

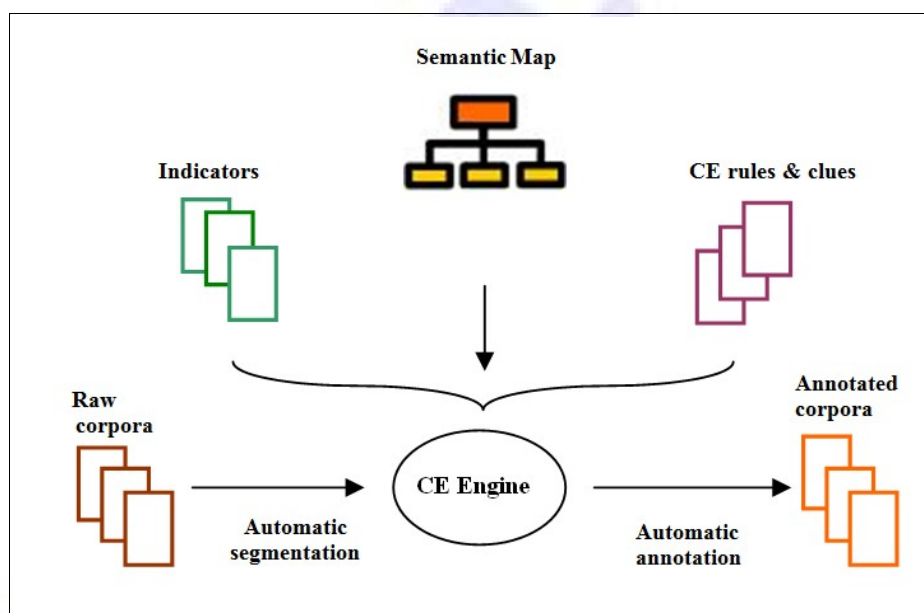
# EXCOM-2

## User's manual

*Draft*

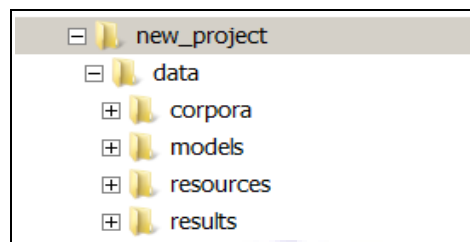
*Written by Motasem Alrahabi*

*Last update :  
March, 2015*

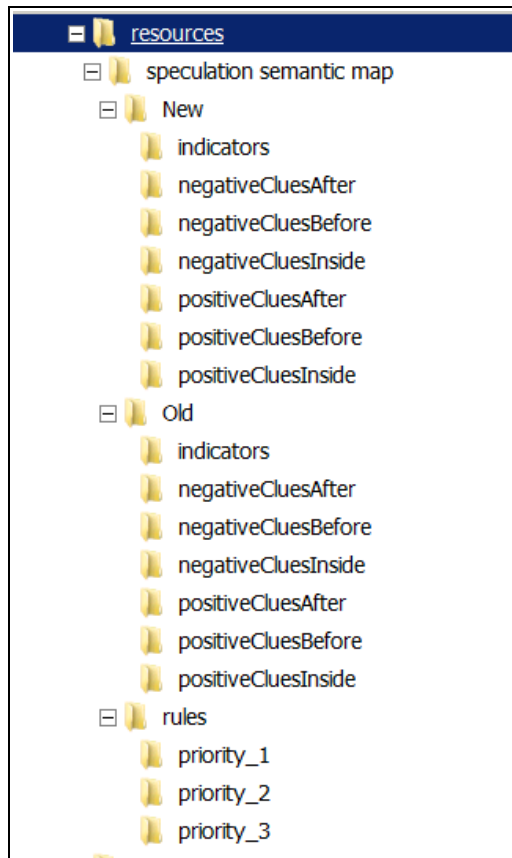


## Préparation des données :

1. Téléchargez sur votre ordinateur le dossier « new-project » à partir du [site](#).
2. Décompressez le dossier « new-project », et renommez-le comme vous voulez. Vous trouverez dedans les éléments suivants :
  - le dossier « data » qui contient les dossiers suivants :
    - « corpora » qui contient des textes à traiter.
    - « models » qui contient des fichiers nécessaires au traitements. Ce dossier et son contenu ne doivent pas être modifiés, supprimés ou déplacés.
    - « resources » qui contient vos ressources linguistiques (marqueurs et règles).
    - « results » qui contient les résultats des traitements, étape par étape.



