# MICN

Tatiana Cogne
tatiana.cogne@epfl.ch

62020

## Contents

# 1 Week 8

## 1.1 Step 1

The objective for this first week is to scrape all HTML pages for the COPS from the IISD website [1]. For each COP we had to extract it's name, date and also location. After that looked at the issues for each of the COP and we collected their number, date and HTML link also on the website.

## 1.2 Details of the implementation and issues

To begin, we collected all the COP and the information about it. We defined a function extract-list-cops who takes as input only the html page processed with the library BeautifulSoup.
All the COP's name does not have the same structure. We had to be very specific to be able to extract each name.

— COP 1 — 28 March - 7 April 1995 — Berlin, Germany

— COP 12 - CMP 2 — 6-17 November 2006 — Nairobi, Kenya

Then to be able to process the string in the same way where all the information is contained we had to modify some strings (remove substrings).
For each COP we extracted the date from the string and convert it into an int.
For this, we created a function extract-date who help the extract the desired date and take only a string as input.
We chose to keep only the starting date of the event (day, month), the year but not the finishing date (more easy). At the end our function returns a list who contains for each COP:

1. The number
2. The date (tuple (day, month, year))
3. The place

After that we had to extract for each COP all the issues discussed during the event. We defined a function called extract-details-cops who also takes as input only the html page processed with the library BeautifulSoup. The challenge in this function was to find each COP and each issues related to it because the webpage's location where we need to collect information is a table with a lot of rows. Those rows contain : a) the COP's name and all the information needed in our first task (extract-list-cops) or b) the issue's number and information about it. But the rows are not connected, they are next to each other only. We had first to identify a row defined for an issue to be able to know in which COP it was discussed and the collect all the issues.
At the end, the function returns a list where each cell, who corresponds to a specific COP (ordered by date), contains a list for each COP where we have :

1. The issue's number
2. The date when it was discussed
3. The html link of the issue.

The two function are specific to the given website : ://enb.iisd.org/enb/vol12/ We had to study the structure of the webpage to be able to collect all the information wanted.

### 1.2.1 Comments

20 February :First meeting
- I will have to learn how to parse an HTML page and do HTTP request in python with multiple libraries.
- I also have to read documents

1. "Old wine in a new bottle? A time-dynamic analysis of country motives to mitigate global climate change". (Marlene's thesis)

2. "Inside and outside the climate negotiations: Contrasting networks of conference diplomacy reporting and media perception".(castro2019inside)
3. Slides for WTI seminar (castro-kamerer-seminar.pptx)
4. Mails

- Install c4science and read the guidelines for my code in python

# References

[1] International Institute for Sustainable Development https://enb.iisd.org/enb/vol12/