Poestry Miner

- Manuel utilisateur -

Réalisé par RINGWALD Célian
Dans le cadre d'un stage de fin d'année
de Master 1 d'informatique

PRESENTATION

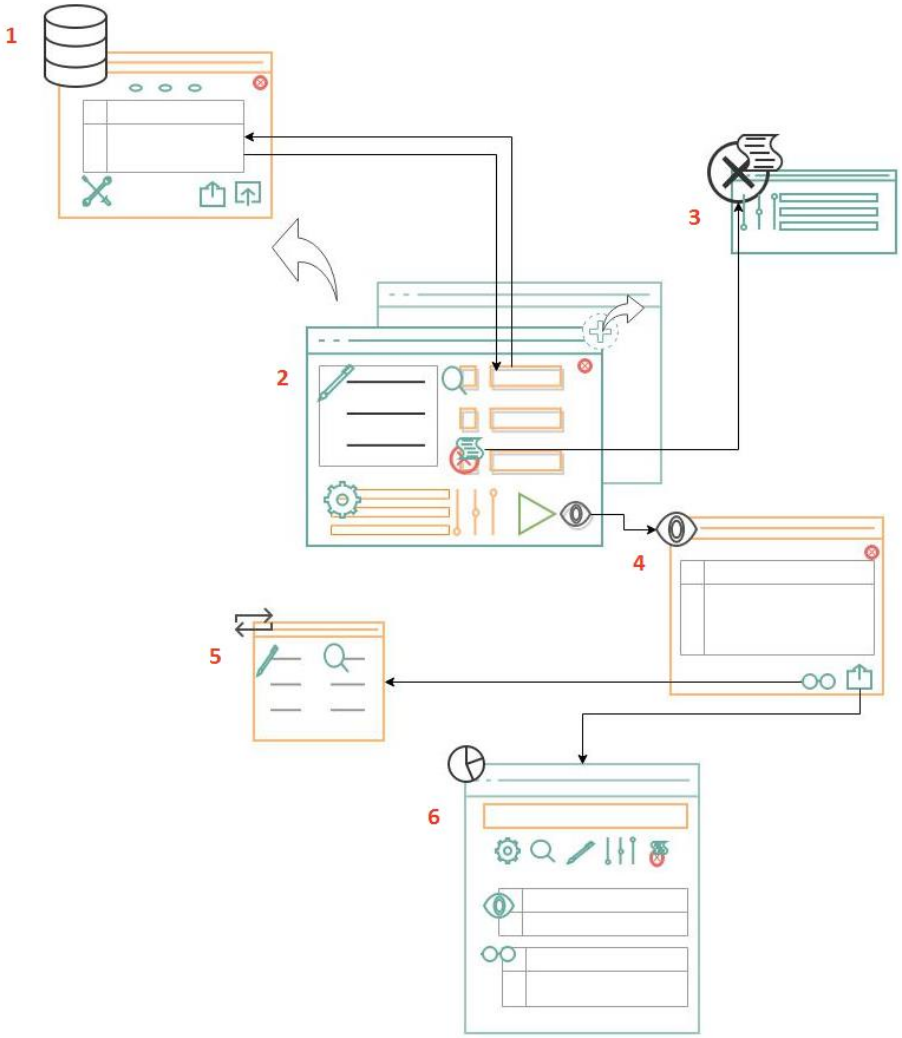
Poestry Miner est un logiciel réalisé en JAVA utilisant une base de données Derby. Ce programme est le résultat d'un stage de 3 mois au CIPL (Centre Informatique de Philosophie et Lettre) de l'Université de Liège, réalisé par Célian RINGWALD dans le cadre de son stage de fin d'année de Master 1 d'informatique à l'Université de Lyon 2. Cet outil permet d'assister M. Gerald PURNELLE, philologue et passionné de poésie, dans son travail de publication des œuvres inédites d'Izoard et Jacqmin.

Adapté à une utilisation tournée vers l'étude de la poésie, il permet à son utilisateur de gérer un corpus de poèmes dans le but de l'utiliser comme base de recherche et d'extraction d'informations. En effet, en fournissant une requête pouvant être basée sur quelques mots significatifs, un texte, un poème ou un recueil ; Poestry Miner fournit une liste de poèmes pour lesquels de nombreux indicateurs sont fournis, permettant de repérer quels poèmes sont les plus pertinents face aux besoins de l'utilisateur : recherche de mots clefs, analyse lexicale…

Ce guide permet à l'utilisateur lors de sa première utilisation de prendre rapidement en main les différentes fonctionnalités du logiciel.