3. Organisez vos ressources linguistiques dans le dossier modèle « resources » :
  - le dossier « resources » peut contenir une ou plusieurs cartes sémantiques (ou simplement « tâche » à traiter). Celles-ci sont organisées sous forme arborescente. A chaque carte sémantique ou tâche correspond la structure suivante :
    - Un dossier racine qui porte le « nom de la carte » ou de la tâche. Exemple : « *speculation semantic map* ». Une carte peut contenir à son tour une ou plusieurs catégories (feuilles dans l'arborescence) :
      - Un dossier qui porte le « nom de la catégorie », exemple : « *New* » ou « *Old* ». Chaque dossier d'une catégorie contient les 7 dossiers de marqueurs suivants :
        - Le dossier des indicateurs « *indicators* »
        - Le dossier des indices positifs avant l'indicateur « *positiveCluesBefore* »
        - Le dossier des indices positifs après l'indicateur « *positiveCluesAfter* »
        - Le dossier des indices négatifs avant l'indicateur « *negativeCluesBefore* »
        - Le dossier des indices négatifs après l'indicateur « *negativeCluesAfter* »
        - Le dossier des indices positifs à l'intérieur de l'indicateur « *positiveCluesInside* »
        - Le dossier des indices négatifs à l'intérieur de l'indicateur « *negativeCluesInside* »Ces sept dossiers ne doivent pas être supprimés, leurs noms ne doivent pas être modifiés non plus.
    - Le dossier « *rules* » qui contient toutes les règles construites pour les catégories de la carte. Pour en savoir plus, consultez la partie « [Priorité entre règles](#) ». Le dossier « *rules* » ne doit pas être supprimé ou renommé. Il faut également au moins un dossier de priorité : « *priority\_1* » pour contenir les règles. Celui-ci, pareil, ne doit pas être supprimé ou renommé.



Exemple d'une carte sémantique

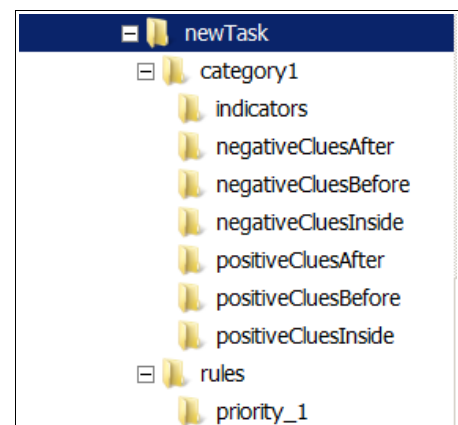
### Remarques importantes :

1. À chaque fois que vous lancez l'application, il faut choisir comme racine de votre projet l'emplacement du dossier « new-project » que vous venez de décompresser sur votre PC.
2. Pour que le programme excom2 fonctionne correctement, il ne faut absolument pas modifier l'organisation hiérarchique de ces dossiers (*sauf bien sûr les noms des cartes et des catégories*).

### Comment organiser une carte sémantique ou une tâche ?

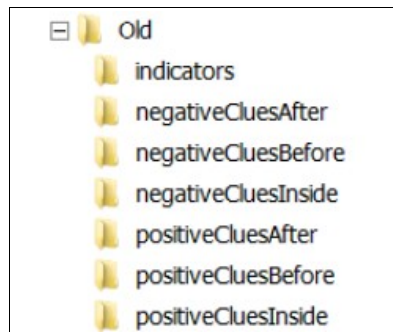
Les dossiers des catégories (ex. *New et Old*) dans excom2 peuvent être organisés de manière à reproduire l'organisation de la carte sémantique. Ils peuvent être placés ou bien imbriqués les uns à côté des autres, selon le cas.

