

Activity 3

Вы можете это повторить! –Сжатие текста

Краткое содержание

Так как компьютеры имеют ограниченные возможности в хранении информации, то необходимо представить информацию настолько эффективно, насколько это допустимо. Этот процесс называют сжатием. Кодируя данные до их сохранения, и декодируя после их восстановления, компьютер может хранить больше данных, или отправлять их быстрее посредством Интернета.

Учебные направления

Английский язык: Распознавание закономерности в словах и в тексте.

Технология: Технологические знания и понимания. Как работают компьютеры.

Умения и навыки

Повторение написанного текста.

Возраст

От 9 и старше

Материалы

ОНР плакат, сделанный из пленки ОНР, для Учителя: Вы сможете это повторить! (стр. 3)

Каждому ребёнку необходимо:

Лист деятельности: Вы сможете это повторить! (стр. 4)

Лист деятельности: Для любознательных (стр. 5)

Лист деятельности: Коротко и ясно (стр. 6)

Лист деятельности: Для очень любознательных (стр. 7)

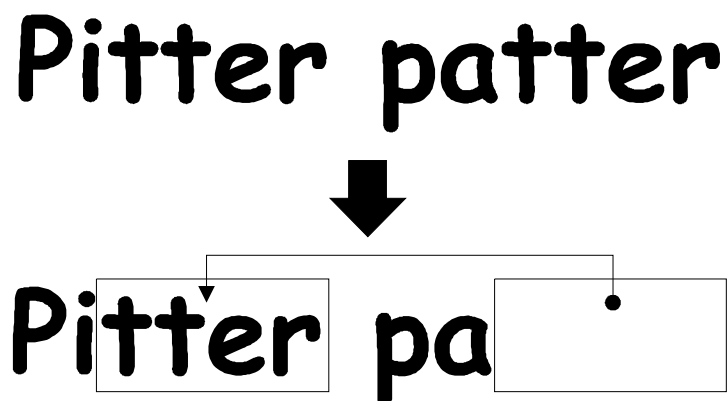
Вы сможете это повторить!

Указания

Компьютеры должны хранить и передавать большой объем данных. Но так как они не могут использовать для хранения данных или для передачи больших массивов информации через сети связи, неограниченное количество ячеек памяти, то они сжимают данные и информацию, и выглядит это следующим образом.

Демонстрация и обсуждение

Покажите «Дождь» ОНР (стр. 3). Найдите одинаковые словосочетания (шаблоны) в этом стихотворении. Сможете ли вы найти группы из 2 или более символов, которые повторяются, а может быть целые слова или фразы? (Замените их полями как показано ниже на диаграмме).



The Rain

Pitter patter

Pitter patter

Listen to the rain

Pitter patter

Pitter patter

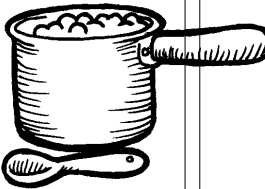
On the window pane

Лист деятельности: Вы сможете это повторить!

Много слов и символов отсутствуют в этом стихотворении. Сможете ли вы заполнить недостающие символы и слова, чтобы получить правильные предложения? Вы найдете их в полях, на которые указывает стрелка.

Peas porridge hot, cold, in the pot,

 Nine days a week,
 Some like it hot, _____,
 _____,
 _____,
 _____.



Сейчас вы выберете простое стихотворение или ритмичную скороговорку и спроектируете свою задачу. Удостоверьтесь, что все ваши стрелки указывают на начальную часть текста. Ваше стихотворение должно быть декодировано слева направо, и сверху донизу, таким образом, как мы всегда читаем текст.

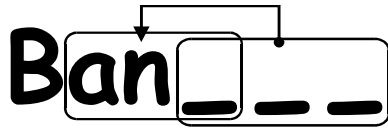
Сложная задача: Обратите внимание, что некоторые оригинальные слова необходимо сохранить!

Вот некоторые из них: Three Blind Mice (Три самоуверенных мышонка), Mary Mary Quite Contrary (Мэри Мэри Абсолютное Упрямство), Hickory Dickory Dock (Хикори Дикори Док) — или другие герои книг Dr. Seuss (др. Сеусса)!

Подсказка: Попытайтесь избежать нагромождения стрелок. Оставьте больше пространства около предложений и слов, для указания полей и записи в них шаблонов. Проще проектировать задачу, если вы вначале записываете стихотворение, а затем решаете, где у вас будет поле.

Лист деятельности: Для любознательных

Как вы будете решать эту задачку?



Иногда недостающий текст указывает на часть себя. В этом случае, чтобы декодировать правильно, необходимо все символы повторить слева направо. Тогда каждый символ можно будет повторить прежде, чем он будет востребован. Это используется в компьютерах, когда он кодирует длинный специфический символ или шаблон.

Попробуйте придумать свой пример.

В компьютерах поля и стрелки представляются числами. Например,

Banana

можно записать, как **Ban(2,3)**. "2" указывает, что необходимо читать с символа, записанного на два символа раньше, чтобы найти точку начала для повторения,

Ban---

и "3" указывает на необходимость повторения трёх последовательных символа:

Bana--

Banana-

Banana



Составьте некоторые слова, закодированные таким же способом, для компьютера, как если бы они были сжаты им. Ваши друзья смогут их декодировать?

