

Groupe de travail Réseau
Request for Comments : 678
NIC: 31524

J. Postel, (SRI-ARC)
19 décembre 1974
Traduction Claude Brière de L'Isle

Formats de fichier standard

Introduction

En essayant de fournir des documents en ligne à la communauté du réseau nous avons rencontré de nombreux problèmes avec le format physique des documents finaux. Beaucoup de ces difficultés tiennent au fait que nous n'avons pas le contrôle ou même la connaissance de toutes les étapes du traitement ou des appareils qui agissent sur le fichier du document. Une grande partie de la difficulté dans le passé a été due à des hypothèses que nous avons faites en considérant que le reste du monde était approximativement comme notre propre environnement. Nous voyons maintenant que les problèmes sont dus à des hypothèses et des traitements différents sur les fichiers à imprimer comme documents. Nous proposons donc de définir certains formats standard pour les fichiers et de décrire la forme finale attendue pour les copies imprimées de tels fichiers.

Ces formats standard ne sont pas des types/modes/structures supplémentaires de données du protocole de transfert de fichier, mais plutôt des descriptions d'utilisations entre l'origine et le receveur ultime du fichier. Il peut être utile ou même nécessaire pour certains hôtes de construire des programmes qui convertissent les fichiers des formats locaux communs dans les formats standard spécifiés ici.

L'intention est que l'auteur d'un document puisse préparer son texte et le mémoriser dans un fichier en ligne, puis annoncer ce fichier par son nom et son format (comme spécifié ici), de telle façon que les individus intéressés puissent copier et imprimer le fichier en pleine connaissance des caractéristiques des commandes de format et de la taille de la page logique.

Éléments de normalisation

Les éléments ou aspects d'un fichier qui sont à normaliser sont le jeu de caractères utilisé, les procédures de contrôle de format, la zone de la page à utiliser pour le texte, et la méthode pour décrire les caractères soulignés ou en surimpression.

La zone d'une page à utiliser pour le texte peut être source de confusion dans la discussion, aussi pour essayer d'être clairs nous définirons une page physique et une page logique. Prière de noter que le principale objet de la présente note est de décrire les formats standard en termes de page logique, et qu'il revient à chaque site de transposer la page logique en page physique de chacun de ses appareils.

Page Physique

La page physique est le support qui porte le texte, la hauteur et la largeur de sa zone sont mesurées en pouces. La page physique normale est une pièce de papier de onze pouces de haut et de huit pouces et demi de large.

La densité normale d'impression est de 10 caractères par pouce en horizontal et de 6 caractères par pouce en vertical. IL en résulte que la page physique normale a une capacité maximum de 66 lignes et de 85 caractères par ligne. Il arrive souvent que les appareils d'impression limitent la zone de la page physique en mettant en application des marges.

Page logique

La page logique est la zone qui contient le texte. La hauteur de cette zone est mesurée en lignes et la largeur est mesurée en caractères.

Une page logique normale fait 60 lignes de haut et 72 caractères de large.

Jeu de code

Le codage des caractères sera le code standard du réseau de terminal virtuel réseau (NVT, *Network Virtual Terminal*) tel qu'utilisé dans les protocoles Telnet et de transfert de fichiers, c'est-à-dire en ASCII en octets de huit bits dont le bit de poids fort est zéro.

Contrôle de format

Le format sera contrôlé par les opérateurs de format ASCII :

Saut de page (*Form Feed*) <FF>

Déplace l'impression au sommet de la prochaine page logique en gardant la même position horizontale.

Retour chariot <CR>

Déplace l'impression au bord gauche de la page logique en restant sur la ligne actuelle.

Saut à la ligne (*Line Feed*) <LF>

Déplace l'impression à la prochaine ligne d'impression, en gardant la même position horizontale.

Tabulation horizontale <HT>

Déplace l'impression à la prochaine marque de tabulation horizontale.

Les marques conventionnelles des tabulations horizontales sont tous les huit caractères, c'est-à-dire les positions de caractère 9, 17, 25, ... au sein de la page logique.

Noter qu'il est difficile de mettre en application ces conventions et il est donc recommandé que les tabulations horizontales ne soient pas utilisées dans les fichiers de documents.

Tabulation verticale <VT>

Déplace l'impression à la prochaine marque de tabulation verticale.

Les marques conventionnelles des tabulations verticales sont toutes les huit lignes en commençant à la première ligne d'impression sur chaque page logique, c'est-à-dire les lignes 1, 9, 17, ... au sein de la page logique.

Noter qu'il est difficile de mettre en application ces conventions et il est donc recommandé de ne pas utiliser les tabulations verticales dans les fichiers de documents.

Espace arrière (*Back Space*) <BS>

Déplace l'impression d'une position de caractère vers le bord gauche de la page logique.

Tous ces opérateurs ne sont pas utilisés dans tous les formats standard ; tout opérateur qui n'est pas utilisé dans un format standard est ignoré.

Longueur de page

La longueur de la page logique sera spécifiée en termes de nombre de lignes de texte.

Largeur de page

La largeur de la page logique sera spécifiée par un nombre de caractères.