Le moteur excom2 ne donne aucune *interprétation* sémantique à l'organisation des catégories d'une carte (suite de catégories ou arborescence). Pour l'outil, c'est une simple organisation de dossiers.



## Comment construire les fichiers de marqueurs linguistiques ?

Vous devez déposer vos marqueurs dans les dossiers des marqueurs d'une catégorie donnée.



Les fichiers doivent être au format texte brute (codage UTF-8).

Vous pouvez bien entendu utiliser des marqueurs sous forme d'expressions régulières. Il existe des outils gratuits qui peuvent vous assister dans cette démarche (voir par exemple [Regex coach](#)).

## Comment créer une règle d'EC ?

Les règles doivent être au format XML. Il existe 3 types de règles dans excom2 :

### Règles de type 1

Les règles de type 1 opèrent dans un espace de recherche continu, c'est-à-dire dans une phrase ou un titre.

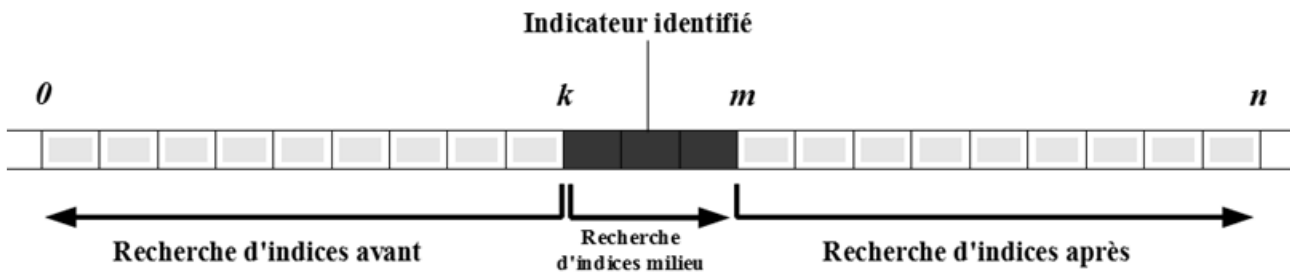
Une règle de type 1 sous excom2 peut identifier une combinaison de marqueurs linguistiques. La place de l'indicateur déclencheur permet de définir plusieurs sous-espace de recherche :

1. Un espace avant l'indicateur où le moteur pourrait identifier :
  - un indice positif
  - un indice négatif
2. Un espace après l'indicateur où le moteur pourrait identifier :
  - un indice positif
  - un indice négatif
3. Un espace à l'intérieur même de l'indicateur où le moteur pourrait également identifier :
  - un indice positif
  - un indice négatif

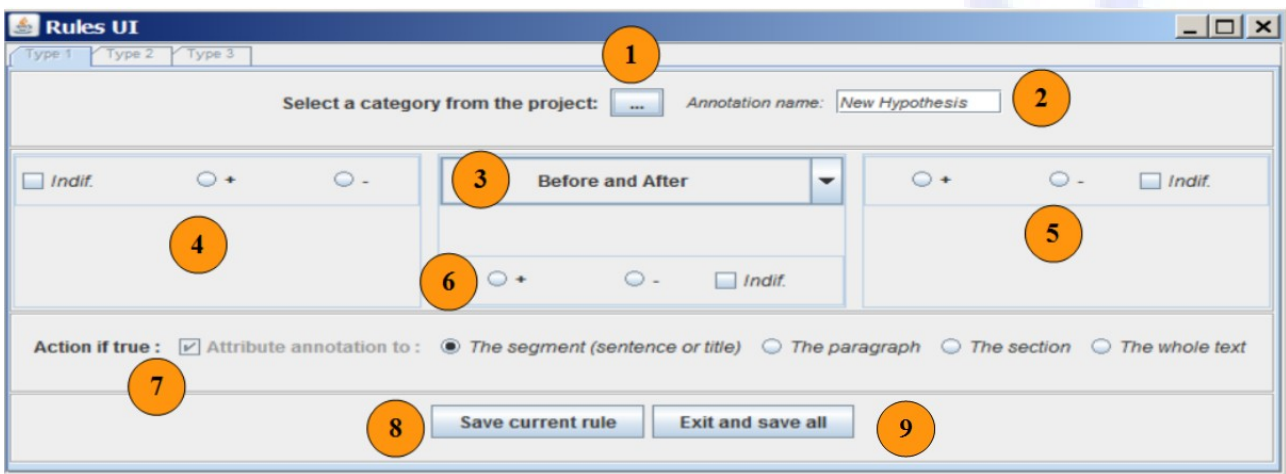
Un indice positif est un indice complémentaire qui vient confirmer la valeur sémantique de l'indicateur. Par contre, la présence d'un indice négatif dans le contexte d'un indicateur annule la règle en cours.

