compare_files

June 3, 2023

1 Magic command to compare files

Some ways to display differences between files.

[1]:
```
from jyquickhelper import add_notebook_menu
add_notebook_menu()
```

[1]:
```
<IPython.core.display.HTML object>
```

1.1 Two functions slighly different

[2]:
```
f1 = '''
def edit_distance_string(s1, s2):
    """
    Computes the edit distance between strings *s1* and *s2*.

    :param s1: first string
    :param s2: second string
    :return: dist, list of tuples of aligned characters
    """
    n1 = len(s1) + 1
    n2 = len(s2) + 1
    dist = numpy.full((n1, n2), n1 * n2, dtype=numpy.float64)
    pred = numpy.full(dist.shape, 0, dtype=numpy.int32)

    for j in range(1, n2):
        dist[0, j] = j
        pred[0, j] = 2
    for i in range(0, n1):
        dist[i, 0] = i
        pred[i, 0] = 1
    pred[0, 0] = -1

    for j in range(1, n2):
        for i in range(1, n1):
            c = dist[i, j]

            p = 0
            if dist[i - 1, j] + 1 < c:
                c = dist[i - 1, j] + 1
                p = 1
            if dist[i, j - 1] + 1 < c:
                c = dist[i, j - 1] + 1
                p = 2
            d = 0 if s1[i - 1] == s2[j - 1] else 1
            if dist[i - 1, j - 1] + d < c:
                c = dist[i - 1, j - 1] + d
                p = 3
            if p == 0:
                raise RuntimeError(
                    "Unexpected value for p=%d at position=%r." % (p, (i, j)))

            dist[i, j] = c
            pred[i, j] = p

    d = dist[len(s1), len(s2)]
    return d
'''
```

[3]:
```
f2 = '''
def edit_distance_string(s1, s2):
    """
    Computes the edit distance between strings *s1* and *s2*.

    :param s1: first string
    :param s2: second string
    :return: dist, list of tuples of aligned characters
    """
    n1 = len(s1) + 1
    n2 = len(s2) + 1
    dist = numpy.full((n1, n2), n1 * n2, dtype=numpy.float64)
    pred = numpy.full(dist.shape, 0, dtype=numpy.int32)

    for i in range(0, n1):
        dist[i, 0] = i
        pred[i, 0] = 1
    for j in range(1, n2):
        dist[0, j] = j
        pred[0, j] = 2
    pred[0, 0] = -1

    for i in range(1, n1):
        for j in range(1, n2):
            c = dist[i, j]

            p = 0
            if dist[i - 1, j] + 1 < c:
                c = dist[i - 1, j] + 1
                p = 1
            if dist[i, j - 1] + 1 < c:
                c = dist[i, j - 1] + 1
                p = 2
            d = 0 if s1[i - 1] == s2[j - 1] else 1
            if dist[i - 1, j - 1] + d < c:
                c = dist[i - 1, j - 1] + d
                p = 3
            if p == 0:
                raise RuntimeError(
                    "Unexpected value for p=%d at position=%r." % (p, (i, j)))

            dist[i, j] = c
            pred[i, j] = p

    d = dist[len(s1), len(s2)]
    equals = []
    i, j = len(s1), len(s2)
    p = pred[i, j]
    while p != -1:
        if p == 3:
            equals.append((i - 1, j - 1))
            i -= 1
            j -= 1
        elif p == 2:
            j -= 1
        elif p == 1:
            i -= 1
        else:
            raise RuntimeError(
                "Unexpected value for p=%d at position=%r." % (p, (i, j)))
        p = pred[i, j]
    return d, list(reversed(equals))
'''
```

1.2 Visual differences: codediff

[4]:
```
%load_ext pyquickhelper
```

[5]:
```
%%html
<style>
table td, table th, table tr {text-align:left !important; white-space: pre;}
</style>
```

```
<IPython.core.display.HTML object>
```

This is slow due to the edit distance computation. It could be improved by a C++ implementation.

[6]:
```
%codediff f1 f2 --verbose 1
```

```
100%|¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿
 ↪¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿
¿ ¿ ¿ ¿ ¿ ¿ ¿ | 47/47 [00:02<00:00, 23.05it/s]
```

[6]:
```
<IPython.core.display.HTML object>
```

[7]:
```
%codediff f1 f2 --verbose 1 --two 1
```

```
100%|¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿␣
 ↪¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿ ¿
¿ ¿ ¿ ¿ ¿ ¿ ¿ | 47/47 [00:02<00:00, 22.99it/s]
```

```
[7]: <IPython.core.display.HTML object>
```

1.3 strdiff

```
[8]: %strdiff f1 f2
```

```
[8]: <IPython.core.display.HTML object>
```

1.4 textdiff

```
[9]: %textdiff f1 f2
```

```
[9]: <IPython.core.display.Javascript object>
```

```
[10]:
```