In [1]:

```
import spacy
from random import random
from bisect import bisect
```

Lab 04 - Edit distance

Implementacja algorytmu wyznaczania odległości edycyjnej wraz z odtworzonymi krokami

In [2]:

```
def delta(a, b):
    return 0 if a == b else 1

def edit_distance(x, y, delta, show_answer = False):
    m = len(x)
    n = len(y)
    table = [[0]*(n+1) for _ in range(m+1)]

    for i in range(m+1):
        table[i][0] = i

    for i in range(n+1):
        table[0][i] = i

    for i in range(1, m+1):
        for j in range(1, n+1):
            table[i][j] = min(table[i-1][j] + 1, table[i][j-1]+1, table[i-1][j-1] + delta(x[i-1], y[j-1]))

    if show_answer:
        return([x + ' -> ' + y, '-'*(len(x) + len(y) + 4)] + get_solution(table, x, y, m, n))
    return table[m][n]

def get_solution(table, s1, s2, i, j):
    if table[i][j] == 0:
        return []

    if i == 0:
        step = s1 + ' -> ' + '|' + s2[j-1] + '|' + s1
        s1 = s2[j-1] + s1
        return [step] + get_solution(table, s1, s2, i, j-1)
    elif j == 0:
        step = s1 + '->' + s1[:-1] + '_'
        s1 = s1[:-1]
        return [step] + get_solution(table, s1, s2, i-1, j)
    else:
        if table[i][j] == table[i-1][j] + 1:
            step = s1 + ' -> ' + s1[:i-1] + '_' + s1[i:]
            s1 = s1[:i-1] + s1[i:]
            return [step] + get_solution(table, s1, s2, i-1, j)
        elif table[i][j] == table[i][j-1] + 1:
            step = s1 + ' -> ' + s1[:i] + '|' + s2[j-1] + '|' + s1[i:]
            s1 = s1[:i] + s2[j-1] + s1[i:]
            return [step] + get_solution(table, s1, s2, i, j-1)

        if s1[i-1] == s2[j-1]:
            return get_solution(table, s1, s2, i-1, j-1)

        step = s1 + ' -> ' + s1[:i-1] + '*' + s2[j-1] + '*' + s1[i:]
        s1 = s1[:i-1] + s2[j-1] + s1[i:]
        return  [step] + get_solution(table, s1, s2, i-1, j-1)
```

In [3]:

```
for line in edit_distance('los', 'kloc', delta, True):
    print(line)
```

```
los -> kloc
-----------
los -> lo*c*
loc -> |k|loc
```

In [4]:

```
for line in edit_distance('Łódź', 'Lodz', delta, True):
    print(line)
```

```
Łódź -> Lodz
------------
Łódź -> Łód*z*
Łódz -> Ł*o*dz
Łodz -> *L*odz
```

In [5]:

```
for line in edit_distance('kwintesencja', 'quintessence', delta, True):
    print(line)
```

```
kwintesencja -> quintessence
----------------------------
kwintesencja -> kwintesencj_
kwintesencj -> kwintesenc*e*
kwintesence -> kwintes|s|ence
kwintessence -> k*u*intessence
kuintessence -> *q*uintessence
```

In [6]:

```
for line in edit_distance('ATGAATCTTACCGCCTCG', 'ATGAGGCTCTGGCCCCTG', delta, True):
    print(line)
```

```
ATGAATCTTACCGCCTCG -> ATGAGGCTCTGGCCCCTG
----------------------------------------
ATGAATCTTACCGCCTCG -> ATGAATCTTACCGCCT_G
ATGAATCTTACCGCCTG -> ATGAATCTTACC_CCTG
ATGAATCTTACCCCTG -> ATGAATCTTA|G|CCCCTG
ATGAATCTTAGCCCCTG -> ATGAATCTT*G*GCCCCTG
ATGAATCTTGGCCCCTG -> ATGAATCT|C|TGGCCCCTG
ATGAATCTCTGGCCCCTG -> ATGAA*G*CTCTGGCCCCTG
ATGAAGCTCTGGCCCCTG -> ATGA*G*GCTCTGGCCCCTG
```

