,
bio((\s|\t|\n)*-(\s|\t|\n)*)?diversité(s)?, biocarburant(s)?, biomasse,
changement(s)?(\s\t\n)+climatique(s)?, changements? climatiques?,
conservation(\s\n\t)+de(\s\n\t)+la(\s\n\t)(nature|bio((\s|\t|\n)* -(\s|\t|\n)*)?diversité(s)?),
conservation|(P|p)r(é|e)servation, convention(\s|\t|\n)+de(\s|\t|\n)+rio,
convention(\s|\t|\n)+de(\s|\t|\n)+vienne, d(é|e)chets(\s\t\n)+(T|t)oxiques,
d(é|e)veloppement(\s\t\n)durable, durabilité, effet de serre, environnemental(isme|iste(s)?), géothermie,
hydroélectrique(s)?, ipcc, land degradation, mar(e|é)e(s)?((\s\t\n)*-(\s\t\n)*)?(n)oire(s)?,
perte(s)?(\s|\t|\n)|((T|r)ou(\s\n\t)+dans(\s|\n|\t)+ la(\s|\t|\n)+)?couche)+d, photovoltaïque(s)?, polluant(s)?,
r(é|e)chauffement(de la planète| de la planete| de la terre| de l, recycl(er|é|ée(s)?|age(s)?),
refroidissement(de la planète| de la planete| de la terre| de l, réchauffement(s)?(\s\t\n)+climatique(s)?,
réchauffements? climatiques?, solaire, te(-)?)?traitement(\s\t\n)+(des|de)(\s\t\n)+d(é|e)chet(s)?

German List of Regular Expressions

#AR5, #TEDxWWF, #oShoot, (Kohlen(stoff)?(dioxid)?|co2)-(abgabe|sequestrierung| (ae|ä|e)quivalent(e)? |-eq(.)?|emission(en)?|fussabdruck|reduktion|handel|speicher(ung|n)?), (REDD|Reducing Emissions from Deforestation and Degradation), (RSPO|roundtable on sustainable palmoil), (Solar|milankovitch)(-|)zykle(n)?, (ant)?arktis(ch)?, (aus|ab|an)schalten, (city|gemeinde)(-|)?challenge(s)?, (denkmal|strassen) (-|)?beleuchtung(en)?, (erd|natur)gas, (gemeinschaft|community), (kyoto|montreal)(-|)protokoll, (müll|kehrricht|abfall)(vermeidung| trennung|verwertung), (nat(ue|ü)rlich|erneuerbar|alternativ| gr(ü|ue)n)e? (quellen?|energien?|ressourcen), (nat(ue|ü)rlich|erneuerbar|alternativ|gr(ü|ue)n)e? (quellen?|energien?| ressourcen?), (nicht-?)?erneuerbare?, (schweizer)?(firmen|unternehmen), (umwelt(-|)?)?tip(p)?s, (umwelt)?(-|)?kriterien, (ö|oe)ko(system|logie|logisch|logen?), (ö|oe)lpest, 1990(-|)level, @TEDxWWF, Alternative energie, Alternativenergie, Annex(-|)l(-|)(l(ä|ae)nder|staat(en)?|partei(en)?), Atmosph(ä|ae)re, Atmosphäre, Atom(-|)?strom, Atomenergie, Atomstrom, Ausstoß, Biodiesel, Biodiversit(ae|ä)t, Biodiversität, COP ?(14|15|16|17|18|19|20|21), Clean Development Mechanism, EH14, EH2014, EartHour2014, Earth hour(blue)?, EarthHour14, Emission, Emissionen, Emissions(-)?handel, Emissions(rechte|(-)?)handel, Energie, Energieeffizienz, Energieträger, Erdgipfel, Erwärmung, Europa(-|)2020, Footprint, Fussabdruck, Globalisierung, Globalisierungsdruck, Klima, Klima(-|)?(variabilit(ae|ä)t| feedback|extreme|anomalie(n)?| oscillation|vorhersage(n)?|beobachtung(en)?|monitor|model(e)? |einsch(ae|ä)tzung|wissenschaft(en)?| forschung|prophezeihung(en)?|politik|adaption| katastrophe(n)?|milderung|scenari(en)?), Klima(wandel)?, Klimakiller, Klimalüge, Klimatologie, Klimawandel, Klimaänderung, Kohlenstoff, Kohlenstoffdioxid, Kopenhagen(-|)Gipfel, Mark(e)?t Transformation(s)? Initiative, Meeres(-)? höhe(n)?