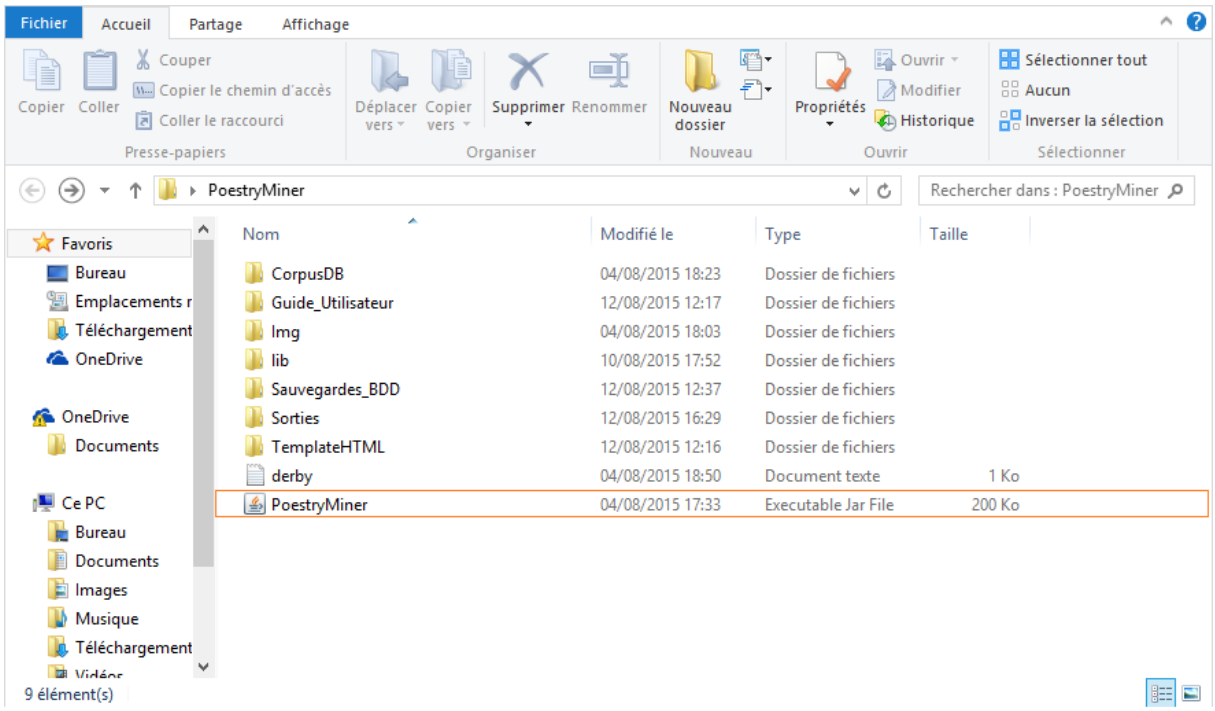
SOMMAIRE

1. VUE GLOBALE DU LOGICIEL



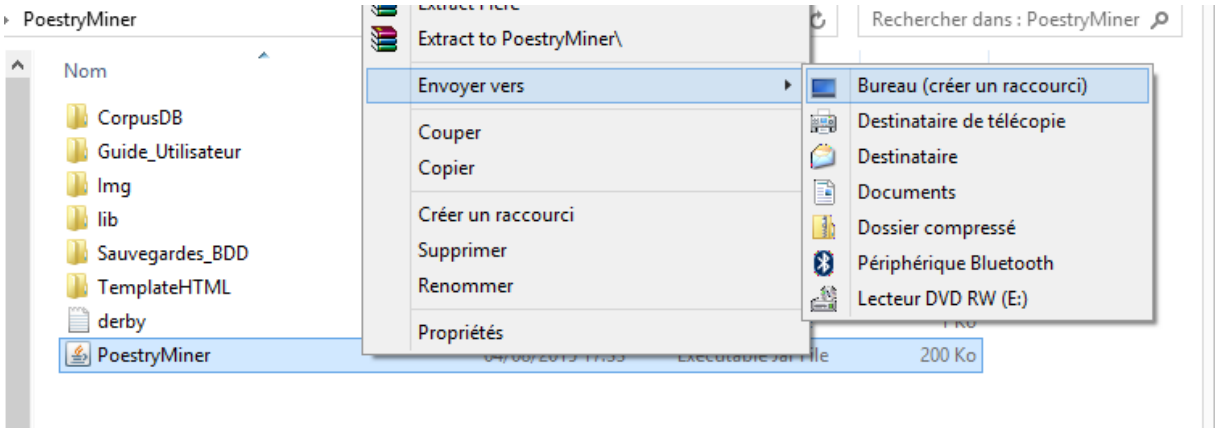
L'utilisation de Poestry Miner est basée sur une base de données que l'utilisateur doit dans un premier temps alimenter afin de pouvoir lui soumettre des requêtes. Son interface (1) et son utilisation y sont décrites dans la partie … de ce guide. Une fois la base de corpus alimentée, l'utilisateur a la possibilité de lui soumettre une requête paramétrée selon son besoin (interface 2), ainsi que d'affecter à celle-ci un anti-dictionnaire (interface 3). Une fois la requête lancée l'utilisateur obtient dans l'interface 4 tous les résultats obtenus, affichant dans une vue tabulaire différents indicateurs calculés pour les poèmes sélectionnés préalablement. Il a ensuite le choix d'exporter les résultats qui lui semblent intéressants (interface 6) ou tout simplement de comparer le contenu de la requête avec les poèmes de la liste (interface 5). Ce guide décrit dans cet ordre les différentes règles d'utilisation de chacune de ces interfaces, ainsi que les règles de calcul des indicateurs mis en place.

2. LANCEMENT DU LOGICIEL



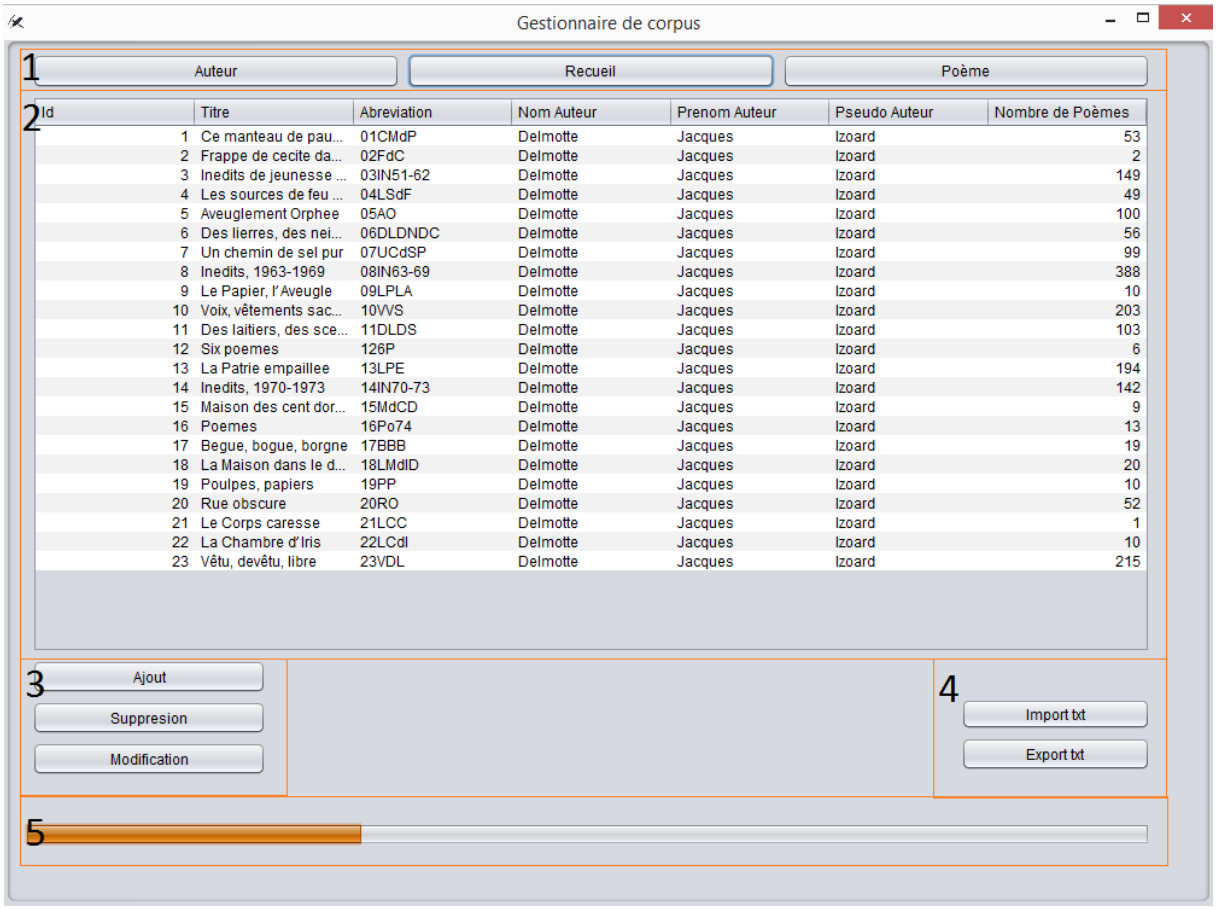
Poestry Miner est fourni au format .rar dans une archive, extrayez donc celle-ci afin de pouvoir lancer par la suite le programme. Afin de lancer Poestry Miner il vous suffit de cliquer sur le fichier .jar mis en valeur ci-dessus.

Si vous souhaitez y accéder depuis votre Bureau, le plus simple et de créer un raccourci pointant vers ce fichier :



Il est conseillé de ne pas toucher aux dossiers « Corpus-DB » contenant la base de données (voir ci-après la partie 3.1. précautions d'utilisation), « Img » contenant les icones du programme, « lib » contenant les librairies java utilisées par le programme ainsi que « Template HTML » contenant le Template d'export de résultats.

