```
# FILE: text_clean.py
# FINAL PROJECT
# Professor Ahmadnia
# Group: Kevin Vuong, Anika Corpus, Christopher Grant
# Description: This program provides functionality for removing comments and formatting the
spaces.

import re


def comment_remover(file_read, file_write):
    """
    Removes the comments from the text file (file_read) and writes it to another file (
file_write).

    :param file_read: The text file being read from
    :param file_write: The text file being written into
    :return: None
    """
    content = file_read.readlines()
    content = ''.join(content)

    # The pattern to remove multi-line comments
    mult_line_comment_pattern = '//.*\n.*//'
    content = re.sub(mult_line_comment_pattern, '', content, 0)

    # The pattern to remove single-line comments
    single_line_comment_pattern = '//.*//'
    content = re.sub(single_line_comment_pattern, '', content, 0)

    file_write.writelines(content)


def space_formatter(expr):
    """
    Formats the spaces of a line of text

    Arguments:
        expr: a line of text

    Returns:
        line: the line of processed text

    """
    token_list = expr.split(" ")
    content = ''

    # This loop removes elements with empty string contents
    for token in token_list:
        if re.match(r'^\s*\n*\s*$', token):
            continue
        content = content + token.strip()

    # This section adds the appropriate spaces for the reserved words
    reserved_pattern = r'(\s*PROGRAM\s*|\s*INTEGER\s*|\s*PRINT\s*|\s*BEGIN\s*|\s*END\.\s*)'
    matched = re.match(reserved_pattern, content)
    word = ''
    if matched is not None:
        word = matched.group()

    if word == 'PROGRAM':
        content = re.sub(r'(\s*PROGRAM\s*)', 'PROGRAM ', content, 0)

    if word == 'INTEGER':
        content = re.sub(r'(\s*INTEGER\s*)', 'INTEGER ', content, 0)

    if word == 'PRINT':
        content = re.sub(r'(\s*PRINT\s*)', 'PRINT ', content, 0)

    if word == 'BEGIN':
```

```
            content = re.sub(r'\s*BEGIN\s*', 'BEGIN', content, 0)

        if word == 'END.':
            content = re.sub(r'\s*END.\s*', 'END.', content, 0)

        # This section adds the appropriate spaces for the symbols
        symbolic_pattern = r'(\=|\*|\-|\,|\:|\(|\)|\<\=|\+|\;)'
        matched = re.findall(symbolic_pattern, content)

        for word in matched:
            if word == '=':
                content = re.sub(r'\s*=\s*', ' = ', content, 0)

            if word == ',':
                content = re.sub(r'\s*,\s*', ' , ', content, 0)

            if word == ';':
                content = re.sub(r'\s*;\s*', ' ;', content, 0)

            if word == '(':
                content = re.sub(r'\s*\(\s*', ' ( ', content, 0)

            if word == ')':
                content = re.sub(r'\s*\)\s*', ' ) ', content, 0)

            if word == '+':
                content = re.sub(r'\s*\+\s*', ' + ', content, 0)

            if word == '-':
                content = re.sub(r'\s*-\s*', ' - ', content, 0)

            if word == '*':
                content = re.sub(r'\s*\*\s*', ' * ', content, 0)

            if word == ':':
                content = re.sub(r'\s*:\s*', ' : ', content, 0)

        return content+'\n'


    def clean_text(filename):
        """
        Cleans up the spaces in the text file.

        :rtype: lines_content: The entire content of the string cleaned up
        :param filename: The name of the file you want to clean
        """
        file = open('finalv2.txt', mode='r+', encoding='utf-8')
        lines_read = file.readlines()
        lines_content = ''

        for line in lines_read:

            # Ignore lines that only contain the newline character
            if re.match(pattern=r'\s*\n\s*', string=line):
                continue

            line = space_formatter(line)
            lines_content = lines_content + line

        # Writes the cleaned up text to the text file
        with open(filename, mode='w+') as new_file:
            new_file.writelines(lines_content.strip())

        return lines_content.strip()


    def main():

        # First, remove all the comments
```

```
    with open("finalv1.txt") as source_file:
        new_file = open('finalv2.txt', mode='w+', encoding='utf-8')
        comment_remover(source_file, new_file)
        new_file.close()

    # Clean the spaces
    content = clean_text('finalv2.txt')
    print(content)


if __name__ == "__main__":
    main()
```