Surimpression

La surimpression (noter que le soulignement est un sous-ensemble de la surimpression) peut être spécifiée d'une des deux façons suivantes, ou pas du tout :

Par ligne

La ligne composite est constituée de segments de texte terminés chacun par la séquence <CR><NUL> excepté le segment final qui est terminé par la séquence <CR><LF>.

Par caractère

Chaque caractère qui subit la surimpression est suivi immédiatement par un <BS> et le caractère qui vient en surimpression.

Fin de ligne

La convention de fin de ligne est la convention Telnet de fin de ligne avec la séquence <CR><LF>. Il est recommandé que l'utilisation de <CR> et <LF> soit évitée dans d'autres contextes que la fin de ligne.

Formats standard

Format 1 [Document de base]

Ce format est conçu pour être utilisé pour des documents à imprimer sur des imprimantes en ligne, qui ont normalement 66 lignes sur une page physique, mais ont souvent des marges forcées de haut et de bas de 3 lignes chacune.

Opérateurs de format actifs : <FF>, <CR>, <LF>.

Longueur de page : 60 lignes.

Largeur de page : 72 caractères.

Surimpression : par ligne.

Format 2 [Terminal]

Ce format est conçu pour une utilisation avec des terminaux à copie papier, qui dans le cas normal ont 66 lignes sur une page physique. On s'attend à ce qu'il n'y ait pas de mise en application de marge de haut et de bas par le terminal ou son système local, et donc toutes les marges autour de la coupure de page physique doivent venir du fichier.

Opérateurs de format actifs : <FF>, <CR>, <LF>, <HT>, <VT>, <BS>.

Longueur de page : 66 lignes.

Largeur de page : 72 caractères.

Surimpression : par caractère.

Format 3 [Imprimante en ligne]

Ce format est conçu pour une utilisation avec une sortie sur une imprimante en ligne de pleine largeur (papier de 11 pouces sur 14).

Opérateurs de format actifs : <FF>, <CR>, <LF>.

Longueur de page : 60 lignes.

Largeur de page : 132 caractères.

Surimpression : Aucune.

Format 4 [Image de carte]

Ce format est conçu pour une utilisation d'entrée de carte simulée. La largeur de page est de 80 caractères, chaque image de carte est suivie par <CR><LF>, et donc chaque carte est représentée par 2 à 82 caractères dans le fichier. Noter que les espaces en queue d'une image de carte n'ont pas besoin d'être présents dans le fichier, et que la première occurrence de la séquence <CR><LF> indique que le reste de l'image de carte contient des caractères d'espace.

Opérateurs de format actifs : <CR>, <LF>.

Longueur de page : Infinie.

Largeur de page : 80 caractères.

Surimpression : Aucune.

Format 5 [Document centré]

Ce format est destiné à une utilisation avec des documents à imprimer sur des imprimantes en ligne qui ont normalement 66 lignes de page physique mais mettent en application des marges de haut et de bas de 3 lignes chacune. Le texte est censé être centré sur le papier. Si la densité d'impression horizontale est de 10 caractères par pouce et que le papier est large de 8 pouces et demi, il y aura alors une marge d'un pouce de chaque côté.

Opérateurs de format actifs : <FF>, <CR>, <LF>.

Longueur de page : 60 lignes.

Largeur de page : 65 caractères.

Surimpression : par ligne.

Format 6 [Document lié]

Ce format est destiné à une utilisation avec des documents à imprimer sur des imprimantes en ligne qui ont normalement 66 lignes de page physique mais mettent en application des marges de haut et de bas de 3 lignes chacune. Si la densité d'impression horizontale est de 10 caractères par pouce et que le papier est large de 8 pouces et demi, le texte devrait alors être positionné de telle façon qu'il y ait une marge à gauche d'un pouce et demi et une marge à droite d'un pouce.

Opérateurs de format actifs : <FF>, <CR>, <LF>.

Longueur de page : 60 lignes.

Largeur de page : 60 caractères.

Surimpression : par ligne.

Suggestions de mise en œuvre

Débordement

Le débordement peut résulter de deux causes : soit si la page physique est plus petite que la page logique, et aussi si le texte réel dans le fichier viole la norme selon laquelle il est traité.

Dans l'un et l'autre cas, les suggestions suivantes sont faites à ceux qui mettent en œuvre les programmes qui traitent des fichiers dans ces formats.

Longueur

Si plus de lignes sont traitées que n'en tient le minimum de la longueur de la page physique et de la page logique depuis les dernier <FF>, l'action <FF> devait alors être forcée.

Largeur

Si plus de positions de caractères sont traités qu'il n'en tient dans le minimum de largeur de la page physique et de la largeur de la page logique depuis le dernier <CR>, les caractères sont alors éliminés jusqu'au prochain <CR>.

ou

Si plus de positions de caractères sont traités que n'en tient la largeur minimum de la page physique et la largeur de la page logique depuis les dernier <CR>, les actions <CR> et <LF> devraient alors être forcées.

Références

A. McKenzie "TELNET Protocol Specification," Aug-73, NIC 18639.

"USA Standard Code for Information Interchange," United States of America Standards Institute, 1968, NIC 11246.