Please keep only these 🡺 portions of the document with your answers.

🡺 Your name: Student No.: UserID: ­­­­\_\_\_\_\_ @mySeneca.ca

**K****indly delete the instructions and notes.** **−10% cost if you make me wade through all the stuff you were asked to delete.**

**Part 1: Compression (40 points)**

old quote from Vangie Beal, managing editor of Webopedia; text contains 449 characters:

*Data compression is particularly useful in communications because it enables devices to transmit or store the same amount of data in fewer bits. There are a variety of data compression techniques, but only a few have been standardized. The CCITT has defined a standard data compression technique for transmitting and a compression standard for data communications through modems. In addition, there are file compression formats, such as ARC and ZIP.*

Compression algorithms build a token/string dictionary and substitute repeating strings with tokens:

*♠compression   
♣here are   
♦communications   
♥data   
☺standard   
☼transmit   
☯technique*

*♥♠is particularly useful in ♦because it enables devices to ☼ or store the same amount of ♥in fewer bits. T♣a variety of ♥♠☯s, but only a few have been ☺ized. The CCITT has defined a ☺ ♥♠☯ for ☼ting faxes and a ♠☺ for ♥♦through modems. In addition, t♣ file ♠formats, such as ARC and ZIP.*

Including dictionary, total size is 368 characters or 82% of original. Compression increases with length of text, i.e. more pattern matches for each dictionary item.

Explore file compression using the [ substitution token to original character string ] dictionary ideas above. The total text of this nursery rhyme is 187 characters including a trailing space after each word:

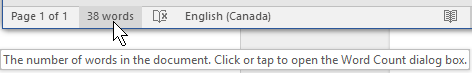
**the itsy bitsy spider crawled up the water spout**

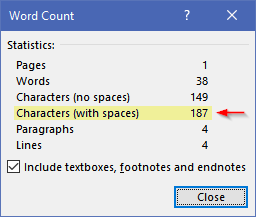
**down came the rain and washed the spider out**

**out came the sun and dried up all the rain**

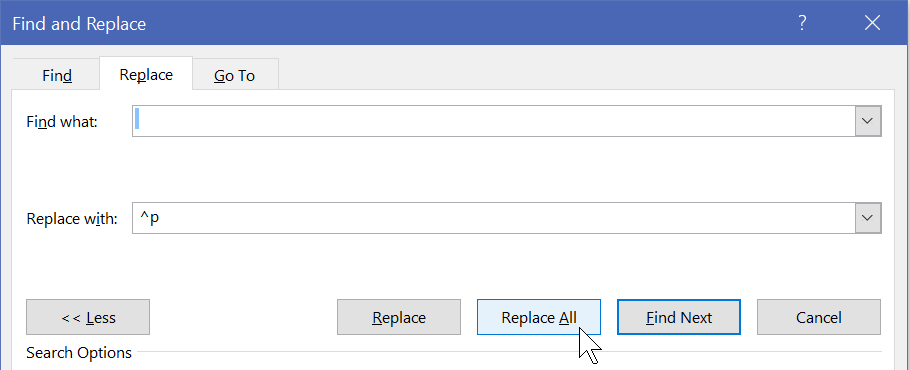
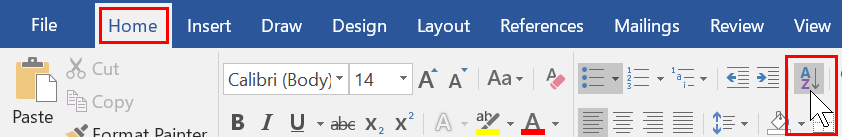
**and the itsy bitsy spider went up the spout again**

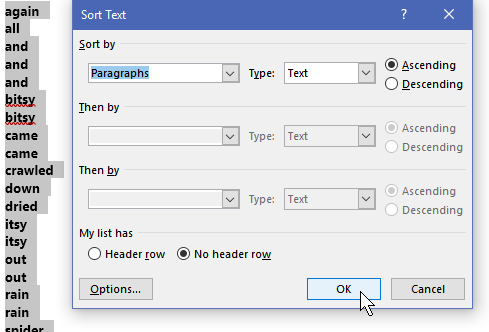
Copy the four lines of the above rhyme to a new MS-Word document (Ctrl-N).  
In the bottom left of the Word display, click “38 words”.  
It should have 187 characters with spaces.