3. GESTION DE LA BASE DE DONNEE

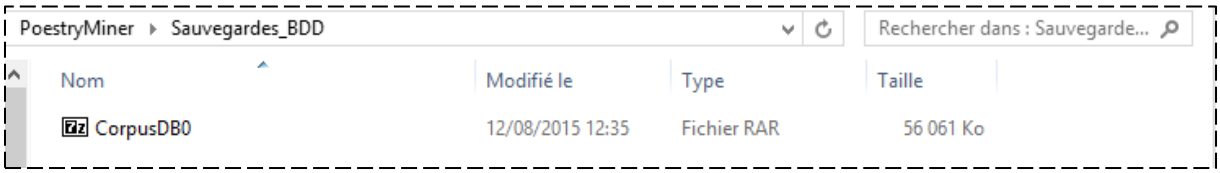


Le gestionnaire de base de corpus permet à l'utilisateur de manipuler de manière triviale les données qu'il souhaite étudier dans le cadre de son travail. Il permet donc gérer : auteurs, recueils et poèmes en explorant la barre de navigation (1). Une vue tabulaire (2) lui offre un aperçu de ce qui est déjà entré dans la base de données. Deux modes de gestion lui sont alors offerts : un mode manuel, via lequel il sera amené à entrer champ par champ les données d'une entité donnée (3) ainsi qu'un mode de gestion par fichier texte (4) lui permettant d'importer directement ses données depuis des fichiers formatés. La barre de progression (5) lui permet de suivre l'avancement de la tâche en cours (parfois longue si celle-ci est par exemple l'import de toute l'œuvre d'un auteur).

3.1. PRECAUTIONS D'UTILISATION

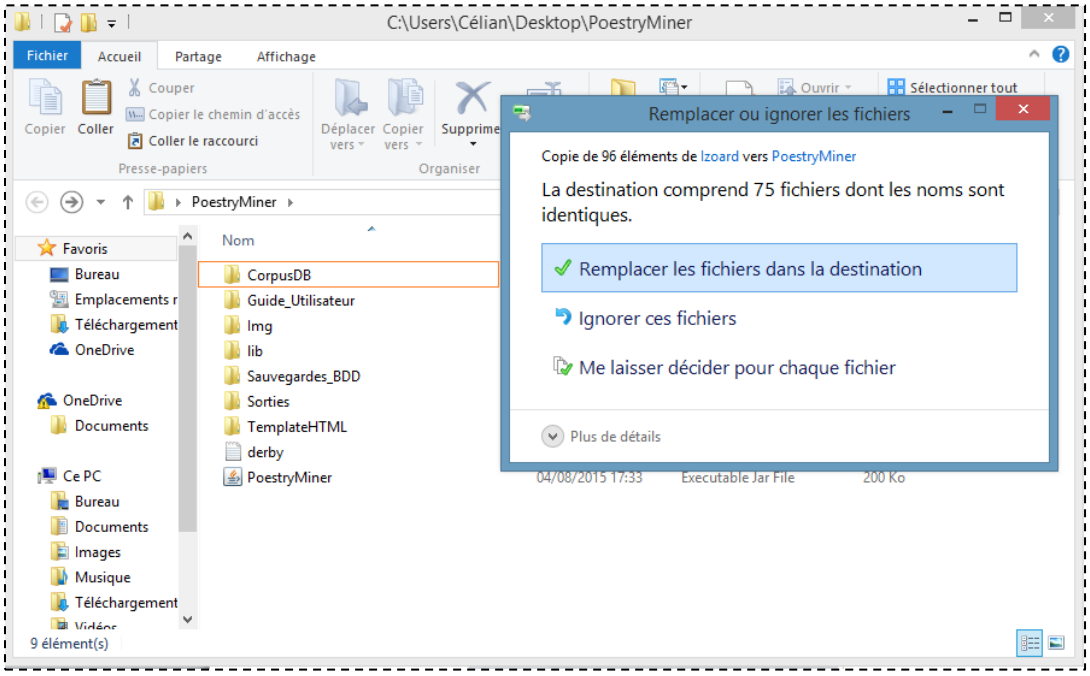
- La gestion de la base de données est une tâche délicate, en effet toute utilisation maladroite de celle-ci (interruption d'un import, d'une suppression…) peut entrainer par la suite des problèmes lors de l'utilisation future de l'outil.

Il est donc conseillé de sauvegarder régulièrement celle-ci. Le dossier *« Sauvegardes_BDD »* du dossier original Poestry Miner est d'ailleurs prévu à cet effet, il contient une copie archivée de la base de données originale (recelant le corpus des œuvres d'Izoard).



Pour Sauvegarder la base de données copiez le dossier *« corpus DB »* dans le dossier *« Sauvegardes_BD »,* archivez le et renommez le en fonction de son numéro de sauvegarde (exemple : CorpusDB5, si 5e sauvegarde).

En cas de problèmes extrayez la base de données de votre dernière archive en date et remplacez le dossier nommé *« CorpusDB »* par votre fichier de sauvegarde :

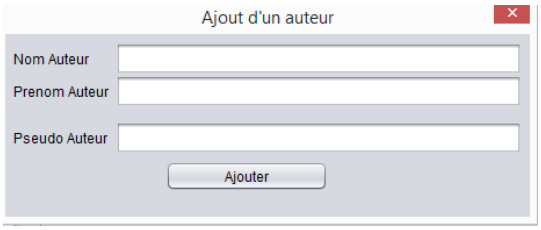


3.2. GESTION MANUELLE

3.2.1. Ajout

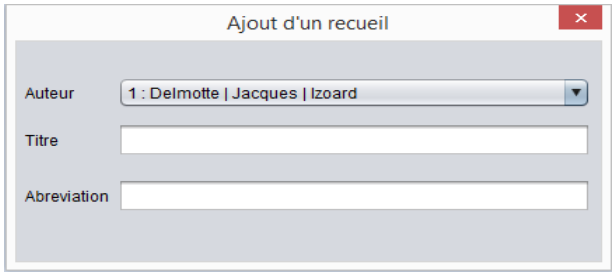
- Ajout d'un Auteur :

  Lors de l'ajout d'un auteur l'utilisateur a l'obligation de préciser soit le nom et prénom de celui-ci, soit son pseudonyme, il sera possible par la suite d'y ajouter des recueils.



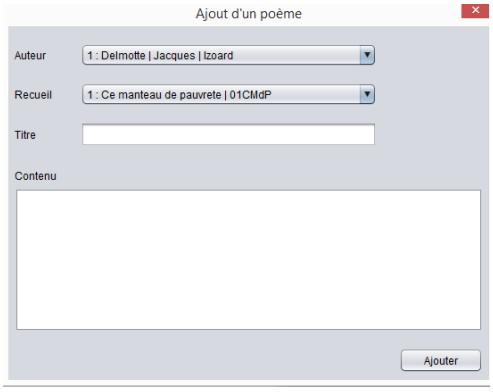
- Ajout d'un recueil :

  Lors de l'ajout d'un recueil l'utilisateur à l'obligation de préciser le titre de celui-ci, il doit notamment sélectionner l'auteur de celui-ci. L'abréviation n'est pas obligatoire mais permettra lors de l'export en format texte de proposer automatiquement ce champ comme nom de fichier. Une fois ceci effectué l'utilisateur pourra lui attribuer des poèmes.



- Ajout d'un poème :

  Lors de l'ajout d'un poème il n'est pas obligatoire de lui trouver un titre si celui-ci n'en possède pas, cependant il est obligatoire d'affecter celui-ci à un recueil et de préciser son contenu. Son numéro, correspond à sa référence dans son recueil d'appartenance, il n'est donc pas demandé de le préciser.



3.2.2. Suppression

Après sélection de l'entité à supprimer un message de confirmation apparait, une fois la suppression confirmée toutes les entités « filles » le sont aussi (exemple : la suppression d'un auteur entraîne la suppression de tous ses recueils). Notez qu'il est possible de supprimer plusieurs entités à la fois.

