

Codage Typographie

E. Jeandel

Emmanuel.Jeandel at lif.univ-mrs.fr

- On essaie de convertir une suite de caractères en un texte imprimé.
- Comment afficher des mots ?
- Comment afficher les paragraphes ?

Aujourd'hui, uniquement le second point.

Justification - Formulation (simple) du problème

- On se donne une liste de “mots”, chacun faisant une longueur w_i ;
- Chaque espace a une longueur idéale de id , mais qui peut varier entre min et max ;
- On peut, si on est forcé, faire une espace dont la longueur est plus grande que max , mais jamais dont la longueur est plus petite que min ;
- Chaque ligne fait une longueur de l .

Le but est de répartir les mots sur les lignes, de façon si possible élégante.

Exemple

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

- $\min = 0.7, \max = 2, \text{id} = 1$
- $l = 10, 20, 30$

- Dans chaque ligne, on met le maximum possible de mots, en mettant tous les espaces à leur longueur id ;
- On essaie d'ajouter le mot qui ne passe pas, si cela est possible sans que les espaces n'aient une longueur trop petite.
- Sinon, le mot passe à la ligne suivante.

Exemple

l=20

Au premier de ces
trois golfes
appartient le port
de Vitylo. Creusé à
la lisière de sa rive
orientale, au fond
d'une anse
irrégulière, il
occupe les premiers
contreforts
maritimes du
Taygète, dont le
prolongement
orographique forme
l'ossature de ce
pays du Magne. La

Exemple

l=30

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers

Exemple

l=40

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

Exemple

l=50

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

Méthode Best-Fit

- Dans chaque ligne, on cherche à s'approcher le plus possible de l , en changeant le moins possible la valeur de l'espace.
- Soit e la largeur de l'espace.
 - $r = (e - id)/(id - min)$ si $e < id$
 - $r = (e - id)/(max - id)$ si $e > id$
 - $r = 0$ dans le cas idéal, $r = -1$ si on est obligé de compresser les espaces au maximum, $r = 1$ si on est obligé de les étendre au maximum.
- Critère possible : Minimiser la valeur de $|r|$
- Note : r ne doit jamais être inférieur à -1 (sinon les mots sont trop collés)

Exemple

l = 20

Au premier de ces
trois golfes
appartient le port
de Vitylo. Creusé à
la lisière de sa rive
orientale, au fond
d'une anse
irrégulière, il
occupe les premiers
contreforts
maritimes du
Taygète, dont le
prolongement
orographique forme
l'ossature de ce
pays du Magne. La

Exemple

l = 30

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers

Exemple

l = 40

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

Exemple

l = 50

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

Méthode Total-Fit

- On se fixe un critère pour chaque ligne qu'on va appeler b_i (erreur sur la ligne i)
- $b_i = 0$ si la ligne est parfaite, $b_i = 10000$ si la ligne est vraiment ratée.
- On trouve un arrangement qui minimise la somme totale des erreurs $e = \sum b_i$

Comment trouver cet arrangement ? (Essayer tous les arrangements est trop coûteux)

Total-Fit - Programmation Dynamique

- Soit $T[i]$ l'erreur e qu'on fait en coupant après le i ème mot, et en disposant les mots précédents de façon à minimiser l'erreur totale.
- Comment calculer $T[i]$ connaissant tous les précédents ?
- $T[i]$ ne dépend que
 - Des mots sur la dernière ligne (celle qu'on coupe après le i ème mot)
 - De l'erreur commise en coupant avant ces mots.
- $M[k]$ est l'erreur qu'on commet en mettant les k derniers mots sur la dernière ligne, et en plaçant les autres pour obtenir une erreur de $T[i - k]$.
- $T[i] = \min_{k \leq i} M[k]$

Total-Fit - Programmation Dynamique

- Soit $T[i]$ l'erreur e qu'on fait en coupant après le i ème mot, et en disposant les mots précédents de façon à minimiser l'erreur totale.
- Comment calculer $T[i]$ connaissant tous les précédents ?
- $T[i]$ ne dépend que
 - Des mots sur la dernière ligne (celle qu'on coupe après le i ème mot)
 - De l'erreur commise en coupant avant ces mots.
- $M[k]$ est l'erreur qu'on commet en mettant les k derniers mots sur la dernière ligne, et en plaçant les autres pour obtenir une erreur de $T[i - k]$.
- $T[i] = \min_{k \leq i} M[k]$