Implementacja algorytmu znajdowania maksymalnego wspólnego podciągu

In [7]:

```
def longest_common_sequence(x, y):
    ranges = [len(y)]
    y_letters = list(y)
    for i in range(len(x)):
        positions = [j for j, l in enumerate(y_letters) if l == x[i]]
        positions.reverse()
        for p in positions:
            k = bisect(ranges, p)
            if k == bisect(ranges, p - 1):
                if k < len(ranges) - 1:
                    ranges[k] = p
                else:
                    ranges[k:k] = [p]
    return len(ranges) - 1
```

In [8]:

```
print('Najdłuższy wspólny podciąg słów los i kloc ma długość: {}'
      .format(longest_common_sequence('los', 'kloc')))
```

```
Najdłuższy wspólny podciąg słów los i kloc ma długość: 2
```

In [9]:

```
print('Najdłuższy wspólny podciąg słów Łódź i Lodz ma długość: {}'
      .format(longest_common_sequence('Łódź', 'Lodz')))
```

```
Najdłuższy wspólny podciąg słów Łódź i Lodz ma długość: 1
```

In [10]:

```
print('Najdłuższy wspólny podciąg słów kwintesencja i quintessence ma długość: {}'
      .format(longest_common_sequence('kwintesencja', 'quintessence')))
```

```
Najdłuższy wspólny podciąg słów kwintesencja i quintessence ma długość: 8
```

In [11]:

```
print('Najdłuższy wspólny podciąg słów ATGAATCTTACCGCCTCG i ATGAGGCTCTGGCCCCTG ma długość: {}'
      .format(longest_common_sequence('ATGAATCTTACCGCCTCG', 'ATGAGGCTCTGGCCCCTG')))
```

```
Najdłuższy wspólny podciąg słów ATGAATCTTACCGCCTCG i ATGAGGCTCTGGCCCCTG ma długość: 13
```

Podział tekstu na tokeny i usunięcie 3% tokenów

In [12]:

```
def text_with_removed_tokens(doc : spacy.tokens.Doc, tokens_to_delete : float):
    reduced_text = []

    for token in doc:
        if random() > tokens_to_delete :
            reduced_text += [token.text_with_ws]
        else:
            cnt = token.text_with_ws.count('\n')
            for _ in range(cnt):
                reduced_text += ['\n']

    return reduced_text
```

In [13]:

```
with open('romeo-i-julia-700.txt', mode='r', encoding='utf-8') as file:
    data = file.read()

    nlp = spacy.load("pl_core_news_sm")
    doc = nlp(data)
    reduced_text1 = text_with_removed_tokens(doc, 0.03)
    reduced_text2 = text_with_removed_tokens(doc, 0.03)


    with open('variant1.txt', mode='w', encoding='utf-8') as save_file:
        for token_text in reduced_text1:
            save_file.write(token_text)

    with open('variant2.txt', mode='w', encoding='utf-8') as save_file:
        for token_text in reduced_text2:
            save_file.write(token_text)

    lcs = longest_common_sequence(reduced_text1, reduced_text2)

    print('Ilość tokenów na które został podzielony tekst: {}'.format(len(doc)))
    print('Ilość tokenów usuniętych z pierwszej wersji podziału: {}'.format(len(doc) - len(reduced_text1)))
    print('Ilość tokenów usuniętych z drugiej wersji podziału: {}'.format(len(doc) - len(reduced_text2)))
    print('Najdłuższy wspólny podciąg tokenów: {}'.format(lcs))
```

```
Ilość tokenów na które został podzielony tekst: 2699
Ilość tokenów usuniętych z pierwszej wersji podziału: 51
Ilość tokenów usuniętych z drugiej wersji podziału: 53
Najdłuższy wspólny podciąg tokenów: 2564
```

In [14]:

```
def show_diff(x, y):
    L = [[0 for _ in range(len(y) + 1)] for _ in range(len(x) + 1)]
    for i in range(1, len(x) + 1):
        for j in range(1, len(y) + 1):
            if x[i - 1] == y[j - 1]:
                L[i][j] = L[i - 1][j - 1] + 1
            else:
                L[i][j] = max(L[i - 1][j], L[i][j - 1])

    lines = []
    i = len(x) - 1
    j = len(y) - 1

    while i >= 0 and j >= 0:
        if x[i] == y[j]:
            i, j = i-1, j-1
        elif L[i][j-1] >= L[i-1][j]:
            lines.append(f"> |{j}| {y[j]}")
            j -= 1
        elif L[i][j-1] < L[i-1][j]:
            lines.append(f"< |{i}| {x[i]}")
            i -= 1

    while j >= 0:
        lines.append(f"> |{j}| {y[j]}")
        j -= 1

    while i >= 0:
        lines.append(f"< |{i}| {x[i]}")
        i -= 1
    lines.reverse()
    for line in lines:
        print(line)
```