(-)?anstieg, Monokultur(en)?, Müll, Nachhaltigkeit, Natur, Non(-|)Annex(-|)l, Ozon, Ozongas, Ozonsmog, Palmkernöl, Palmöl, Pflanzenöl, Recycelbarkeit, Recyclebarkeit, Recyclierbarkeit, Recycling, Scorecard(s)?, Temperatur(-|)?anomalit(ä|ae)t, Treibhaus, Umwelt, Umwelt(-aktion|schutz), Umwelt- aktion, Umweltbeobachtung, Umweltbilanz, Umweltschutz, Verbrauch(en)?, Zertifizierung(ssystem)?, abfall(vermeidung|trennung|verwertung), abholz(en|ung), abholzen, abholzung, albedo, anthropogen, anthropogen(e(s)?|ische?), anthropogenische?, apcc, bedrohte art(en)?, bedrohte(n)? art(en)?, bedrohten arten, bio(brennstoff(e)?|krafststoff(e)?| treibstoff(e)?), biobrennstoff, biokraftstoff, biomasse, biotreibstoff, cap and trade, carbon finance, co2, desertif[^]+, desertifikation, detailhandel, earth hour, earthhour, earthhour.org, el ni(n|ñ)o, emission(en)?, emission(en|s)?(beschr(ä|ae)nkung(en)?| reduktion(en)?), emissions(-)?szenari(o|en), energie(-|)?effizien(z|te?r?), energie(-|)?effizien(z|te?r?s?n?), energie(verschwendung|träger), energieeffizient, energieeffiziente, energieeffizienten, energieeffizienter, energieeffizientes, energieverschwendung, erdgas, erneuerbar, erneuerbare, extremes wetter, flächen(-)?wandel, fossile, fossiler (energieträger|brennstoff), fossiler brennstoff, fracking, gemeinde((-|)rating)?, gemeinde-rating, gemeinderating, geothermal, giftige abfälle, giftige(r)? abf(a|ä)lle?, giftiger? abf(a|ae|ä)lle?, global, globale abkühlung, globale erw(ae|ä)rmung, globale erwärmung, greenpeace, i will if you will, ipcc, iwiyw, kleine eiszeit, klima(|-)(gespr(ä|ae)che| debatte(n)?), klima-killer, klimadebatte, klimaforschung, klimagespräche, klimakatastrophe, klimakiller, klimalüge, klimapolitik, klimatologie, kohlendioxid, korallen bleichung, la ni(n|ñ)a, land degradation, licht löschen, meer(es)?versauerung, meereshöhenanstieg, monokultur, monokulturen, nachhaltig, nachhaltig(keit)?, nachhaltige, nachhaltige entwicklung, nachhaltigkeit, nicht-erneuerbar, orang-utan, orangutan, ozon(-|)?(loch|schicht|schwund), pakt mit dem panda, palmöl-beschaffung, palmölbeschaffung, photovoltaik, polare Eiskappe(n)?, polare eiskappen, re(c|z)y(c|k)(e)?l(n|e|ing|ed|es|ieren), recyceln, recycl(e|ing|ed|es), recyclen, rio+20, roundtable, roundtable on sustainable palmoil, rspo, scorecard, scorecards, strassenbeleuchtung, temperatur anomalität, treibhaus(-|)(effekt(e)?)?, treibhaus(-|)effekte?, treibhaus(-|)gase?, treibhauseffekt, treibhausgase, tropen, tropisch, umwelt(schutz)?, umweltbelastung(s|spunkte), umweltbewusst, umweltfreundlich, umweltgefährdend, umweltsch(ue|ü)tzer(in)?, umweltschutz, umweltschützer, unfccc, verantwortungsvoll(e)?(r)?, vers(a|ä|ae)uer[^]+, verschmutzen, verschmutzung, verschmutzung(en)?, wasserkraft, wirtschaft, world wide fund for nature, wwf, wwf (international))?,(palm(ö|oe)l)?, zertifizierungssystem, Ökologie, Ökosystem, ökologisch