3.2.3. Modification

Après sélection de l'entité à modifier, il n'est pas possible de modifier tous les attributs de celui-ci : les caractéristiques attrayant à une l'appartenance : comme l'appartenance à un recueil ou à un auteur ne sont pas modifiables. Lors d'une erreur d'insertion il est donc préférable de supprimer cette entité et de la recréer.

3.3. GESTION PAR FICHIERS TEXTE

3.3.1. Structure des fichiers texte

Le format de fichier utilisé pour la gestion des poèmes par texte est le *format .txt*. Il est très important que celui-ci soit formaté en UTF-8 sans cela des erreurs d'affichage de caractères s'opèreront !

Structure d'un poème :

Un poème formaté contient obligatoirement une balise de titre (ligne 3) annonçant le début du poème. Entre la balise ouvrante et fermante peut être précisé, se le poème en possède, son titre (dans notre exemple le poème n'a pas de titre). L'ID précédé du symbole « $ » représente son référencement dans son recueil d'appartenance. Les lignes 1 et 2 ne sont pas obligatoires à l'import. En effet celles-ci sont ajoutées à l'export afin de contextualiser le poème si par exemple il est extrait seul. Les lignes 4 à 11 sont quant à elles le contenu du poème : il est obligatoire.

Structure d'un recueil :

```
1-  @Auteur_nom : Delmotte @Auteur_prenom : Jacques @Auteur_pseudo : Izoard
2-  @Titre_recueil : La Patrie empaillee
3-  <Titre_poeme id=$105></Titre_poeme>
4-  Les lilas, les nerfs
5-  la main les touche ;
6-  la maison dans le poing
7-  serre les vieux habits.
8-  À présent, l'embellie,
9-  la jambe exacte.
10- Et tu respires
11- sans y penser.
```

Un recueil au format txt est composé obligatoirement d'un titre entre balise (ligne 2), ainsi que de poème(s) : dans notre exemple nous avons 3 poèmes. Les poèmes y sont alors formatés de la même manière que les fichiers ne contenant qu'un seul poème : avec une balise de titre (ligne 3, 13 et 21) indiquant le début de ceux-ci, dans lequel figure son id précédé d'un signe « $ », et un contenu (de la ligne 4 à 12 pour le premier poème). Le titre encore une fois peut ne pas être précisé.

De la même manière que pour les poèmes la ligne (1) est une contextualisation créée à l'export elle n'est pas nécessaire lors de l'import des recueils.

```
1-  @Auteur_nom : Delmotte @Auteur_prenom : Jacques @Auteur_pseudo : Izoard
2-  <Titre_recueil>Six poemes</Titre_recueil>
3-  <Titre_poeme id=$1>null</Titre_poeme>
4-  Toucher.
5-  Envenime l'épiderme ou le bossu,
6-  venin malin des aulnes.
7-  Bohémien.
8-  Bleue, l'esquirre.
9-  Le rite ensanglanté
10- que le duc dame.
11- Botte ou suture
12- ou mec au fronton.
13- <Titre_poeme id=$2>null</Titre_poeme>
14- Qui casse l'eau, casse l'aine
15- ou le héron tendu qui grogne
16- ou devient fétu, feu, fille.
17- De quel tandem, de quel idem
18- garder les noeuds ?
19- Noyés sont les conspirateurs
20- dès qu'on sonne l'alarme.
21- <Titre_poeme id=$3>null</Titre_poeme>
22- Bleu balbutieur sans enclume,
23- tu sais voyelles et corneilles,
24- Derechef. La ronde et douce.
25- et l'amère, et la douce.
26- La balbutie des reinettes
27- sous la saillie, la saisie.
```

3.3.2. Import de fichiers texte

- Recueil :

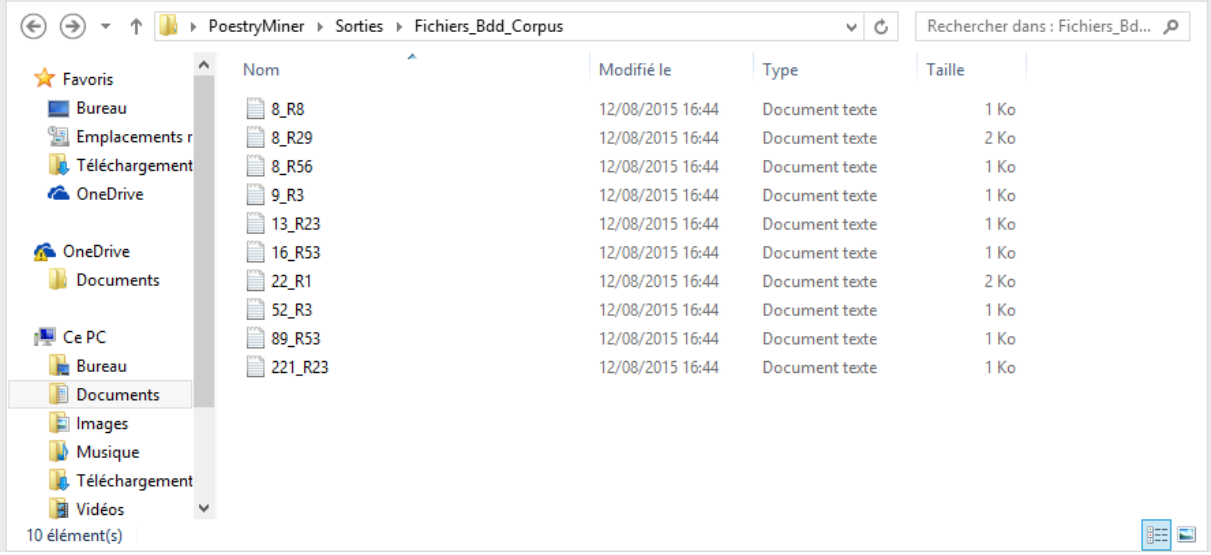
  Après sélection d'un auteur, il est possible de sélectionner plusieurs fichiers et de les attacher à celui-ci. Notez que plus l'utilisateur importera de recueils, plus la tâche sera longue. Afin d'importer pour la première fois le corpus d'un auteur il est donc conseillé de créer un auteur via l'interface « Auteur » puis de lui affecter par paquets successifs ses recueils.

- Poème :

  Après sélection d'un auteur et d'un recueil, il est possible de sélectionner plusieurs fichiers et de les attacher à celui-ci.

3.3.3. Export de fichiers texte

L'export de fichiers se fait après sélection d'une seule entité. Exporter un auteur correspond à enregistrer tous ses recueils dans un dossier. Les exports de poèmes et de recueils sont formatés selon les normes exposées dans la partie 2.2.1 de ce guide (Structure des fichiers texte), ils sont par défaut placés dans le dossier originel de Poestry Miner : « .\PoestryMiner\Sorties\Fichiers_Bdd_Corpus »



Notez que plus le nombre de fichiers à enregistrer est important plus cette tâche sera longue.

4. INTERFACE DE CREATION DE REQUETES

4.1. LES DIFFERENTS TYPES DE REQUETES

4.1.1. Les requêtes libres

Il existe deux types de requêtes libres celles portant sur un texte, ainsi que celles portant sur des mots. L'avantage majeur de ce type de requête réside dans le fait que leur computation est très rapide.

