Database contents

4 main technical parameters:

* Permeance (GPU)
* Selectivity CO2/N2
* Selectivity CO2/H2
* Selectivity CO2/H20

Secondary technical parameters:

* Area(s)
* Pressures of permeate and filtrate
* Feed composition

Economic parameters:

* Reactants
* Cost (hard to find)

Pipeline

Tuesday 12

1. Grobid to clean the PDF into XML format (not needed with CDE v2 that works on pdf)
2. Tabula (or other) to retrieve tabular data (also not needed with CDE v2 which handles tables)
3. ChemDataExtractor to scrape the paragraphs

Has issues on reading PDFs 🡪 use pdfminer to parse before treatment

Wednesday 13

1. CDE v2 works best on HTML or JATS files