

Team ESNAware - Milestone 2

Introduction

Climate change impacts all of us and the world's population has a central role to play in the fight against it. For this reason, it is important to inform people about the different issues related to climate change and to improve engagement on the battlefield. We wish to implement a platform that allows users to have a global understanding of issues relating to climate change. We want to display information in a way that makes navigation through the issues easy and interactive. The idea would be to give them the ability to view basic data stories about each of them expliciting their causes, their actors, their consequences and the solutions being brought up. One of the aims of this project is therefore to raise awareness in a different and novel way.

Throughout this project, we are going to implement a platform that allows easy navigation through the interconnected, underlying issues of climate change and easy access to interactive information about the different issues. This project is intended for any person who is curious to know more about the topic and for people who have difficulties linking the dots and grasping the whole picture.

Visualizations

The data we are using is extracted from Wikipedia and is linked together using Natural Language Processing techniques. To display this linkage, we are going to build a knowledge graph, with nodes representing Wikipedia articles. The links between nodes are created and weighted using different methods:

- The position of the articles in Wikipedia's category tree. Each article belongs to a category and each category can be linked to other categories.
- Jaccard similarity between outgoing links on a Wikipedia page.
- Topic similarity using Latent Dirichlet Allocation (LDA)

The preprocessing is done using Python and the following packages: gensim, spacy, scikit-learn, networkx. Moreover, we started out by reading lecture 10 about network graphs and decided to use D3.js, and in particular D3's force layout graph.

The visualization comes in two different components : a node-link graph representing issues and their relationships (Figure 1), and a page used to display detailed information about an issue (Figure 2). Initially we are faced with just the graph spread out across the page, but upon clicking a node its information page will pop up and cover about 75% of the screen.

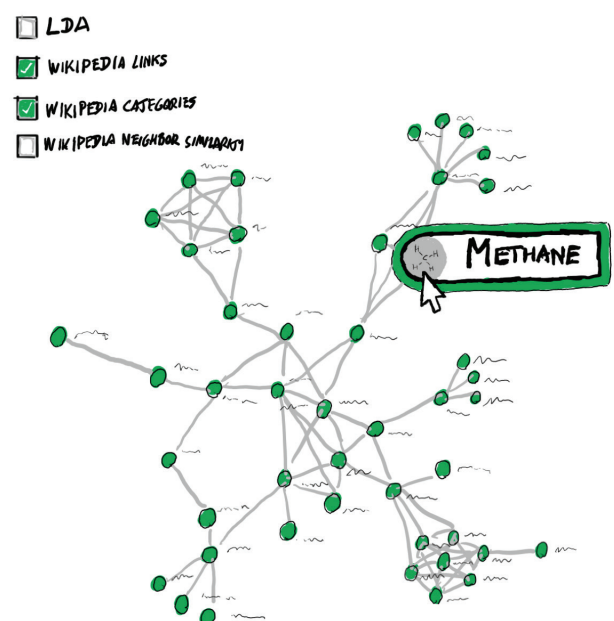


Figure 1

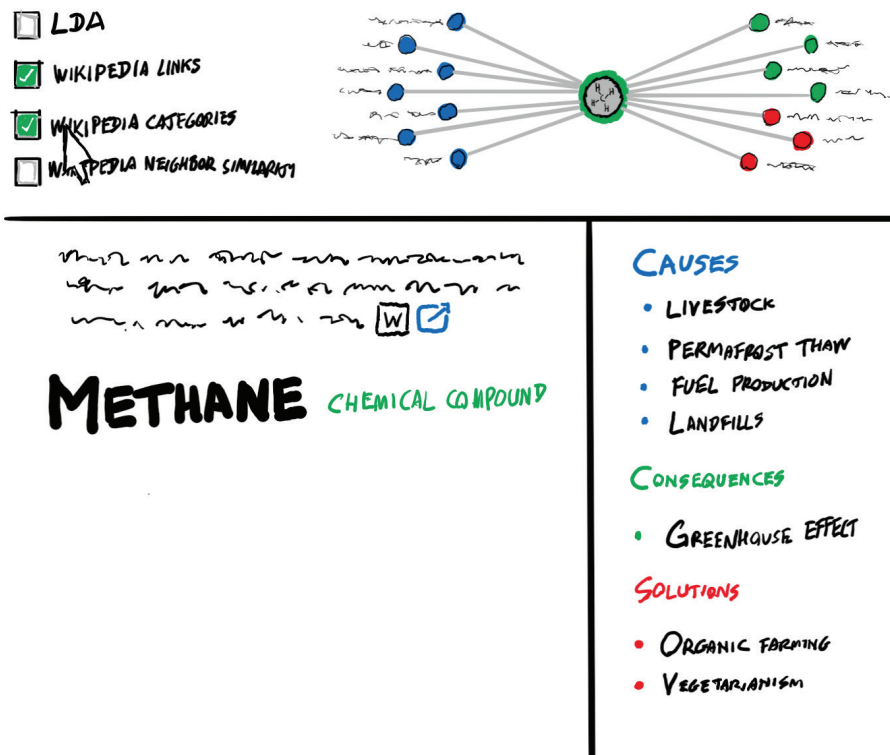


Figure 2

The graph will be squeezed in the remaining upper 25% and take the following determined shape :

- Middle : the current node
- Left : the nodes representing the immediate causes of the issue
- Right : the nodes representing the immediate consequences and solutions of the issue
- Background, position uncontrolled : the rest of the nodes

The user can navigate to causes, consequences and solutions by clicking them in the information page or directly on the graph, which will rearrange itself to display the chosen node in the same fashion. It is possible using checkboxes to toggle the different linkage methods listed higher (LDA and others).

Note that the specifics of this layout are prone to changes as we understand more and more about our dataset. We do not know at which extent we will be able to extract accurate and relevant relationship data with our current methods, therefore we might alter the way we categorize in order for our visualization to be interesting.

As mentioned above, the goal of the project is not only to allow easy navigation through different subjects related to climate change but also to allow easy understanding of those subjects. In order to implement that, we opted for a certain layout for key information corresponding to a particular issue. It would incorporate among others yet to be determined:

- A description of the subject
- A title
- Causes, consequences, solutions

The causal relations are extracted semi-automatically from Wikipedia articles. We apply pattern matching using a dictionary of causal terms and verbs, and then we manually select the most relevant results.