Total-Fit - Programmation Dynamique

- Soit $T[i]$ l'erreur e qu'on fait en coupant après le i ème mot, et en disposant les mots précédents de façon à minimiser l'erreur totale.
- Comment calculer $T[i]$ connaissant tous les précédents ?
- $T[i]$ ne dépend que
 - Des mots sur la dernière ligne (celle qu'on coupe après le i ème mot)
 - De l'erreur commise en coupant avant ces mots.
- $M[k]$ est l'erreur qu'on commet en mettant les k derniers mots sur la dernière ligne, et en plaçant les autres pour obtenir une erreur de $T[i - k]$.
- $T[i] = \min_{k \leq i} M[k]$

Total-Fit - Programmation Dynamique

- Soit $T[i]$ l'erreur e qu'on fait en coupant après le i ème mot, et en disposant les mots précédents de façon à minimiser l'erreur totale.
- Comment calculer $T[i]$ connaissant tous les précédents ?
- $T[i]$ ne dépend que
 - Des mots sur la dernière ligne (celle qu'on coupe après le i ème mot)
 - De l'erreur commise en coupant avant ces mots.
- $M[k]$ est l'erreur qu'on commet en mettant les k derniers mots sur la dernière ligne, et en plaçant les autres pour obtenir une erreur de $T[i - k]$.
- $T[i] = \min_{k \leq i} M[k]$

- Pour $i = 0$ à n
- Pour $k = 1$ à i
 - Calculer la valeur esthétique de la dernière ligne en supposant qu'elle contient les mots de $i + 1 - k$ à $i + 1$.
 - On appelle cette quantité $\beta[k]$
- $T[i + 1]$ est alors le minimum des $(T[i + 1 - k] + \beta[k])$

Exemple

l=20

Au premier de
ces trois golfes
appartient le port
de Vitylo. Creusé
à la lisière de
sa rive orientale,
au fond d'une
anse irrégulière,
il occupe les
premiers contreforts
maritimes
du Taygète, dont
le prolongement
orographique forme
l'ossature de ce
pays du Magne.

Exemple

l=30

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers

Exemple

l=40

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

Exemple

l=50

Au premier de ces trois golfes appartient le port de Vitylo. Creusé à la lisière de sa rive orientale, au fond d'une anse irrégulière, il occupe les premiers contreforts maritimes du Taygète, dont le prolongement orographique forme l'ossature de ce pays du Magne. La sûreté de ses fonds, l'orientation de ses passes, les hauteurs qui le couvrent, en font l'un des meilleurs refuges d'une côte incessamment battue par tous les vents de ces mers méditerranéennes.

Total-Fit - Conclusion

- Total-Fit donne en moyenne des meilleurs résultats.
- Il dépend (comme les autres) beaucoup des définitions qu'on donne pour les erreurs b_j et e
- En pratique, on prend plutôt $b_j = 1 + |r|^3$
- Mais ces facteur dépendent d'autres paramètres. . .

Pertinence du modèle

Le modèle simple (mots, espace) est bien, mais pas suffisant.

- Les lignes n'ont pas toutes la même longueur (dessin à côté, indentation)
- Les espaces n'ont pas toutes la même longueur
- Certaines espaces ne peuvent pas être remplacées par un retour à la ligne (l'espace insécable, par exemple avant le symbole “ :”)
- On peut choisir de couper des mots (césure, *hyphenation* en anglais) ;
- Couper des mots a un prix : on ne veut pas que tous les mots soient coupés.

On peut facilement changer le modèle pour tenir compte d'espaces de longueur variable. Le reste est plus compliqué.

Modèle colle-pénalité (glue-penalty)

- Une boîte représente un mot ou un bout de mot, un caractère. . .
- Une boîte a une longueur fixe l .

Modèle colle-pénalité (glue-penalty)

- Une colle représente une espace ;
- Une colle est représenté par 3 valeurs (id, min, max) ;
- Ces nombres peuvent être négatifs ;
- Lorsqu'on effectue un retour à la ligne, la colle **après** le retour à la ligne est supprimée ;
- En pratique, la notion de colle est plus compliquée ;
- Comment représente-t-on un espace de longueur fixe ?

Modèle colle-pénalité (glue-penalty)

- Une pénalité est un endroit où on peut couper le texte ;
- Cependant, couper à cet endroit représente un coût p ;
- Couper à cet endroit ajoutera à la ligne une longueur l .
- Exemple typique : une césure.

Comment voir un texte dans ce contexte ?

- Chaque mot correspond à une boîte
- Si un mot a une césure, par exemple `maison`, on le transforme en deux boîtes, (une pour `mai`, l'autre pour `son`) séparée par une pénalité de coût modéré, et de longueur l correspondant à la longueur de la césure (le “-”)
- Intuition : on peut couper la ligne à cet endroit, mais ça va nous coûter.