Le moteur excom2 recherche toujours l'indice positif en premier, ensuite, il recherche l'indice négatif. Si celui-ci est trouvé, la règle est annulée.

La figure suivante montre le sens de parcours d'un segment à la recherche de marqueurs.



Excom2 offre une interface pour créer automatiquement des règles au format XML (bouton « Rules »).



Voici les étapes à suivre :

- (1) Choisir la catégorie à partir de la racine du projet.
- (2) Modifier le nom de l'annotation généré par défaut (option)
- (3) Choisir le connecteur entre l'espace Avant et l'espace Après : Et, Ou, Ou exclusif.
- (4) Définir ce que le moteur doit trouver dans l'espace Avant : indice positif et indice négatif.
  - Définir également l'ordre entre ces deux indices. Exemples :
    - Si vous cochez uniquement « + » cela signifie que le moteur doit rechercher un indice positif avant l'indicateur.
    - Si vous cochez « + » et « - » et que vous déplacez (glissez) le bouton « - » avant le bouton « + », cela signifie que le moteur va rechercher dans l'espace en question un indice positif et ensuite, dans l'espace qui précède la position de l'indice positif trouvé, il va chercher un indice négatif.
    - Si vous cochez « indif », cela signifie que le moteur doit rechercher les deux indices positif et négatif dans un ordre indifférent.
    - etc.
- (5) Même chose que (4)
- (6) Même chose que (4)

- (7) L'action par défaut dans cette version est « l'annotation » de l'espace de recherche. Vous pouvez cependant attribuer l'annotation à l'une de ces 4 espaces :
- le segment en cours : phrase ou titre
  - le paragraphe en cours
  - la section en cours
  - le texte entier
- (8) Après la création de chaque règle, il faut enregistrer en appuyant sur le bouton « save current rule ». Puis vous commencez à créer la règle suivante.
- (9) Si vous avez fini de créer toutes vos règles, vous pouvez appuyer sur « exit and save all ». Une fenêtre apparaît alors pour indiquer le chemin d'emplacement où vous voulez stocker le fichier créé (consultez aussi la partie « [Priorité entre règles](#) »).

## Règles de type 2

*Ce type de règles récursives n'est pas disponible avec cette version du programme.*

## Règles de type 3

*Ce type de règles hiérarchiques n'est pas disponible avec cette version du programme.*

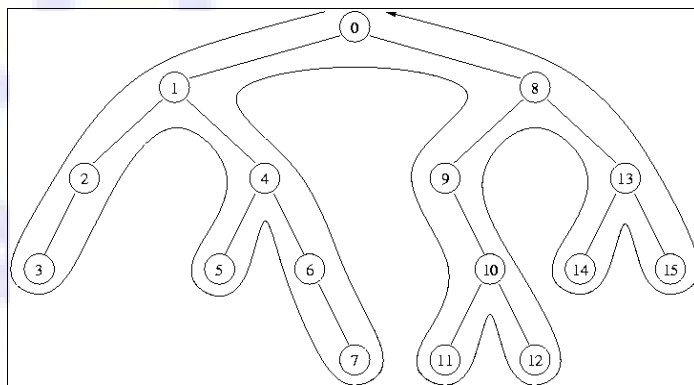
### Remarque :

Après avoir construit les règles, si vous modifiez le fichier des règles, le nom d'un dossier dans les ressources ou bien l'organisation des ressources, il faudrait recréer (ou modifier) les règles concernées par ce changement.