This may help with your substitution analysis:

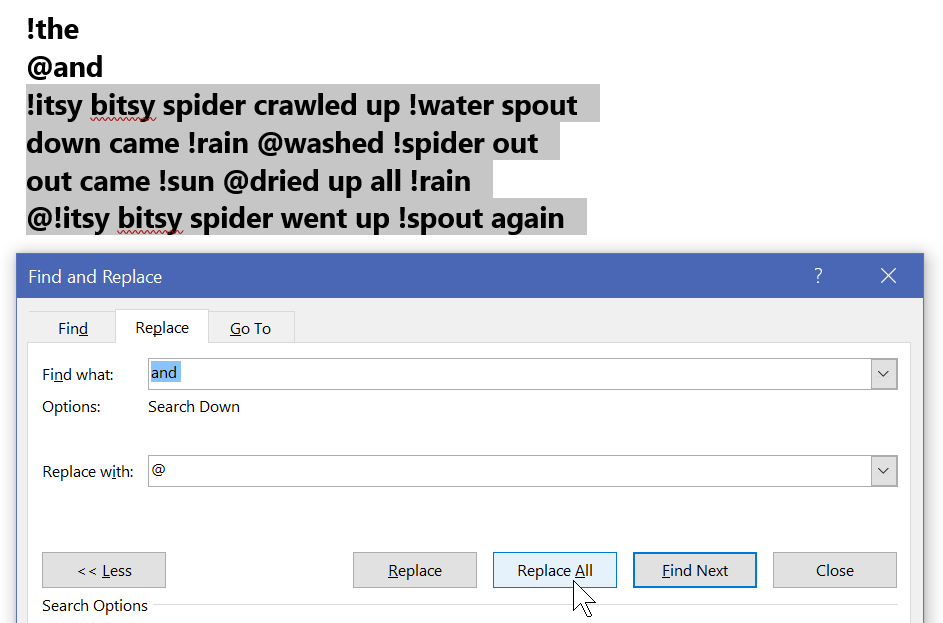
* copy the rhyme to another new document (Ctrl-N) used only for investigation
* Find and Replace a **space** with a **space** **+ paragraph marker** **^p** (Ctrl+H)  
  Find what: □ Replace with: □^p   
  
* Then sort the lines to see repeating words. (Alt H S O)  
  

Sort by Paragraphs resulting in the rhyme's words, one per line, in alphabetical order;   
this makes it easy to see repeating words:

Anything occurring only once is not worth substituting with a token and including in the dictionary; you will just be adding two characters (the tokens) to the file. Any string with a length of 2 or 3 and occurring only twice is similarly not worth it.

Here is an example of saving many characters by replacing repeating words **including their trailing spaces** ("the ", "and ") with tokens (!,@). Remember:

* a space is a character that must be preserved
* the compression dictionary with token/string must be included in order to decompress the text back to its original.



**FYI, this is the formula for characters saved:**  
For the ***length*** of any string occurring ***n*** times and replaced by a single character token:

Saving ***n*** occurrences \* ***length*** costs the dictionary entry (token char + string ***length***) plus the ***n*** token placeholders in the data stream.

e.g. "the " occurs 8 times with a length of 4 including the trailing space (32 characters)  
less the compression overhead of a 5 char dictionary entry + 8 tokens (13 characters)  
is 32 original less 13 overhead = 19 characters saved.

**How much more can you compress the text?**

* Use digits and the special characters on the shifted top row for tokens: **1234567890!@#$%&\*()\_+** Note: do not use **^**, it is a Microsoft escape char.
* A separator in the *token to string* dictionary will inflate your character count. A decompression algorithm would read the first character as the substitution token and the second char to end-of-line as the original character string. Do keep the trailing blank with the original word – it is part of the replaced string.   
  Do this: **\*the**  not this \***=**the
* Copy and paste the following into your answer document for 10 points each.

🡺 your dictionary of compression token to string characters, one entry per line.

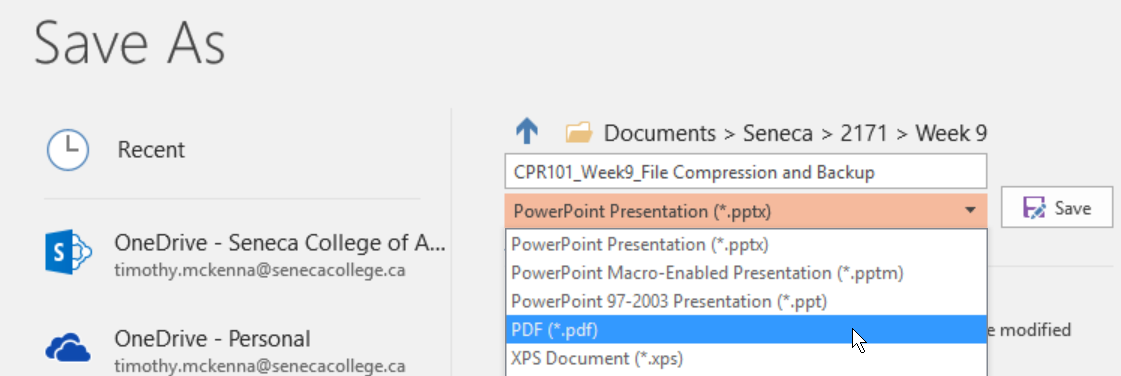
🡺 the compressed rhyme with the token substitutions,   
 i.e. the compressed text (not the sorted analysis list of words)

🡺 how many characters are in the dictionary + compressed text and what is it as percentage of the original’s 187?

