

Dossier JSON

Dans le dossier JSON on y retrouve l'ensemble des fichiers Json générés pour tous les articles allant de 2004 à 2018.

Ces fichiers concernent uniquement comme demandés les articles ayant une méta-session comprise entre 1 et 9.

Dossier Langage

Dans ce dossier, il est question d'une api qui permet de détecter la langue d'un texte.

Cette api est utile pour éventuellement travailler dans l'avenir sur des textes d'autres langues.

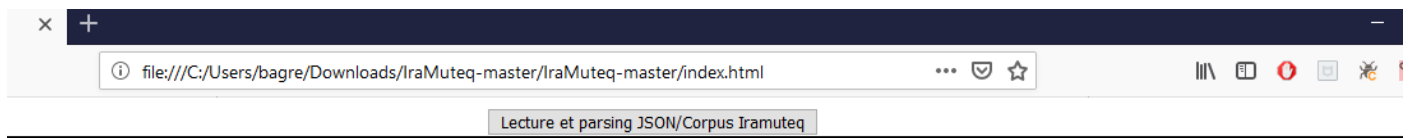
Le fichier index.html

Ce fichier contient l'essentiel du programme pour la génération d'un corpus IraMuteQ selon le modèle désiré.

Dans le but d'éviter des erreurs de restrictions concernant le navigateur chrome au cours de la génération il est vivement conseillé de l'ouvrir avec le navigateur Firefox.

Comment utiliser le fichier index.html

Après donc avoir ouvert le fichier index.html dans le navigateur Firefox , vous obtiendrez l'image ci-dessous.



Après avoir cliqué sur le bouton affiché , la génération du corpus peut prendre un certains temps (en moyenne trente secondes maximum).

On obtient donc une génération de ce type .

```
**** *annee_2014 *VilleConf_Rennes
Dans cet article, nous présentons une approche de fouille de textes ainsi qu'une interface de visualisation afin d'explorer une large collection de chansons françaises à partir des paroles. Dans un premier temps, nous collectons paroles et métadonnées de différentes sources sur le Web. Nous utilisons une approche combinant clustering et analyse sémantique latente afin d'identifier différentes thématiques et de déterminer différents descripteurs significatifs. Nous transformons par la suite le modèle afin d'obtenir une visualisation interactive permettant d'explorer la collection de chansons
**** *annee_2014 *VilleConf_Rennes
Les graphes orientés attribués sont des graphes orientés dans lesquels les noeuds sont associés à un ensemble d'attributs. De nombreuses données, issues du monde réel, peuvent être représentées par ce type de structure, mais encore peu d'algorithmes sont capables de les traiter directement. La fouille des graphes attribués est difficile, car elle nécessite de combiner l'exploration de la structure du graphe avec l'identification d'itemsets fréquents. De plus, du fait de l'explosion combinatoire des itemsets, les isomorphismes de sous-graphes, dont la présence impacte énormément les performances des algorithmes de fouille, sont beaucoup plus nombreux que dans les graphes étiquetés. Dans cet article, nous présentons une nouvelle méthode de fouille de données qui permet d'extraire des motifs fréquents à partir d'un ou de plusieurs graphes orientés attribués. Nous montrons comment réduire l'explosion combinatoire provoquée par les isomorphismes de sous-graphes en traitant de manière particulière les motifs automorphes.
**** *annee_2014 *VilleConf_Rennes
Les représentations condensées ont fait l'objet de nombreux travaux depuis 15 ans. Tandis que les motifs maximaux des classes d'équivalence ont reçu beaucoup d'attention, les motifs minimaux sont restés dans l'ombre notamment cause de la difficulté de leur extraction. Dans ce papier, nous présentons un cadre générique concernant l'extraction de motifs minimaux en introduisant la notion de système minimisable d'ensembles. Il permet de considérer des langages variés comme les motifs ensemblistes ou les chaînes de caractères, mais aussi différentes métriques dont la fréquence. Ensuite, pour n'importe quel système minimisable d'ensembles, nous introduisons un test de minimalité rapide permettant d'extraire en profondeur les motifs minimaux. Nous démontrons que l'algorithme proposé est polynomial-delay et polynomial-space. Des expérimentations sur les benchmarks traditionnels complètent notre étude.
```

Ensuite il suffit juste de sélectionner tous l'ensemble des textes de la page (CTRL+A) de le copier (CTRL+C) et de le coller (CTRL+V) dans un fichier .txt créé manuellement.

Comment modifier le fichier index.html

```
$(function() {
    $('#lecture').on('click', function(){
        // Il y a 1330 fichier Json
        // On parcourt l'ensemble des fichiers JSON : Ici de 1 à 1330
        for (var i=1; i<= 1330; i++) {
            $.when(
                // Dans le Répertoire JSON se trouve les articles
                $.getJSON("JSON/article_"+i+".json")
            ).done(function(data) {

                guessLanguage.info(data.abstract, function(info) { // vérifie la langue du texte
                    // data.year permet de connaître l'année du fichier JSON en cours d'exécution
                    // info[2]=="French" vérifie si la langue du résumé est en français
                    if ((data.year == "2018") && (data.abstract != null) && (data.metaSession >= 1 && data.metaSession <= 9) && (info[2]=="French"))
                    {
                        // on écrit dans la page html ce qu'on souhaite
                        // '<br>' permet d'aller à la ligne suivante
                        $('#zone').append('**** *annee_'+data.year+" "+"+"*VilleConf_"+data.place+ '<br>' + data.abstract + '<br>');
                        console.log(info[2]);
                        // la commande ci-dessous donnera les lignes suivantes :
                        /* **** *annee_2018 *VilleConf_Luxembourg
                           texte texte texte texte texte
                           texte texte texte texte texte
                        */
                    }
                    if ((data.year == "2017") && (data.abstract != null) && (data.metaSession >= 1 && data.metaSession <= 9) && (info[2]=="French"))
                    {
                        $('#zone').append('**** *annee_'+data.year+" "+"+"*VilleConf_"+data.place+ '<br>' + data.abstract + '<br>');
                        console.log(info[2]);
                    }
                })
            })
        }
    });
});
```