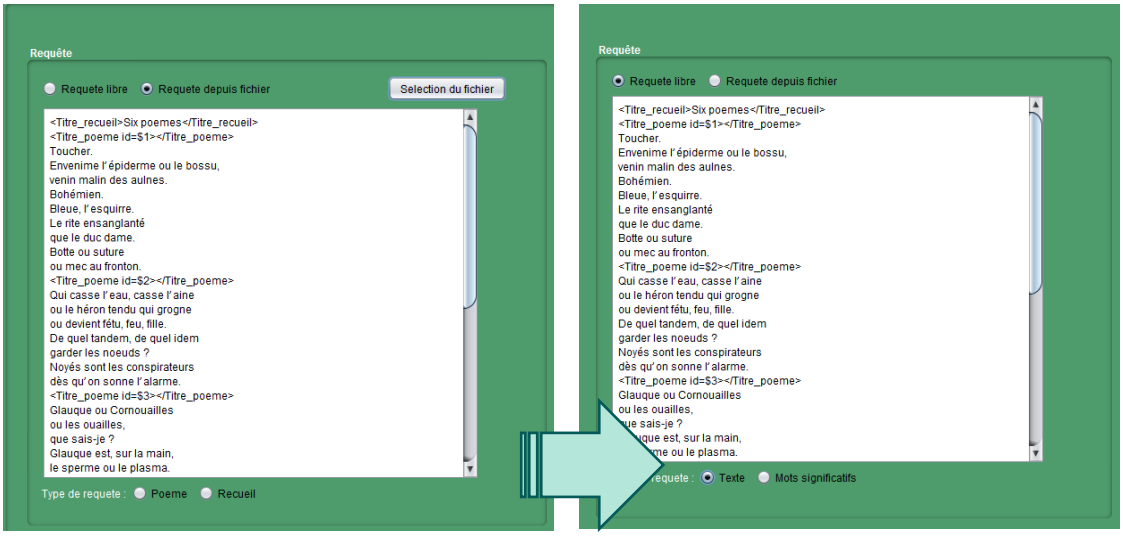
- Les requêtes textuelles :

  Elles sont adaptées aux textes que l'on souhaitera comparer avec notre corpus. Il s'agit ici de phrases ou d'extrait de poèmes formant une suite de mots cohérente.

- Les requêtes formées de mots significatifs :

  Elles sont quant à elles, des mots qui auront été choisis par l'utilisateur comme étant assez significatifs pour être comparés avec notre corpus.

➢ Il est notamment possible de considérer un poème ou un recueil comme une requête textuelle, cependant si un recueil est entré en tant que texte, les résultats découlant de la requête recueil seront traités comme un texte à part entière, dans lequel les poèmes ne seront pas reconnus. De plus, les titres de ceux-ci feront partie intégrante des termes de la requête… Cette solution est envisageable dans le cas où l'on souhaite raccourcir un recueil trop long ou le corriger à la suite d'une erreur d'entrée dans le fichier importé afin de l'éditer avant de le faire repasser par la suite comme une requête recueil.



4.1.2. Les requêtes depuis fichiers

Les requêtes depuis fichiers permettent de comparer un poème ou un recueil de poème avec le corpus entré en base de données à condition que ces fichiers soient formatés comme exprimé dans la partie 2.2.1 de ce guide (Structure des fichiers texte). Bien que permettant une plus petite intervention de l'utilisateur et un export aisé des résultats, les requêtes de type recueils peuvent parfois demander un long temps de calcul. En effet, le logiciel transforme le fichier en une suite de poèmes, elle-même divisée en suite de mots… Cependant entrer une requête en tant que poème ou recueil permet à celle-ci d'intégrer les métadonnées du texte afin de contextualiser celui-ci lors d'un export de résultats par exemple.

4.2. LE CHAMP DE RECHERCHE

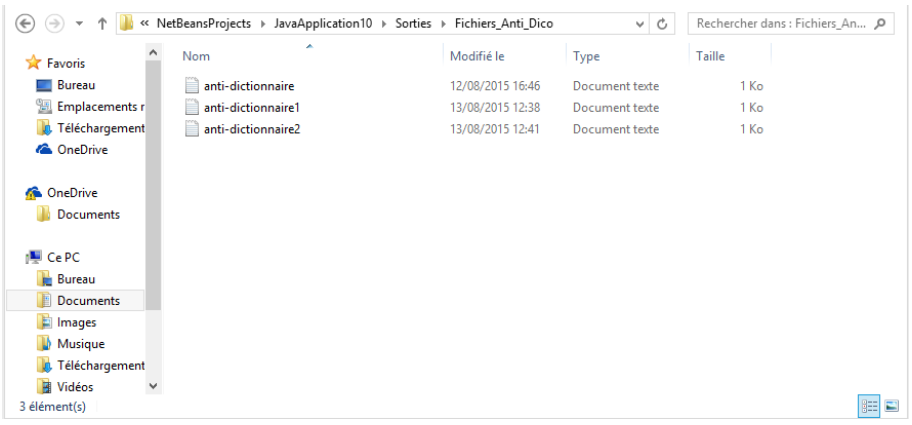
Le champ de recherche permet à Poestry Miner de savoir où chercher. Il est donc possible pour l'utilisateur d'effectuer sa recherche sur tous les recueils d'un auteur via la check box « Tous les recueils de l'auteur », si celui-ci veut affiner sa recherche. Il est aussi possible pour lui de sélectionner plusieurs recueils en maintenant la touche Ctrl enfoncée.

4.3. L'ANTI DICTIONNAIRE

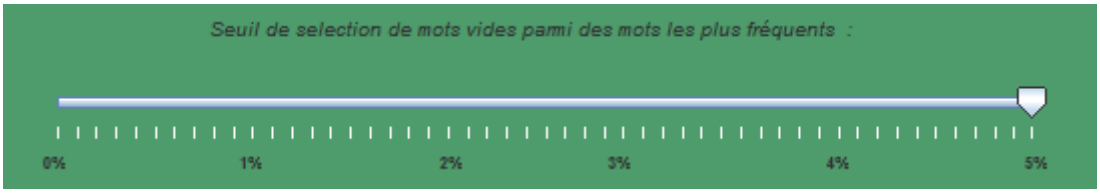
L'anti-dictionnaire, permet de ne pas prendre en compte les « mots vides » lors d'une recherche. En effet, de nombreux mots de liaison tendent à revenir très souvent au fil des recherches et parasitent parfois les résultats obtenus.

Afin d'utiliser un anti-dictionnaire, l'utilisateur doit cocher la check box « Utiliser un anti-dictionnaire » puis ensuite cliquer sur « Choix de l'anti-dictionnaire ».L'utilisateur a alors le choix entre un mode manuel ou un mode automatique.

