

Activité 3

Peux-tu répéter? - Compression de texte

Résumé

Puisque les ordinateurs disposent d'un espace limité pour contenir les informations, ils doivent les représenter d'une manière aussi efficace que possible. C'est ce que l'on appelle la compression. En codant les données avant de les stocker, puis en les décodant lors de leur restitution, l'ordinateur peut en stocker davantage et les envoyer plus rapidement par Internet.

Liens pédagogiques

- ✓ Français : reconnaître des séquences dans les mots et les textes.
- ✓ Technologie : connaissance et compréhension technique.
Comment fonctionnent les ordinateurs?

Compétences

- ✓ Copier un texte écrit

Âge

- ✓ 9 ans et plus

Matériels

- ✓ Transparents préparés à partir du transparent destiné au professeur :
Peux-tu répéter? (page 25)

Chaque enfant a besoin de :

- ✓ L'exercice : Peux-tu répéter? (page 26)
- ✓ L'exercice : Pour les plus forts (page 27)
- ✓ L'exercice : Vite fait, bien fait! (page 28)
- ✓ L'exercice : Pour les champions des champions (page 29)

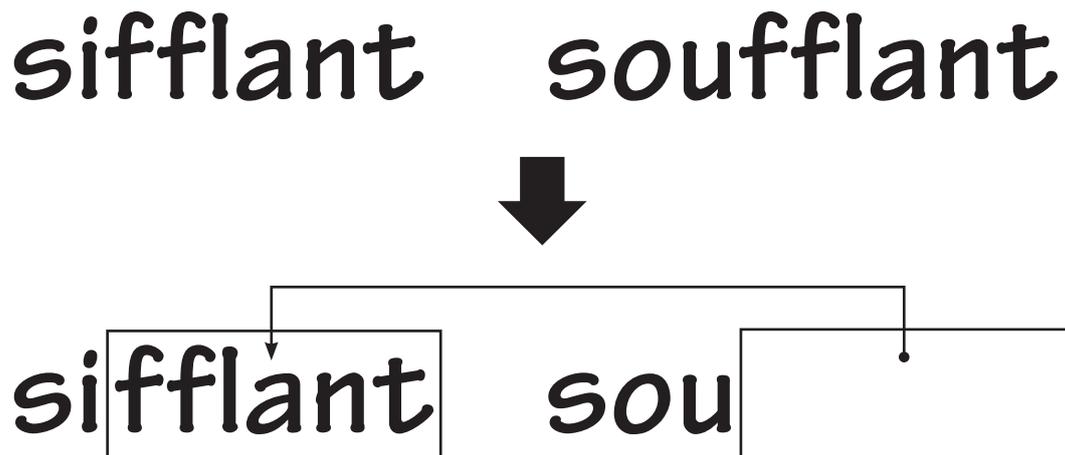
Peux-tu répéter ?

Introduction

Les ordinateurs stockent et transmettent de grandes quantités de données. Pour éviter d'utiliser trop d'espace de stockage ou mettre trop de temps à envoyer les informations par le modem, ils compressent le texte d'une manière similaire à celle présentée ci-dessous.

Présentation et discussion

Montrez le transparent « Vive le vent » (page 25). Cherchez les séquences de lettres répétées dans ce poème. Pouvez-vous trouver des groupes de 2 lettres ou plus qui se répètent, ou même des mots entiers ou des phrases ? (Remplacez ces groupes de lettres répétées par des cases, comme dans l'exemple ci-dessous).



Transparent du professeur :
Peux-tu répéter ?