Лист деятельности: Коротко и ясно

Как много слов вам необходимо в этом примере? Представьте себе, что вы - компьютер, пытающийся вместить на свой диск столько информации, сколько это возможно. Вычеркните все группы двух и более символов, которые уже встречались. Здесь, они такие длинные, уже не нужны, так как могут быть

I know an old lady who swallowed a bird
How absurd! She swallowed a bird!
She swallowed the bird to catch the
spider
That wriggled and jiggled
and tickled inside her
She swallowed the spider to catch the fly
I don't know why she swallowed a fly
Perhaps she'll die...

заменены указателем. Ваша цель состоит в том, чтобы получить как можно больше вычеркнутых символов, настолько, насколько это возможно.

Лист деятельности: Для очень любознательных

Готовы к варианту, действительно жёсткого сжатия?

Нижеприведенный текст, с помощью компьютерной программы был сжат, и вычеркнуты, по крайней мере, 1 633 символа. А сколько вы можете найти? Помните, только группы двух или большего количества повторяемых символов могут быть выброшены. Удачи!

Once upon a time, long, long ago, three little pigs set out to make their fortunes. The first little pig wasn't very clever, and decided to build his house out of straw, because it was cheap. The second little pig wasn't very clever either, and decided to build his house out of sticks, for the "natural" look that was so very much in fashion, even in those days. The third little pig was much smarter than his two brothers, and bought a load of bricks in a nearby town, with which to construct a sturdy but comfortable country home.

Not long after his housewarming party, the first little pig was curled up in a chair reading a book, when there came a knock at the door. It was the big bad wolf, naturally.

"Little pig, little pig, let me come in!" cried the wolf.

"Not by the hair on my chinny-chin-chin!" squealed the first little pig.

"Then I'll huff, and I'll puff, and I'll blow your house down!" roared the wolf, and he *did* huff, and he *did* puff, and the house soon collapsed. The first little pig ran as fast as he could to the house of sticks, and was soon safe inside. But it wasn't long before the wolf came calling again.

"Little pig, little pig, let me come in!" cried the wolf.

"Not by the hair on my chinny-chin-chin!" squealed the second little pig.

"Then I'll huff, and I'll puff, and I'll blow your house down!" roared the wolf, and he *did* huff, and he *did* puff, and the house was soon so much firewood. The two terrified little pigs ran all the way to their brother's brick house, but the wolf was hot on their heels, and soon he was on the doorstep.

"Little pig, little pig, let me come in!" cried the wolf.

"Not by the hair on my chinny-chin-chin!" squealed the third little pig.

"Then I'll huff, and I'll puff, and I'll blow your house down!" roared the wolf, and he huffed, and he puffed, and he huffed some more, but of course, the house was built of brick, and the wolf was soon out of breath. Then he had an idea. The chimney! He clambered up a handy oak tree onto the roof, only to find that there *was* no chimney, because the third little pig, being conscious of the environment, had installed electric heating. In his frustration, the wolf slipped and fell off the roof, breaking his left leg, and severely injuring his pride. As he limped away, the pigs laughed, and remarked how much more sensible it was to live in the city, where the only wolves were in the zoo. And so that is what they did, and of course they all lived happily ever after.

Как это понимать?

Емкость запоминающего устройства компьютеров растет с невероятной скоростью – за 25 лет, объём памяти на типовом компьютере вырос в миллион раз, но мы всё равно хотим хранить все больше и больше информации. Компьютеры могут хранить целые книги или даже библиотеки, и теперь и музыку, и кинофильмы также, если достаточно памяти для этого. Большие файлы – проблема Интернета, так как им необходимо много времени для загрузки. Мы также пытаемся сделать компьютеры меньших размеров – даже в сотовых или наручных часах, вы ожидаете, что такие компьютеры будут хранить много информации! Но решение этой проблемы есть, как бы, то ни было. Вместо того, чтобы покупать больше пространства памяти, или более быстрое сетевое подключение, мы можем *сжать* данные так, чтобы они заняли меньше места. Этот процесс упаковки и распаковки данных, компьютером производится автоматически. Всё, на что мы могли бы обратить внимание – то, что пространство диска вмещает больше данных, или, что web-страница отображает быстрее, т.е. компьютер фактически больше обрабатывает информации. Были изобретены различные методы сжатия. Метод, используемый здесь нами, основанный на принципе, чтобы указывать на более раннее вхождение участков памяти текста, часто упоминается как ‘Ziv-Lempel coding,’ или ‘LZ coding’, изобретенный двумя израильскими профессорами в 1970-ых. Этот метод может использоваться для любого языка и может легко быть разделен на два размера сжимаемых данных. Этот метод иногда упоминается на персональных компьютерах как ‘zip’, а также используется для изображений формата ‘GIF’ (формат обмена графическими данными), также и в высокоскоростных модемах. В случаи модемов, этот метод уменьшает количество данных, которые должны быть переданы по телефонной линии, тем самым процесс идет быстрее. Один из методов, основан на идее, что у символов, которые используются чаще, должны быть более короткие коды. Азбука Морзе использует эту идею.

Решения и подсказки

Вы сможете это повторить! (стр. 4)

**Pease porridge hot,
Pease porridge cold,
Pease porridge in the pot,
Nine days old.**

**Some like it hot,
Some like it cold,
Some like it in the pot,
Nine days old.**