

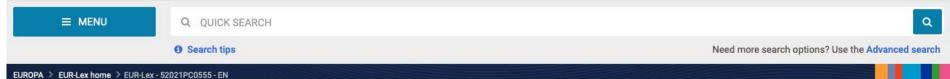
DIS18: Green Deal Annotations

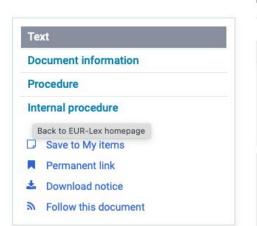
02: Use-Cases

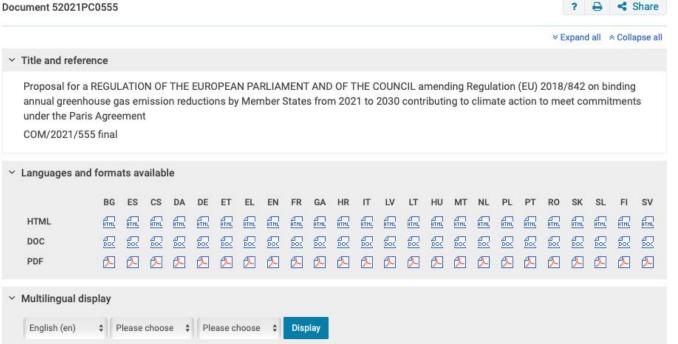
Philipp Schaer, Björn Engelmann, and Meik Bittkowski Technische Hochschule Köln, Cologne, Germany Technology Arts Sciences TH Köln











Let's break out

We will mix you in groups of 3-4 people in Zoom Breakout Rooms

You get 5 minutes to talk about the following topics

- Describe one Snorkel detail that worked out really good.
- What was a disaster while doing the Snorkel tutorial?
- While scanning through the EU documents, what did you find interesting or worth investigating?

After the breakout session each group has to report on the results of your discussions.

Jamboard for notes:

https://jamboard.google.com/d/1W08s9nz9PkcKyhpA7v5aTs4zkjl

le processing (I) Konzept Souse/ Blocment Seprente PUT Extract - HTML?

De-Concat Text Sild Tabelle PDF- linker & "Kontext"/Passajano Topic us Tein IX We did the same... B Doly - La lossy hadrian (III) La Evaluer 3/Mariler Ashanj heiten (zeitlich, thouasish ...) (E) I abollan - Extiation (A) Link-Generiel-1 Projoht man agmant @ Zahlen Extraction (5) Glosar/Honof Pattern Dalenmodellier Founde Le

Overview on Use-cases

- Link generation
- Document classification
- Number/NER extraction and linking
 - Stakeholder / Dates / Budget
- Glossary
 - local vs. global definitions
- Table extraction

Link generation

Many documents contain references to other documents or sections

- Find these references
- Automatically create a link to the referenced source

Link generation

sustainability potential, such as sustainable aviation fuels produced from was (feedstock listed in Annex IX Part B of RED II), should be eligible in order to la market and allow for emissions reductions already in the short term. Whereas such a is expected to reduce substantially emissions from the transport sector as a whole, at the large majority of liquid biofuels are produced for the road sector. Therefore, a feedstock for production of biofuels in the road sector to production for the aviation may occur, but would be limited. Indeed, it estimate that the potential displace biofuels produced from Annex IX Part B of RED II would be around 3.2% by 2030 5.

Document classification

It is helpful if documents, sections or passages are categorized

- Define categories based on the corpus that allow a reader to quickly classify the text (law, report, target statements)
- Classify entire texts or individual sections of the document using these categories

Document classification

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

on ensuring a level playing field for sustainable air transport

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS

A strategic rollout plan to outline a set of supplementary actions to support the rapid deployment of alternative fuels infrastructure

https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0561&from=EN https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021DC0560&from=nl

Number extraction

Numbers often provide important information such as deadlines or specific costs

- Find relevant numbers in the documents
- Connect the numbers with the related information

Number extraction

Overall POs C1 and C2 lead to an increase of the costs of €20.3 billion (C1) and €14.6 billion (C2) relative to the baseline over the period 2021 to 2050. These costs are largely driven by an increase in jet fuel cost relative to the baseline, i.e. €103.5 billion (C1) and €88.2 billion (C2), expressed as present value over 2021-2050. The increase in fuel costs is reflected on air fares, which are estimated to increase by around 8.1-8.2% by 2050. Higher air fares lead to a

Glossary

A global glossary helps a reader better understand the text, as many concepts are not directly understandable

- Find explanations of concepts in the documents
- Build a glossary that connects concepts and explanations

Glossary

- 'aircraft operator' means a person that operated at least 729 commercial air transport flights departing from Union airports in the reporting period or, where that person may not be identified, the owner of the aircraft;
- 'commercial air transport flight' means a flight operated for the purposes of transport of passengers, cargo or mail for remuneration or hire, or business aviation flights;
- 'aviation fuel' means the fuel manufactured for direct use by aircraft;

Table extraction

Tables contain structured information that allows a reader to get an overview of interrelationships

- Find tables in the documents
- Extract the relational information

Table extraction

1.4.4. Indicators of performance

Specify the indicators for monitoring progress and achievements.

Objectives	Indicators	Measurement tools/data sources
Reduce GHG emissions	Level of emissions in the EULevel of emissions globally	Emission statisticsSector statistics
Incentivise cleaner production processes in third countries	 Evolution of actual emissions for CBAM sectors in third countries 	 Level of emissions demonstrated by third country producers subject to CBAM
Prevent carbon leakage	 As indicators of GHG emissions above Level of emissions in the EU relative to level of emissions globally Trade flows in CBAM sectors Trade flows downstream 	Emission statisticsTrade statisticsSector statistics

https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0564&from=en

Provisional Schedule for Winter Term

Date	Туре	Topics	
12.10.21	Zoom	Introduction and kick-off, organization paper distribution	
19.10.21	Zoom	Introduction to Snorkel and Label Studio	
26.10.21	Self-organized	Paper reading / Work on Snorkel and Label Studio tutorials	
03.11.21	Workshop@TH	Student talks on background	
09.11.21	Self-organized	Work on Snorkel and Label Studio tutorials	
16.11.21	Zoom	Introduction to the data sets	
23.11.21		Project week – no meeting	
01.12.21	Workshop@TH	Workshop: Usecase definitions and further planning	
07.12.21	Zoom	Work on prototype	
14.12.21	Zoom	Work on prototype	
21.12.21	Zoom	Work on prototype	
28.12.21		Christmas holidays	
04.01.22	Zoom	Work on prototype	
12.01.22	Workshop@TH	Student talks – Milestone presentation	

What to do until 1 December 2021

- Form an internal group structure
 - Project management?
 - Data collectors?
 - Preprocessing?
- Take some sample data or take the whole data set and
 - Preprocess the data accordingly (how?)
 - Develop a paper prototype/first demo for your specific usecase
 - Develop a concept that describes to problem and your idea to tackle it
 - Think about an evaluation strategy
- Prepare to pitch your concept/prototype on 1 December 2021
 - 2-3 min presentation/demo

Teams!

Group 1: Number/NER extraction and linking

Puddu, Blum, Kubon, Bateva, Wolf

Group 2: Document Linking

Nebatz, Jennert, Paramalingam

Group 3: **Document Classification**

Pycha, Hilbrecht, Schmer, Novak

Group 4: Table Extraction

Tanriver; Celik; Alarslan; Romer