

Activités du Lundi 22 avril au Vendredi 26 avril :

- Préparation à l'extraction du target-path des events :
 - o Transformation des fichiers txt sans annotation en structure de dépendances.

Résultat de la transformation en dépendances :

```
defaultdict(<function DependencyGraph.__init__.<locals>.<lambda> at 0x000000B668585620>,  
  {0: {'address': 0,  
      'ctag': 'TOP',  
      'deps': defaultdict(<class 'list'>, {'root': [1]}),  
      'feats': None,  
      'head': None,  
      'lemma': None,  
      'rel': None,  
      'tag': 'TOP',  
      'word': None},  
  1: {'address': 1,  
      'ctag': 'CD',  
      'deps': defaultdict(<class 'list'>, {}),  
      'feats': '-',  
      'head': 0,  
      'lemma': '-',  
      'rel': 'root',  
      'tag': 'CD',  
      'word': '1998-08-07'}})defaultdict(<function DependencyGraph.__init__.<locals>.<lambda> at 0x000000B668585EA0>,  
  {0: {'address': 0,  
      'ctag': 'TOP',  
      'deps': defaultdict(<class 'list'>, {'root': [2]}),  
      'feats': None,  
      'head': None,  
      'lemma': None,  
      'rel': None,  
      'tag': 'TOP',  
      'word': None},  
  1: {'address': 1,  
      'ctag': 'NNS',  
      'deps': defaultdict(<class 'list'>, {}),  
      'feats': '-',  
      'head': 2,  
      'lemma': '-',  
      'rel': 'nsubj',  
      'tag': 'NNS',  
      'word': 'Explosions'},  
  2: {'address': 2,  
      'ctag': 'VB',  
      'deps': defaultdict(<class 'list'>,  
                          {'dobj': [4],  
                           'nmod': [8],  
                           'nsubj': [1]}),
```

On remarque que pour chaque dictionnaire, la clé est l'identifiant de mot. A chaque phrase, cet identifiant se remet à 0.

- Création d'un fichier csv pour retrouver les events dans les fichiers de dépendances qui utilisent des données non-annotées.
 - o Même base que pour le csv des identifiants artificiels
 - mais sans volonté de rassembler les timex par leur « > » ou autre ponctuation comme « ' » pour « year's »
 - Le but est de ne pas fausser les identifiants de mots
 - o On laisse la tokenization telle quelle car elle correspond bien à celle du stanford parser dependency
- Recherche sur comment exploiter les dépendances qui ont un format type dictionnaire de dictionnaires (mettre au format json pour pouvoir utiliser les dictionnaires ?)