## Comment fonctionnent les priorités entre les règles ?

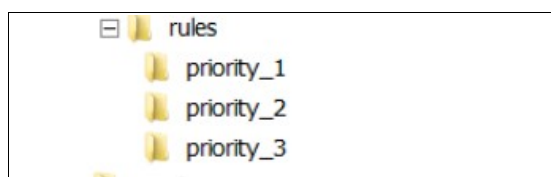
Excom2 permet de gérer deux types de priorités entre les règles :

1. Priorité en lecture descendante, définie par l'arborescence de la carte sémantique. Dans ce cas, le moteur excom2 parcourt l'arbre d'une carte selon un mode en « profondeur préfixe ». Le parcours préfixe de l'arbre ci-dessous, par exemple, parcourt les nœuds dans l'ordre [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15].



Ce type de priorité n'est pas disponible dans cette version du programme.

2. Priorité définie par l'utilisateur. Celui-ci doit organiser ses règles dans le dossier « Rules », selon un ordre de priorités de 1 à n (du plus important vers le moins important) : *priority\_1*, *priority\_2*, *priority\_3*, etc.



Dans ce mode de priorité, le passage d'un niveau « *priority\_n* » à un mode de niveau « *priority\_{n+1}* » dépend du choix de l'utilisateur :

- Lecture « Minimale » → pour chaque segment, le moteur va parcourir les dossiers des priorités les uns après les autres. Dès la première annotation attribuée au segment, le moteur arrête le processus et passe au segment suivant.  
→ *Donc, dans un segment donné, il y aura une seule annotation maximum, provenant de la première catégorie trouvée.*
- Lecture « Multiple » → pour chaque segment, le moteur va parcourir les dossiers des priorités les uns après les autres. Quand il identifie une catégorie dans le segment, le moteur l'annote et passe à la catégorie suivante du même niveau de priorité. Ensuite il passe aux niveaux supérieurs, avant de passer au segment suivant.  
→ *Donc, dans un segment donné, il y aura : une seule annotation maximum pour chaque catégorie trouvée.*

*Ce type de lecture n'est pas disponible dans cette version du programme.*

Par défaut, si l'utilisateur ne souhaite activer aucune forme de priorité, il peut se contenter de choisir « No, by default » dans la fenêtre de dialogue qui s'affiche avant l'annotation. Le moteur parcourera alors tous les sous dossiers de « Rules », l'un après l'autre, sans appliquer de priorités.

→ *Dans ce cas-là, pour un segment donné, il y aura : toutes les annotations potentielles pour toutes les catégories trouvées.*

## Quelle est la structure de texte traité dans excom2 ?

Voici, de manière simplifiée, les éléments de la structure d'un document traité sous excom2 :

- Article
  - ArticleInfo
    - Date, Authors, SourceName, SourceLink, DocumentId, DocumentType, KeyWords, Abstract
  - Body
    - MainTitle
    - Section(s)
      - Title
      - Paragraph(s)
        - Title
        - Sentence(s)

Vous pouvez retrouver la DTD adoptée dans excom2 à l'emplacement de votre « new-project » que vous avez téléchargé : `\data\models\dtd\dtdArticle.dtd`

## Le schéma d'annotation dans EXCOM-2 :

...

## Lancer EXCOM-2 :

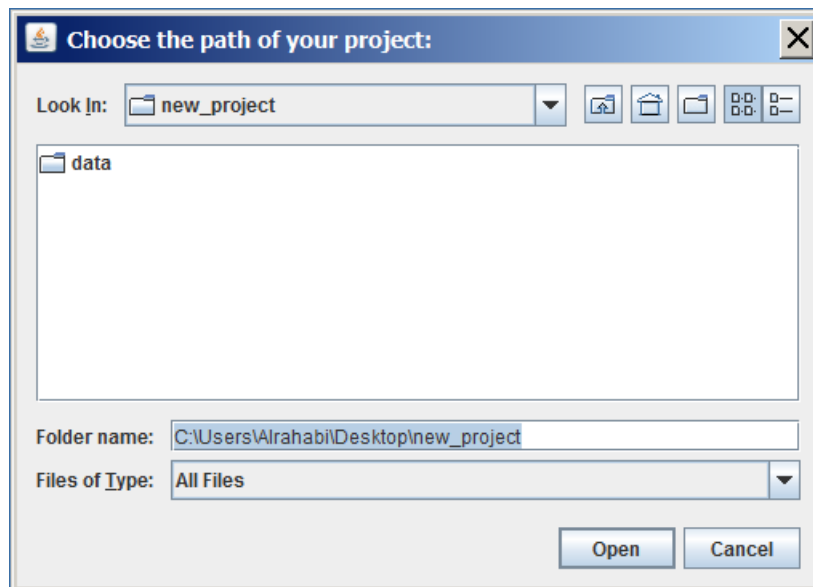
### 1 - Pré-requis généraux :

