

Master Ingénierie Linguistique

Gestion Informatique du Multilinguisme

TD 2

Soit 2 documents contenant le même texte et occupant respectivement 13 Ko et 140 octets sur mon disque dur :

- « representation.odt »
- « representation.txt » qui a été construit à partir du précédent.

J'ouvre chacun avec les applications utilisées par défaut par mon système.

- Quelle est l'application qui s'ouvre avec « representation.odt » ?
- Quelle est l'application qui s'ouvre avec « representation.txt » ?

À première vue, le texte semble bien être le même, mais le rendu est tout à fait différent :

- Comment peut-on expliquer ces différences d'affichage, ainsi que les différences d'occupation de l'espace disque.

Afin de m'assurer que le texte est bien le même, je commence par utiliser dans chaque application, les fonctions qui me renseignent sur la composition du texte édité :

- Dans mon traitement de texte, j'active la fonction « Propriétés » qui m'annonce 133 caractères, 15 mots et 5 lignes
- En revanche, la barre de statut de mon éditeur de texte (TextWrangler) me dit 140 caractères, 15 mots et 8 lignes.

Je vérifie caractère par caractère et je compte : il y a bien 5 mots et je ne vois que 5 lignes qui comportent respectivement :

- ligne 1 : 18 signes graphiques + 5 blancs
- ligne 2 : 14 signes graphiques
- ligne 3 : 0, la ligne semble vide, mais dois-je compter un blanc ?
- ligne 4 : 56 signes graphiques + 3 blancs
- ligne 5 : 32 signes graphiques + 3 blancs

Dans les 2 cas, le texte comporte bien 120 signes graphiques et 11 blancs ou 12 si j'en compte un sur la ligne 3, soit 131 ou 132 caractères.

Ne sachant à quels chiffres me fier, je passe sous Linux et j'édite « representation.txt » avec « Kate ». dont la barre de statut annonce 8 lignes et 133 caractères, puis avec « Gedit » dont la fonction « Statistiques du document » déclare 7 lignes, 15 mots, 139 caractères avec les espaces, 120 sans et 144 octets.

- Que faire ? Comment expliquer ces différents décomptes et comment puis-je vérifier le nombre de caractères que contient mon texte ?