Comment voir un texte dans ce contexte ?

- Chaque espace est représentée par une colle ;
- Les colles peuvent avoir des longueurs différentes
- Par exemple : la colle avant le symbole ‘ : ’ est deux fois plus petite que la colle après le symbole ;
- Une indentation de la ligne initiale sera représentée par une boîte sans contenu d’une certaine longueur ;
- On représente la fin du paragraphe d’une certaine façon (à voir plus loin)

Exemple

Au pre † mier de ces trois golfes ap † par † tient
le port de Vitylo. Creusé à la lisière de sa
rive orientale, au fond d'une anse ir † ré † gu † lière,
il oc † cupe les pre † miers contre † forts ma † ri † time
du Taygète, dont le pro † lon † ge † ment oro † gra † phiq
forme l'os † sa † ture de ce pays du Magne. La
sûreté de ses fonds, l'orien † ta † tion de ses passe
les hau † teurs qui le couvrent, en font l'un de
re † fuges d'une côte in † ces † sam † ment bat † tue par
tous les vents de ces mers mé † di † ter † ra † néennes.

Exemples

- Représentation d'un saut de ligne interdit
- Représentation d'un saut de ligne obligatoire
- Une espace insécable

- Représentation d'un saut de ligne interdit
 - Une pénalité de coût infini ∞ .
- Représentation d'un saut de ligne obligatoire
- Une espace insécable

Exemples

- Représentation d'un saut de ligne interdit
 - Une pénalité de coût infini ∞ .
- Représentation d'un saut de ligne obligatoire
- Une espace insécable

Exemples

- Représentation d'un saut de ligne interdit
 - Une pénalité de coût infini ∞ .
- Représentation d'un saut de ligne obligatoire
 - Une pénalité de coût $-\infty$
- Une espace insécable

Exemples

- Représentation d'un saut de ligne interdit
 - Une pénalité de coût infini ∞ .
- Représentation d'un saut de ligne obligatoire
 - Une pénalité de coût $-\infty$
- Une espace insécable

Exemples

- Représentation d'un saut de ligne interdit
 - Une pénalité de coût infini ∞ .
- Représentation d'un saut de ligne obligatoire
 - Une pénalité de coût $-\infty$
- Une espace insécable
 - Une pénalité de coût $+\infty$
 - Une colle

- On peut représenter la fin de paragraphe de la façon suivante :
 - Une pénalité de coût $+\infty$
 - Une colle de longueur nulle, mais qui peut devenir aussi grande que l'on veut ;
 - Une pénalité de coût $-\infty$

Aligner à gauche et à droite

- On peut aligner des lignes à gauche avec des colles
- On remplace chaque espace de longueur c par
 - Une colle de longueur 0 mais qui peut s'étendre de 80
 - Une pénalité
 - Une colle de longueur c mais qui peut s'étendre de -80 .
- Idée : Si on coupe, c'est sur la pénalité (c'est idiot de couper avant, puisqu'on a tout intérêt à accepter en plus la colle)
- Si on coupe, la première colle reste mais pas la deuxième
- Si on ne coupe pas, l'expression revient à une colle de longueur c .
- (Note : pour que ça marche, il faut définir plus précisément les colles)

Ajustement à l'algorithme Total-Fit

- On change la formule.
- Le facteur r doit être calculé pour chaque colle (puisque ce ne sont pas toutes les mêmes)
- On obtient un facteur β qui permet d'évaluer à quel point les colles sont proches de leur longueur idéale
- On ajoute un terme π si jamais on a coupé à une pénalité
- Chaque ligne contribue donc un terme $\beta + \pi$
- On peut également corriger les formules pour éviter trois césures consécutives.

Texte aligné à gauche

- On peut utiliser les algorithmes précédents lorsque le paragraphe sera aligné à gauche.
- Problème : une ligne trop longue suivie d'une ligne trop courte forment un effet indésirable

- On appelle rivière une succession d'espaces sur plusieurs lignes consécutives.
- Comment éviter les rivières ?

Du paragraphe à la page

- Le découpage en pages peut être effectué de la même manière avec le même modèle colle/pénalité
- Comment éviter les veuves et les orphelins ? (ligne isolée en bas et haut de page)
- Éviter aussi les césures entre une page et la page suivante.

Du paragraphe à la page

- L'algorithme total-fit est inexploitable (pas assez de mémoire pour stocker tout un document)
- On imprime une page dès qu'elle est prête
- Utilisation d'un des deux premiers algorithmes.