1. Ce logiciel a été testé sous Windows et Mac.
2. Vérifiez que Java soit bien installé sur votre ordinateur<sup>1</sup>.
3. Sous Windows, utilisez de préférence « Internet Explorer » comme navigateur par défaut pour avoir un meilleur affichage des résultats.
  - Pour activer de manière permanente l'affichage des graphes générés dans IE :
    - *Select « Tools » - « Internet options ». Then go to the « Advanced » tab and scroll down to the « Security » section and check the little checkbox labeled « Allow active content to run in files on My Computer. ».*
4. Lors du premier lancement du programme à partir du [site](#), ouvrez-le avec l'application `javaws.exe` qui se trouve dans le répertoire `bin` de java (si cela ne se fait automatiquement).

---

<sup>1</sup> <https://www.java.com/fr/download/>





5. Dans la fenêtre qui s'affiche, pointez vers le dossier « new-project » que vous venez de décompresser et que vous avez éventuellement renommé. Le logo de l'application doit s'afficher correctement.

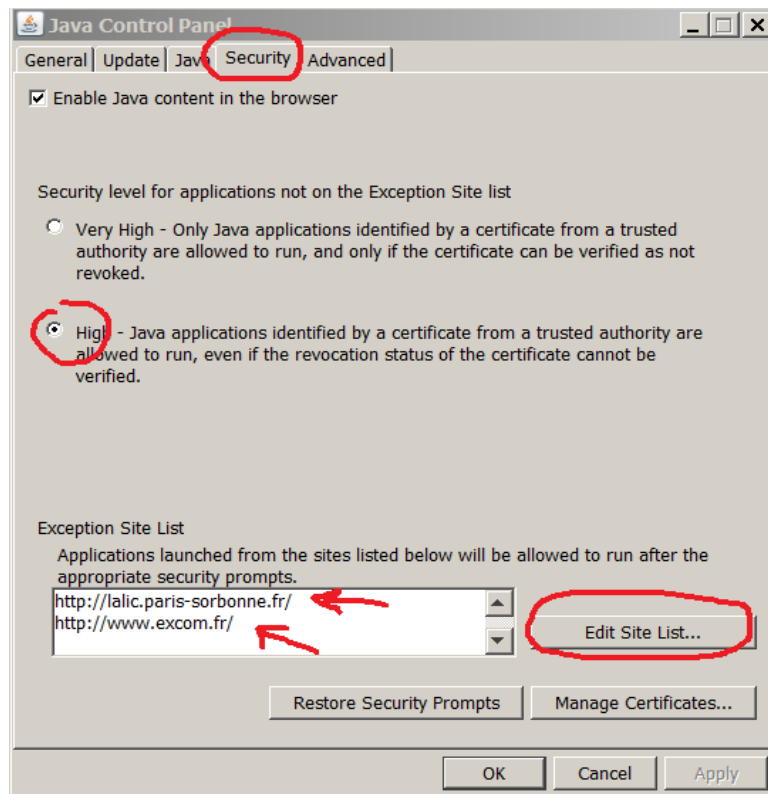


6. Au lancement de l'application, si vous avez un message d'erreur comme :

*Les applications Java sont bloquées par vos paramètres de sécurité*

il faut alors utiliser la fonctionnalité « Liste des sites avec exception » en suivant les étapes suivantes :

