

Limes

Format pris en charge par Limes pour les fichiers mono-méthodes.

Limes : version 1.3 du 15/07/19

document : version du 07/08/19

auteur : J. Ducasse

Le terme « blanc » désigne le caractère espace ou la tabulation.

Dans tous les cas :

- Les lignes blanches ou vides sont ignorées ;
- Pour les mots clés (Group, Species, n, id, etc.), la casse est indifférente ;
- Le fichier est rejeté si :
 - il ne contient aucun spécimen ;
 - plusieurs spécimens portent le même nom.

Format ABGD

Les lignes ne commençant pas par :

Group [

avec un nombre quelconque de blancs (y compris 0) avant et après Group, sont ignorées ; seules les lignes commençant ainsi sont prises en compte. Ces lignes doivent avoir la forme suivante :

Group [<codespèce>] n : <nb> ; id : <liste de spécimens>

Il peut y avoir nombre quelconque de blancs (y compris 0) de part et d'autre de chaque terme.

<codespèce> est le code espèce : ce doit être un numéro quelconque, ≥ 0 . On ne vérifie pas la séquence numérique des codes espèce.

<liste de spécimens> est formé de mots séparés par un ou plusieurs blanc ou tab. Chaque mot constitue le nom d'un spécimen ; le nom d'un spécimen ne peut donc pas contenir de blancs. La liste des spécimens d'un Group est donnée sur une seule ligne ; elle ne peut être scindée sur plusieurs lignes.

Le fichier est rejeté si :

- Une ligne Group n'a pas la forme donnée ci-dessous ;
- Il n'y a aucune ligne Group ;
- Un même code espèce est utilisé dans plusieurs Group ;
- Le nombre de spécimens dans un Group ne correspond pas au nombre annoncé par « n : ».

Format PTP

Le contenu utile du fichier commence à la première ligne de la forme :

`Species <codespèce> (`

avec un nombre quelconque de blancs (y compris 0) avant et après chaque terme, sauf `Species` et `<codespèce>` qui doivent être séparés par au moins un blanc.

Toutes les lignes situées avant la première ligne `Species` sont ignorées.

Le fichier est ensuite formé d'une alternance de lignes `Species` et de lignes donnant la liste des spécimens.

Les lignes `Species` doivent avoir la forme suivante :

`Species <codespèce> (.....)`

La paire de parenthèses doit être présente, mais son contenu est ignoré.

`<codespèce>` est le code espèce : ce doit être un numéro quelconque, ≥ 0 . On ne vérifie pas la séquence numérique des codes espèce.

Les lignes de spécimens débutent obligatoirement par un blanc ou une tabulation. Les noms de spécimens sont séparés par une virgule, avec un nombre quelconque de blancs avant et après celle-ci. La liste des spécimens d'un `Species` est donnée sur une seule ligne ; elle ne peut être scindée sur plusieurs lignes.

Le fichier est rejeté si :

- Une ligne `Species` n'a pas la forme donnée ci-dessous ;
- Il n'y a aucune ligne `Species` ;
- Un même code espèce est utilisé dans plusieurs `Species` ;
- Une ligne de spécimens ne débute par un blanc ou une tabulation.

Format GMYC

Le contenu utile du fichier commence à la première ligne de la forme :

`## GMYC_spec sample_name`

avec un nombre quelconque de blancs (y compris 0) avant et après chaque terme, à l'exception des deux mots qui doivent être séparés par au moins un blanc. La casse est indifférente.

Toutes les lignes situées avant celle-ci sont ignorées.

Toutes les lignes suivantes doivent être de la forme :

`## <numéro> <codespèce> <spécimen>`

avec un nombre quelconque de blancs (y compris 0) avant et après chaque terme, sauf le code espèce séparé par au moins un blanc avant et après.

<codespèce> est le code espèce : ce doit être un numéro quelconque, ≥ 0 .

Le nom du spécimen va du premier caractère non blanc suivant le <codespèce> au dernier caractère non blanc de la ligne ; il peut donc contenir des blancs.

La valeur du <numéro> est ignorée (on ne vérifie pas la séquence).

Mise en correspondance des Méthodes.

Lorsque l'on charge plusieurs fichiers pour construire un espace de N méthodes, chaque fichier apportant une méthode, il est nécessaire de faire correspondre les noms des échantillons entre les différents fichiers.

Pour mettre en correspondance les noms des spécimens (échantillons) entre les méthodes, ceux-ci sont normalisés :

- toutes les lettres sont mises en minuscules, ce qui permet d'ignorer les variations de casses ;
- toutes les séquences de caractères autres que lettres, chiffres ou souligné sont remplacées par un souligné unique.

S'il existe des spécimens qui existent dans certaines méthodes mais pas dans toutes, après cette normalisation, ils sont écartés ; leur liste est affichée. Les calculs se font avec les spécimens restant, présents dans toutes les méthodes.

Le chargement de l'ensemble des N fichiers échoue :

- si la normalisation des noms génère des doublons au sein d'un même fichier (d'une même méthode) ;
- s'il n'y a aucun spécimen commun à l'ensemble des méthodes.