Le mode manuel lui permet d'entrer soit directement des mots vides, soit de charger un fichier déjà enregistré au préalable. Les anti-dictionnaires sauvegardés sont enregistrés dans le fichier originel de Poestry Miner : PoestryMiner/Sorties/Anti_Dico :



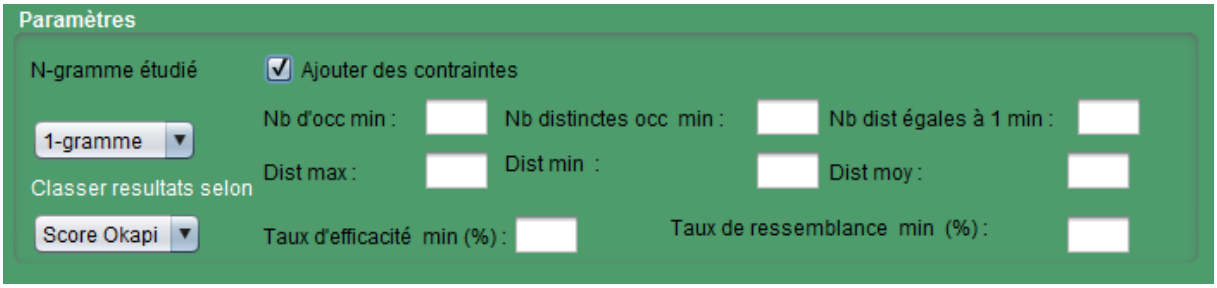
Le mode automatique permet de sélectionner automatiquement des mots vides à partir de la base de données (soit parmi le vocabulaire entier d'un auteur, ou seulement parmi les recueils sélectionnés pour la recherche) :



L'anti-dictionnaire récupère la liste des sacs de mots (1-gramme/2-gramme ou 3-gramme selon les paramètres de la requête) du champ de recherche classés selon leur fréquence dans le champ de recherche. Le seuil sélectionné permet donc à l'utilisateur de récupérer au maximum les 5% de cette liste.

➢ Cette fonctionnalité peut bien sûr être détournée afin de repérer les expressions récurrentes à un auteur à ou une série de recueils. Ainsi l'utilisateur poura utiliser ces sacs de mots en tant qu'entrés lors d'une requête basée sur des mots significatifs.

4.4. LES PARAMETRES



4.4.1. Le n-gramme

Le n-gramme est une séquence de n mots qui se suivent. Par défaut Poestry Miner intègre lors de l'import d'un poème dans sa base de données chaque 1-gramme, 2-gramme et 3-gramme. Le choix du n-gramme permettra de choisir la longueur des sacs de mots sur lesquels Poestry Miner effectuera ses comparaisons et calculs. Cependant, le texte d'entrée n'est pas découpé de la même manière selon le type de requête.

- Cas des requêtes texte, poème et recueil :

  Dans ce cas le programme découpe les textes en séquence de n mots qui se suivent : par exemple la séquence « le chat mange la souris » sera transformé en séquences de 1-gramme comme suis : « le » « chat » « mange » « la » « souris », en séquences de 2-gramme comme ceci : « le chat », « chat mange », « mange la », « la souris » et en séquences de 3-gramme de cette manière : « le chat mange », « chat mange la »,

« mange la souris ». Ce paramètre permet à l'utilisateur de se focaliser sur des segments de mots dans le corpus.

- Cas des requêtes de mots significatifs :

  Dans ce cas les n-grammes ne sont pas calculés de la même manière. En effet Poestry Miner considère le texte entré comme une séquence de n-gramme. Par exemple, la séquence « le chat noir danse le tango », dans le cas d'un calcul par 2-gramme sera considéré comme : « le chat », « noir danse », « le tango ». Les entrés sont donc considérées comme des n-grammes significatifs.

4.4.2. Le score de classement

Deux scores fréquemment utilisés en recherche d'information ont été implémentés dans le logiciel : le Tf-idf ainsi que le score BM 25 appelé aussi Okapi (du nom du premier système sur lequel celui-ci a été implémenté). Ces scores sont des indicateurs permettant de mesurer la pertinence d'un document en fonction de son contenu face à une requête. Le logiciel n'affichant que les 50 premiers poèmes les plus pertinents, classer les résultats via l'utilisation de l'un de ces indicateurs permet d'effectuer un premier écrémage assez précis. Le détail du calcul de ces indicateurs est proposé dans la partie 6.1. (Mode de calcul des indicateurs) de ce manuel.

- Le score TF-IDF :

  Cet indicateur est un calcul de poids prenant en compte le nombre d'occurrences trouvées dans un poème ainsi que le nombre d'occurrences de ce terme dans le corpus. En effet, plus un terme apparait dans un corpus moins il parait pertinent.

- Okapi BM25 :

  Cet indicateur est une version pondérée du tf-idf prenant mieux en compte la longueur des documents. En effet, plus un document est long, plus nous avons de chance d'y retrouver un terme courant.

4.4.3. Les contraintes

Les contraintes visent à limiter le nombre de résultats en fonction d'un désir utilisateur elles sont basées sur les indicateurs fournis en résultats dont le détail de leurs calcul est fourni parti 6.1. de ce manuel. Il en existe deux types :

- Celles basées sur un minimum à atteindre

  - le nombre d'occurrences minimum
  - nombre d'occurrences distinctes minimum
  - le nombre de distances égale à 1.
  - Le taux d'efficacité (à fournir en pourcentage)
  - Le taux de ressemblance (à fournir en pourcentage)

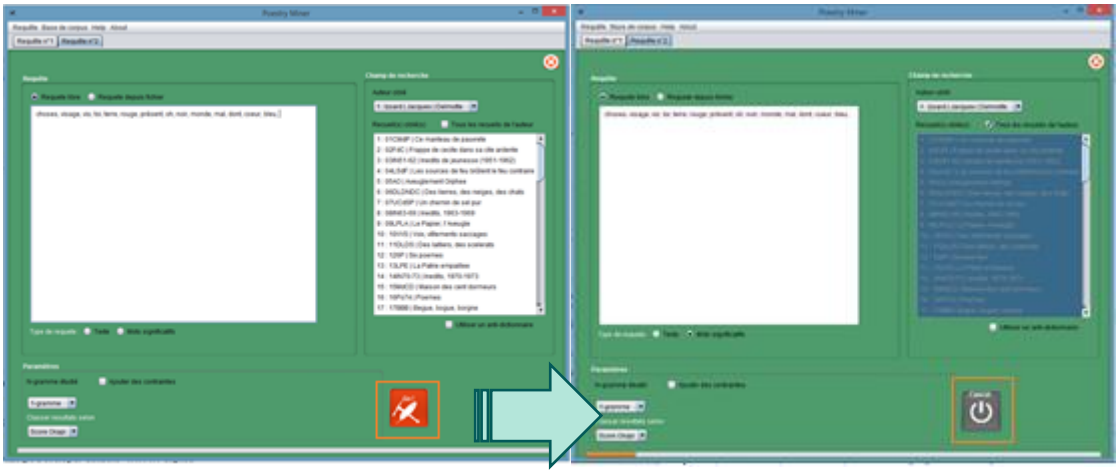
Le nombre d'occurrences minimum est par défaut égal à un, en effet nous ne pouvons calculer nos indicateurs de décisions sur les poèmes n'ayant aucun terme commun avec la requête.