🡺 **Now test your compression dictionary.** Reverse the process to see if your compression dictionary is accurate. Process dictionary items from the bottom up: find the compression character in the compressed data and replace it with the original string. **Paste the decompressed version below** – *even if it is not perfect*. **What modifications, if any, does the compression dictionary need to return the compressed data back into its original state?**

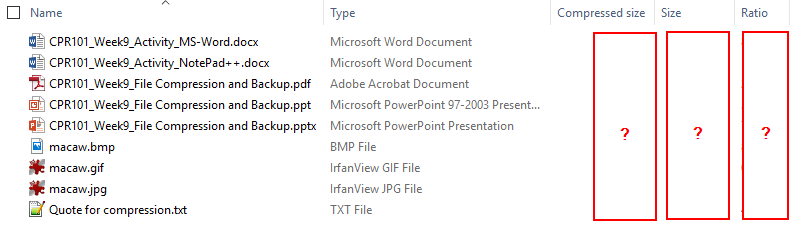
**Part 2:**

* If it is not already there, extract / decompress the files in this week’s activity .zip archive to your Desktop.
  + Remember that compressed files must be decompressed before they can be used. Windows does this automatically into the %temp% folder if you open a file directly from a .zip archive.
* Download this week’s PowerPoint slides and save to your Desktop folder
  + open it, File menu, Save As, PDF (\*.pdf)
  + File menu, Save As, PowerPoint 97-2003 Presentation (\*.ppt)  
    *If you see the Microsoft PowerPoint Compatibility Checker, click Continue.*



Compress all files from your Desktop folder into a zip archive:

* select the files then right click and use the   
  S*end to* > *Compressed (zipped) folder* option or use 7zip.

Open the .zip archive with Windows Explorer.   
Use the Snipping Tool ( + “snip”) to copy only the information seen below.

The compression ratio is (1 – compressed / original size) showing the space saved by the zipping, i.e. compressing, the original files into the archive.

FYI: opening the .zip archive with 7zip will show exact bytes for original Size and Packed (compressed) size. 7zip does not show the compression ratio.

See <https://www.noupe.com/design/everything-you-need-to-know-about-image-compression.html>

**🡺** Paste the image of the Windows Explorer .zip archive information.

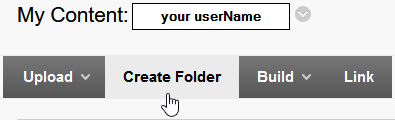
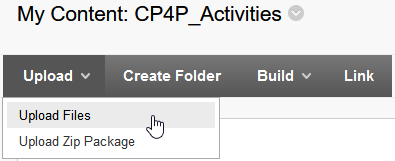
Use the Snipping Tool ( + “snip”) to copy only the information seen above.

🡺 Files with the **lowest** ratios were compressed the **least**. Ratio indicates % of space saved.  
Which file types compressed the least? Why would that be? (10 pts)

🡺 Files with the **highest** ratios were compressed the **most**.   
Which file types compressed the most? Why would that be? (10 pts)

**Part 3: Backup**

In order to preserve important information from accidental deletion, make a copy in a geographically separate location on an independent platform. One option would be to store your data, somewhere on the college’s network such as Office 365's OneDrive (as long you have not set up synchronization). Blackboard's Content Area is also an option. (If Blackboard does not work, contact ITS; in the meantime, use OneDrive.)

* 1. Login to Blackboard.
  2. Click on the **Content System** tab near the top right
  3. From the menu under “My Content”, click on Create Folder  
     
  4. Name the folder **CP4P\_Activities** and click Submit
  5. Click on the new folder to open it up  
     
  6. Select Upload, then select Upload Files from the menu  
     
  7. Browse your CP4P files and upload something to the folder, e.g. the zip archive you created in Part 2…being compressed and a single file upload, it would take the minimum amount of time.   
       
     Congratulations. You just backed up something to the “cloud”.

🡺 paste a screen shot of your backup results. (use the Screen Snip tool) **(10 points)**

**Imagine your laptop just stopped working and could not be restarted**after you completed a great many hours of work today and yesterday.   
You need a backup & restore strategy. **(30 points)**

🡺 What is (or what should have been) your backup routine? How do you ensure your backup is current?

🡺 How does your backup routine address the three characteristics of a real backup and fulfill the 3-2-1 backup check?

🡺 Now that you have a backup *but no laptop*, how will you access and work with the current version of your backed up files? What is your restore/recovery strategy?

🡺 How long would this all take…and what if you a had a big assignment due tomorrow?

<https://www.google.com/search?q=7zip+full+and+differential+backup+script>