Vive le vent

Vive le vent, vive le vent

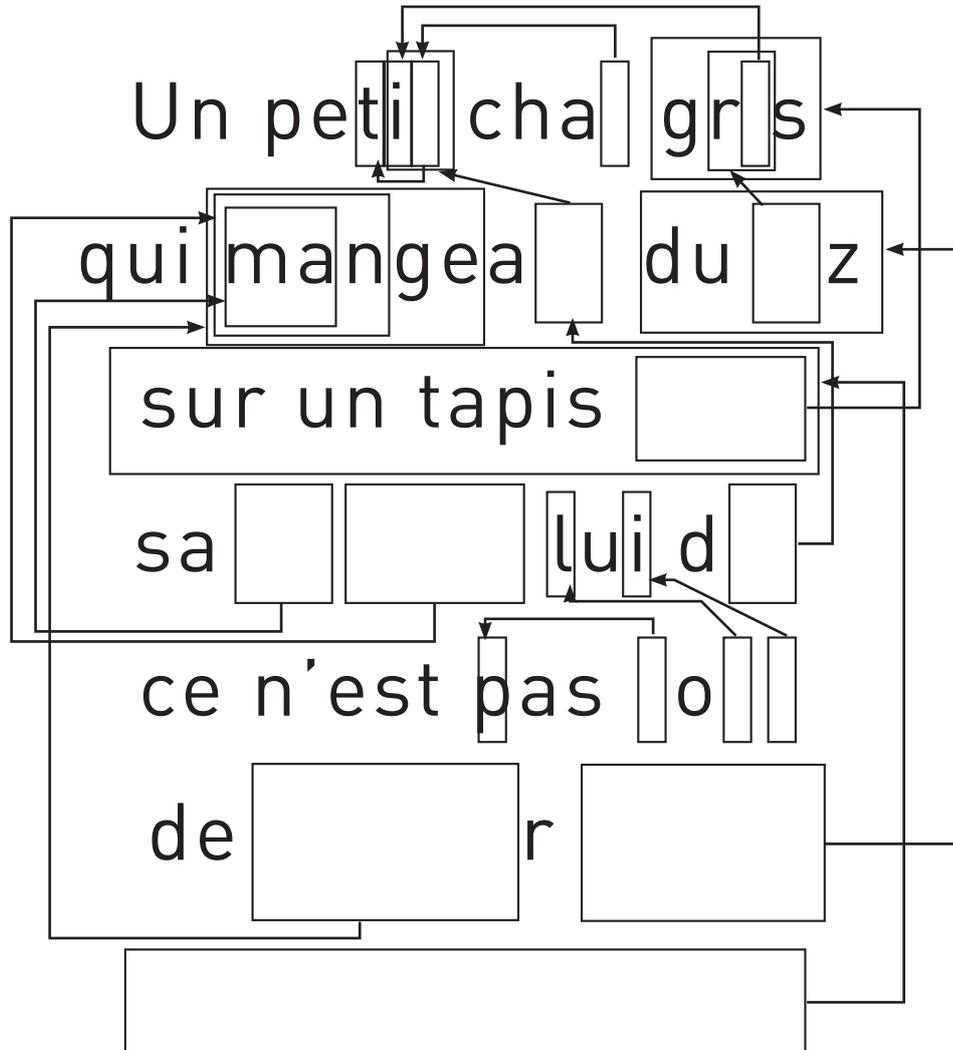
Vive le vent d'hiver

Qui s'en va sifflant
soufflant

Dans les grands
sapins verts

Exercice : Peux-tu répéter ?

Il manque de nombreuses lettres et de nombreux mots dans ce poème. Peux-tu remplir les blancs pour le compléter correctement ? Tu les trouveras dans les cases indiquées par la pointe de la flèche.



Choisis à présent un poème simple ou une comptine et crée ton propre schéma. Assure-toi que les flèches pointent toujours vers une partie du texte déjà lue. Ton poème doit, pour être décodé, être lu dans le sens de lecture habituel : de gauche à droite et du haut vers le bas.

Défi : Compte le peu de mots qui restent !

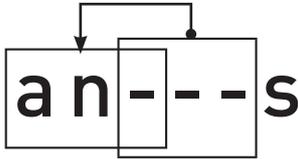
Voici quelques suggestions : Les comptines et chansons traditionnelles se prêtent bien à cet exercice, mais aussi les poèmes appris en classe, de Jacques Prévert, de Robert Desnos, ou d'autres auteurs !

Astuce : Évite de surcharger le poème avec trop de flèches. Espace suffisamment les lettres et les mots quand tu les écris pour garder de la place pour tracer des cases imbriquées et les flèches correspondantes.