- Celles basées sur un maximum à atteindre

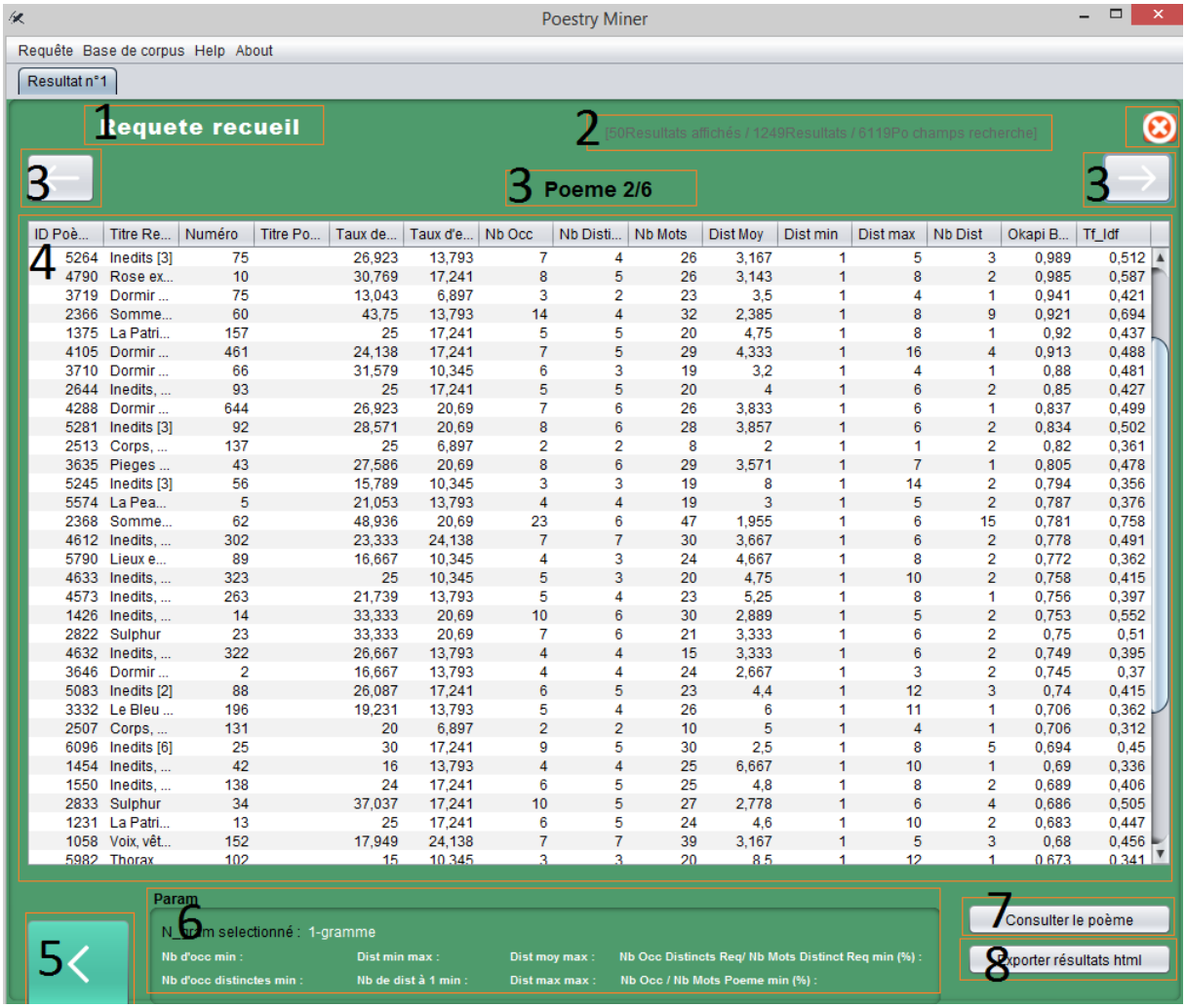
  - la distance maximale entre deux occurrences
  - la distance moyenne entre toutes les occurrences du poème
  - la distance minimale entre deux occurrences

5. LANCEMENT DE LA REQUÊTE

Une fois que l'utilisateur a fini de paramétrer sa requête il lui suffit d'appuyer sur le bouton orange. Il peut alors suivre l'avancement de celle-ci via la barre de progression, ainsi que l'annuler à tout moment via le bouton gris si celle-ci est par exemple trop longue.



6. INTERFACE DE RESULTATS



Une fois le calcul des résultats terminé l'utilisateur accède à l'interface de résultats. Elle offre une vue analytique à l'utilisateur, lui rappelant le type de requête ayant été utilisée (1), le nombre de poèmes affichés, par rapport au nombre de poèmes ayant au moins un motif en commun avec la requête et le nombre de poèmes du champ de recherche (2). Les résultats sont ainsi présentés sous forme tabulaire (4) listant les indicateurs des 50 poèmes les plus vraisemblables classés selon le score tf-idf ou Okapi. Pour chacun de ceux-ci, sont calculés divers indicateurs. L'utilisateur a la possibilité de trier ce tableau en cliquant sur l'indicateur qu'il souhaite utiliser en tant que critère de tri lors de l'affichage. Afin de visualiser un poème en particulier, il est possible de consulter celui-ci en cliquant sur le bouton de consultation (7), un export de la liste des résultats est ensuite possible via le bouton (8) permettant d'obtenir un fichier HTML présentant ceux-ci. Vous noterez que lors du lancement d'une requête de type recueil un tableau de résultat est fourni pour chacun des poèmes constituant le recueil. L'utilisateur a ensuite la possibilité de naviguer parmi les résultats via deux flèches (3) présentes en haut à droite et en haut à gauche du tableau. Enfin, si l'utilisateur souhaite modifier sa requête ou revenir sur celle-ci le bouton (5) lui permet d'accéder de nouveau à l'interface de création de requête.

6.1. MODE DE CALCUL DES INDICATEURS

- Nombre d'occurrences distinctes :

$Nb\ Occ\ (P_i)_{distinctes}$ est le nombre de termes apparaissant dans le poème courant et la requête.

- Nombre d'occurrences :

$$Nb\ Occ(P_i) = \sum_{Nb\ Occ\ (P_i)_{distinctes}}^{j=1} f(t_j) \quad \{t_j \in requête, et\ t_j\ \in P_i)\}$$

Où P_i est le poème courant, $f(t_j)$ la fréquence documentaire du terme t_j appartenant à la requête et au poème.

- Indice de ressemblance :

$$\frac{Nb\ Occ(P_i)}{Nb\ mots(Pi)} \quad \{t_j \in requête, et\ t_j\ \in P_i)\}$$

- Indice d'efficacité :

$$\frac{Nb\ Occ\ Distinctes\ (P_i)}{Nb\ mots(requête)} \quad \{t_j \in requête, et\ t_j\ \in P_i)\}$$

- Distance inter-occurrence :

$$d(t_i, t_{i+1}) = |Pos(t_{i+1}) - Pos(t_i)| \quad \{t_i, t_{i+1} \in requête, et\ t_i, t_{i+1} \in P_i)\}$$

Où $Pos(t_i)$ est la position du terme dans le poème

- Distance moyenne :

$$d(Pi)_{moyenne} = \frac{\sum_{Nb\ occ-1}^{i=1} d(t_i, t_{i+1})}{Nb\ occ - 1}$$

- Distance minimale :

$$d(Pi)_{minimale} = Arg\min(|Pos(t_{i+1}) - Pos(t_i)|)$$

- Distance maximale :

$$d_{(Pi)maximale} = Arg\max(|Pos(t_{i+1}) - Pos(t_i)|)$$

- Le score Tf-Idf :