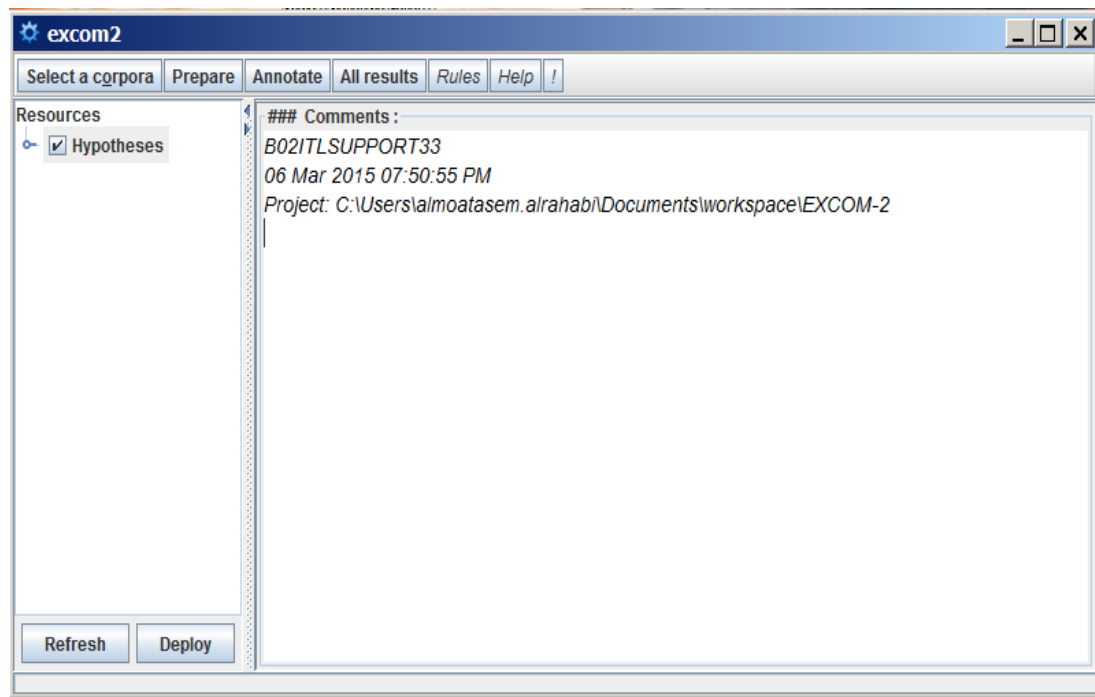
- Accédez au panneau de configuration Java (*sous Windows, cliquez sur Démarrer, puis sur Configurer Java*)
- Cliquez sur l'onglet **Sécurité**.
- Cliquez sur le bouton **Modifier** la liste des sites
- Cliquez sur **Ajouter** dans la fenêtre Liste des sites avec exception



- Cliquez dans le champ vide sous le champ Emplacement pour entrer les 2 URL suivants :
  - <http://www.excom.fr/>
  - <http://lalic.paris-sorbonne.fr/>

## 2 - Utiliser l'application:

1. Après avoir effectué toutes les étapes pré-requises, lancez maintenant la chaîne de traitement :
  - Choisir un texte ou un dossier de textes
  - Préparer (segmenter)
  - Annoter
  - Afficher l'ensemble des segments annotés, et le graphe de visualisation (option).



2. Le volet gauche de la fenêtre permet de visualiser l'organisation des ressources linguistiques et de choisir la ou les catégories à traiter avant chaque annotation.
3. Les résultats des différentes étapes du traitement peuvent être visualisés dans les sous-dossiers du répertoire « results ».

### **Besoin d'aide ?**

Pour toute question technique ou suggestion, n'hésitez pas à m'écrire à l'adresse : [motasem.alrahabi@gmail.com](mailto:motasem.alrahabi@gmail.com)

Consultez également les sites :

<http://www.excom.fr/>

<http://lalic.paris-sorbonne.fr/PAGESPERSO/alrahabi/>

### **Référence:**

Alrahabi M., 2010. *EXCOM-2 : plateforme d'annotation automatique de catégories sémantiques. Applications à la catégorisation des citations en français et en arabe*. Thèse de doctorat, sous la direction du Prof. Jean-Pierre Desclés, Université Paris-Sorbonne.