Il est plus simple d'écrire d'abord le poème et de décider ensuite de l'emplacement des cases.

Exercice : Pour les plus forts

Comment résoudre cette énigme ?



Parfois, la partie de référence et la partie manquante ont des lettres en commun : elles se chevauchent. Dans ce cas, tu peux la déchiffrer correctement en copiant les lettres de gauche à droite. Alors chaque lettre visible peut être copiée avant de servir à déchiffrer la suite. C'est très utile pour l'ordinateur lorsqu'un même caractère (ou une même séquence) se répète plusieurs fois de suite.

Essaie de trouver des exemples.

Dans l'ordinateur, les cases et les flèches sont représentées par des nombres.
Par exemple :

ananas

peut s'écrire **an(2,3)s**. Le « 2 » signifie qu'il faut compter deux caractères vers la gauche pour trouver où l'on doit commencer à copier,

an - - - s

et le « 3 » signifie qu'il faut copier trois caractères consécutifs :

ana - - s

anan - s

ananas

Comme il faut deux nombres pour coder les mots, il est préférable de coder seulement des groupes d'au moins deux lettres pour arriver à une compression, sinon, il n'y a pas d'économie d'espace. En fait, la taille du fichier pourrait même augmenter si deux nombres sont utilisés pour coder une seule lettre.

Trouve des exemples de mots et écris-les comme le ferait un ordinateur s'ils étaient compressés. Demande à tes amis de les décoder !

Exercice : Vite fait, bien fait !

De combien de mots as-tu vraiment besoin ?

Imagine que tu es un ordinateur et que tu dois stocker autant d'informations que possible sur ton disque. Raye tous les groupes d'au moins deux lettres qui se répètent. Ces groupes ne sont plus nécessaires puisqu'on peut les remplacer par des flèches. Le but est d'obtenir le plus grand nombre possible de lettres rayées.

L'escargot

Un escargot s'en allait à la foire
Pour s'acheter une paire de souliers,
Quand il arriva, il faisait déjà nuit noire
Il s'en retourna nu pieds.

Un escargot s'en allait à l'école
Car il voulait apprendre à chanter
Quand il arriva, ne vit que des herbes folles
C'était les vacances d'été.

Un escargot s'en allait en vacances
Pour visiter l'Inde et le Japon
Au bout de sept ans, il était toujours en France
Entre Dijon et... Lyon !

Exercice : Pour les champions

Prêt pour une compression très complexe ?

L'histoire suivante a été analysée par un programme informatique qui a trouvé plus de 2000 lettres pouvant être rayées parmi les 2 500 environ composant ce texte. Combien peux-tu en trouver ? Souviens-toi, seuls des groupes répétés d'au moins deux lettres peuvent être éliminés. Bonne chance !

Il était une fois trois petits cochons qui vivaient avec leur maman dans une petite chaumière. Un jour, la maman appela ses trois fils et leur dit : « Je voudrais que vous partiez d'ici et construisiez chacun votre maison, mais prenez garde qu'elle soit bien solide pour que le grand méchant loup ne puisse pas entrer et vous manger. » La maman embrassa ses trois petits cochons et ils s'en allèrent construire leurs maisons.

Le premier petit cochon rencontra un homme portant une botte de paille. « Puis-je avoir un peu de paille pour construire ma maison ? » demanda le petit cochon. Et l'homme lui donna de la paille.

Le deuxième petit cochon rencontra un homme qui portait un chargement de bois. « Puis-je avoir quelques bouts de bois pour construire ma maison ? » demanda le petit cochon. Et l'homme lui donna du bois.

Le troisième petit cochon, lui, rencontra un homme chargé de briques. « S'il vous plaît, Monsieur, demanda le troisième petit cochon, puis-je avoir quelques briques pour construire ma maison ? » L'homme lui donna assez de briques pour bâtir une grande et solide maison, avec une cheminée près de laquelle on pouvait s'asseoir.

Les trois petits cochons rentraient joyeusement chez eux quand le grand méchant loup les aperçut.