$$w_{TF-IDF}(t,d) = tf(t,d) * \log(N/df(t))$$

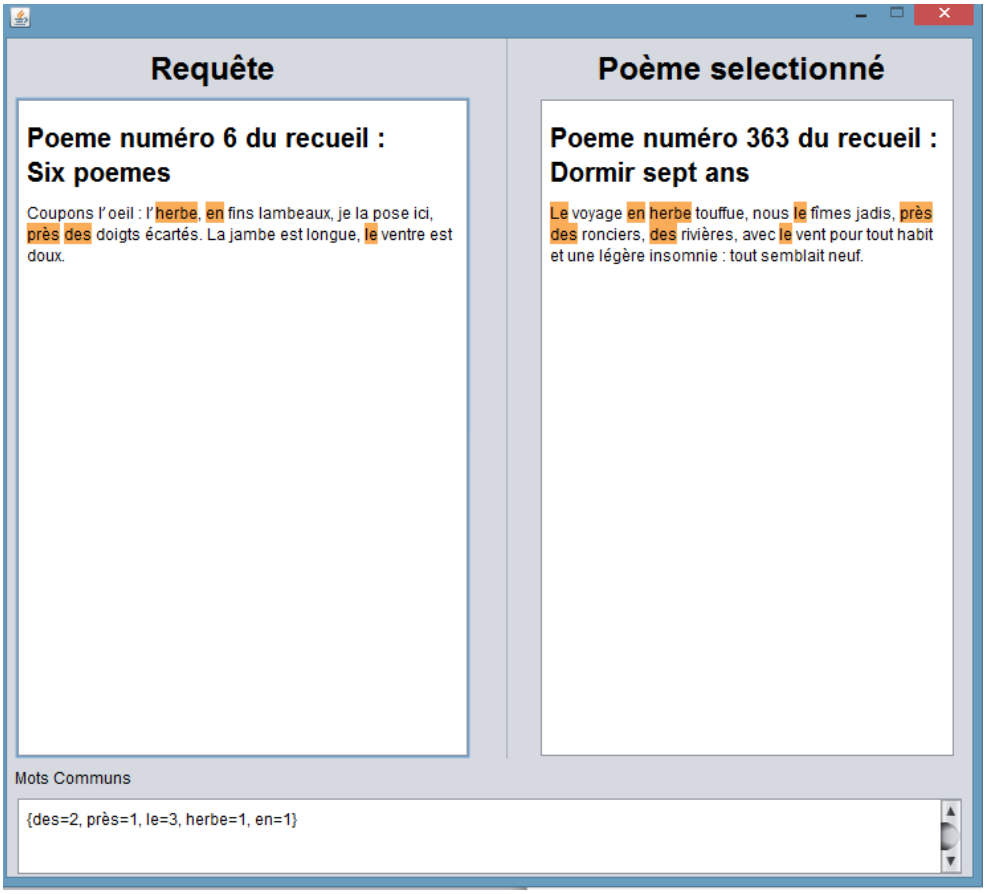
Où tf(t,d) correspond à la fréquence d'apparition du terme (soit le nombre d'occurrences présentes dans le poème t), et l'idf (log(N/df(t))) correspond au poids du terme considéré dans tout le corpus : N étant le nombre de documents et df(t) le nombre de poèmes dans lequel le terme t apparait. Pour un poème donné son tf-idf correspond donc à la somme de tous les tf-idf des termes le composant.

- Le score okapi :

$$\begin{aligned} w_{BM25}(t,d) &= TF_{BM25}(t,d) * IDF_{BM25}(t) \\ &= \frac{tf(t,d) * (k_1+1)}{tf(t,d) + k_1 * (1 - b + b * dl(d)/dl_{avg})} * \log \frac{N - df(t) + 0.5}{df(t) + 0.5}, \end{aligned}$$

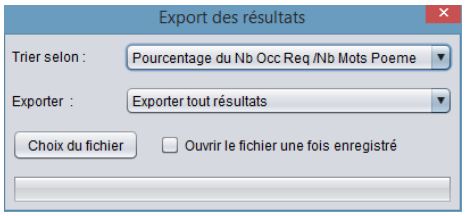
Où k1=2 et b=0.75 sont des constantes, dl(d) la longueur du poème considérée, dlavg la moyenne des poèmes du champ de recherche, N le nombre de documents. Le score Okapi d'un poème donné correspond donc à la somme de tous les scores okapi des termes le composant.

6.2. LA CONSULTATION D'UN POEME



Après avoir cliqué sur le bouton de consultation de poème cette fenêtre s'ouvre. Elle permet dans ce cas de comparer rapidement le poème de la requête ainsi que le poème sélectionné. Sont exposés face à face la requête ainsi que le poème sélectionné, les deux sont mis en forme selon les mots qu'ils possèdent en communs, qui sont de plus listé avec leurs nombre d'occurrences respectives dans l'encart de bas de fenêtre.

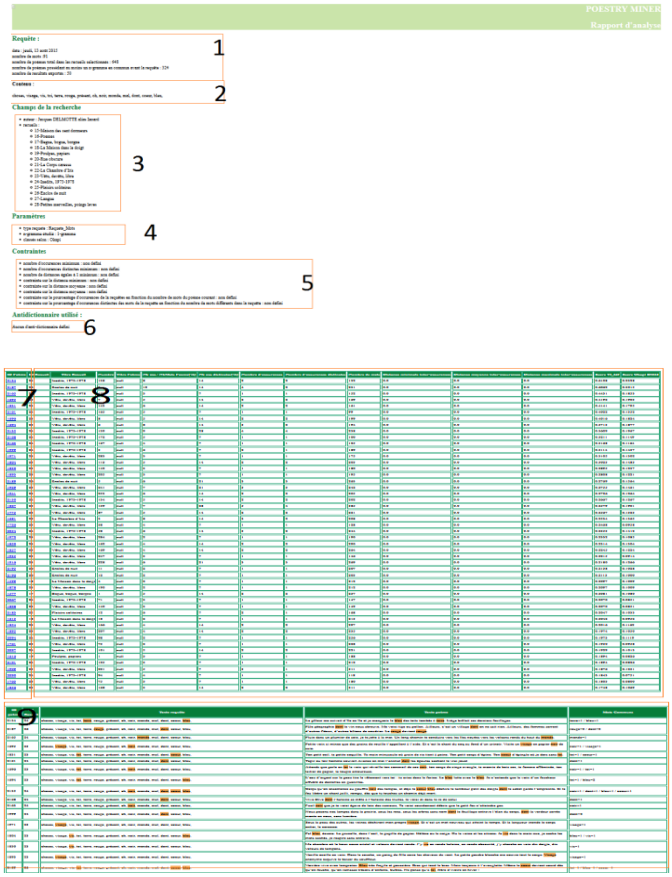
6.3. L'EXPORT DES RESULTATS



Une fois que l'utilisateur choisit d'exporter les résultats il a le choix de tous les exporter ou d'exporter les N premiers poèmes classés selon :

- le score tf-idf,
- le score okapi,
- le rapport entre le nombre d'occurrences et le nombre de mots du poème
- le rapport entre le nombre d'occurrences distinctes et le nombre de mots distincts présents dans la requête.

Il est possible pour celui-ci d'ouvrir le résultat exporté via la check box « Ouvrir le fichier une fois enregistré ». Les résultats une fois exportés sont présentés ceci :



L'utilisateur retrouve sur ces rapports d'analyse toutes les informations se rapportant à la recherche effectuée :

(1) sa date, le nombre de mots présents dans la requête, le nombre de poèmes sur lesquels a été effectuée la recherche, le nombre de poèmes ayant au moins un terme en commun avec la requête, et le nombre de résultats exportés
(2) Le contenu de la requête
(3) Le champ de recherche de la requête : l'auteur et les recueils sélectionnés par l'utilisateur
(4) Les paramètres de la requête : type, n-gramme étudié, classement des résultats présent dans le tableau
(5) Les contraintes définies par l'utilisateur
(6) Les mots présents dans l'anti dictionnaire
(7) L'id des poèmes permettant d'accéder directement à la comparaison mise en forme avec la requête
(8) Le tableau de résultats
(9) La comparaison de chaque poème avec la requête.