« Comme ils doivent être tendres ! Lequel vais-je manger en premier ? Je vais commencer par le petit cochon dans la maison de paille ! » Il frappa à la porte.

Petit cochon, gentil petit cochon, laisse-moi entrer !

Non, Non ! Par le poil de mon petit menton !

Alors, je vais souffler et ta maison s'envolera !

Le loup gonfla ses joues, souffla, souffla de toutes ses forces, et la maison de paille s'envola.

Au secours ! cria le premier petit cochon en courant vers la maison de bois de son frère.

À peine celui-ci eut-il refermé la porte que le loup frappa.

Petits cochons, gentils petits cochons, laissez-moi entrer !

Non, non ! Par le poil de nos petits mentons ! répondirent les deux frères.

Alors, je vais souffler, souffler, et votre maison s'envolera !

Le loup se gonfla les joues, souffla, souffla de toutes ses forces, et la maison de bois s'envola.

Au secours ! crièrent les deux petits cochons en courant aussi vite que possible vers la maison de briques de leur frère.

Bientôt, la voix du loup résonna.

« Petits cochons, gentils petits cochons, je peux entrer ? »

Non ! non ! Par le poil de nos petits mentons !

Alors, vous allez voir, hurla le loup, je vais souffler sur votre maison, et je vais la démolir !

Il prit alors sa plus profonde respiration et souffla comme un fou.

Mais cette fois-ci, il ne réussit pas à mettre la maison par terre. Cela le rendit fou furieux. « Il faut absolument que j'attrape ces cochons » se dit-il, et il décida de passer par la cheminée.

Pendant ce temps, le troisième petit cochon, qui était très rusé, alluma un grand feu dans la cheminée et y posa un chaudron rempli d'eau. Quand le loup descendit dans la cheminée, il tomba tout droit dedans. Il poussa un hurlement qu'on entendit à des kilomètres à la ronde et repartit comme il était venu, par la cheminée. On n'entendit plus jamais parler de lui.

Ce qu'il faut retenir

La capacité de stockage des ordinateurs augmente à un rythme incroyable (environ un million de fois sur un ordinateur classique au cours des 25 dernières années), mais nous trouvons toujours plus à y stocker. Ils peuvent contenir des livres entiers et même des bibliothèques, et maintenant de la musique et des vidéos, à condition d'avoir suffisamment de place. Des fichiers volumineux posent également un problème pour leur transmission par Internet car ils sont plus longs à télécharger. Nous essayons aussi de construire des ordinateurs toujours plus petits : il est même possible de stocker beaucoup d'informations sur des téléphones mobiles ou des montres-ordinateurs !

Cependant, il existe des solutions à ces problèmes. Plutôt que d'acheter plus d'espace de stockage ou un modem plus rapide, nous pouvons *compresser* les données de manière à ce qu'elles occupent moins d'espace. L'ordinateur effectue en général automatiquement ce procédé de compression et de décompression des données. Nous voyons simplement que le disque stocke plus de choses ou que les pages Internet s'affichent plus vite mais en fait, c'est l'ordinateur qui travaille plus.

De nombreuses méthodes de compression ont été imaginées. La méthode utilisée pour cette activité, basée sur le principe des flèches dirigées vers des occurrences de portions de texte, est souvent appelée « algorithme de Ziv-Lempel » ou « algorithme LZ », car il a été découvert par deux professeurs israéliens dans les années 1970. Cet algorithme s'applique à tout langage et peut facilement diminuer de moitié la taille des données compressées. Il est parfois appelé « zip » sur les PC et sert aussi pour les images « gif » et les modems haut débit. Dans le cas des modems, l'algorithme réduit la quantité de données transmises par la ligne téléphonique et rend la transmission beaucoup plus rapide.

Selon d'autres méthodes, les lettres utilisées le plus souvent sont représentées par un code plus court que les autres. C'est le principe du morse.

Solutions et astuces

Peux-tu répéter ? (page 26)

Un petit chat gris
qui mangeait du riz
sur un tapis gris
sa maman lui dit
ce n'est pas poli
de manger du riz
sur un tapis gris.