In [15]:

```
with open('variant1.txt', mode='r', encoding='utf-8') as file:
    lines1 = file.read().splitlines()
with open('variant2.txt', mode='r', encoding='utf-8') as file:
    lines2 = file.read().splitlines()

show_diff(lines1, lines2)
```

```
< |2| Romeo Julia
> |2| Romeo i Julia
< |11|  * PARYS — młody Weroneńczyk szlachetnego rodu, krewny księcia
> |11|  * PARYS — młody Weroneńczyk szlachetnego rodukrewny księcia
< |12| * MONTEKI, KAPULET — naczelnicy dwóch domów nieprzyjaznych sobie
< |13|  * STARZEC — stryjeczny brat Kapuleta
< |14|  * ROMEO — syn Montekiego
> |12|  * MONTEKI, KAPULET — naczelnicy dwóch domów nieprzyjaznych sobie
> |13|  * STARZEC — stryjeczny brat Kapuleta
> |14|  * ROMEO — Montekiego
< |18|  * LAURENTY ojciec franciszkanin
> |18| * LAURENTY — ojciec franciszkanin
< |19|  * JAN — brat z tegoż zgromadzenia
< |20|  * BALTAZAR — służący Romea
> |19|  * JAN — brat z tegoż zgromadzenia
> |20|  * BALTAZAR — Romea
< |25|  * PAŹ PARYSA
> |25| * PAŹ PARYSA
< |37| Rzecz odbywa się przez większą część sztuki w Weronie, przez część piątego aktu Mantui.
> |37| Rzecz odbywa się przez większą część sztuki w Weronie, przez część piątego w Mantui.
< |45| Dwa rody, zacne jednako i sławne
> |45| Dwa rody, zacne jednako i sławne —
< |50| Z łon tych dwu wrogów wzięło życie,
> |50| Z łon tych dwu wrogów wzięło bowiem życie,
< |53| Śmierć ich stłumiła rodzicielskie boje.
> |53| Śmierć ich rodzicielskie boje.
< |54|
< |55| Tej ich miłości przebieg zbyt bolesny
> |54|
> |55| Tej ich miłości przebieg zbyt
< |56| I jak się ojców nienawiść nie zmienia,
< |57| Aż ją zakończy dzieci zgon przedwczesny,
> |56| I jak się ojców nienawiść nie zmienia,
> |57| Aż ją zakończy dzieci przedwczesny,
< |61| Jest w nim co złego, my usuniem błędu…
> |61| Jest w nim co , my usuniem błędu…
< |72| / Plac publiczny. Wchodzą Samson i Grzegorz uzbrojeni w tarcze i miecze. /
> |72| / Plac publiczny. Samson i Grzegorz uzbrojeni w tarcze i miecze. /
< |78|
> |80| GREGORZ
< |88|
> |90| GREGORZ
< |97| Mam zwyczaj drapać zaraz, jak mię kto rozrucha.
> |97| zwyczaj drapać zaraz, jak mię kto rozrucha.
< |102| Tak, ale nie zwykłeś się dać rozruchać.
> |102| Tak, ale nie zaraz zwykłeś się dać rozruchać.
< |112| Rozruchać się tyle znaczy co ruszyć się z ; być walecznym jest to stać nieporuszenie: pojmuję , że skut
kiem rozruchania się twego będzie - drapnięcie.
> |112| Rozruchać się tyle znaczy co ruszyć się z miejsca; być walecznym jest to stać nieporuszenie: pojmuję ,
że skutkiem rozruchania się twego będzie - drapnięcie.
< |117| Te psy z domu Montekich rozruchać mię mogą tylko stania na miejscu. Będę jak dla każdego mężczyzny i ka
żdej kobiety z tego domu.
> |117| Te psy z domu Montekich rozruchać mię mogą tylko do stania na miejscu. Będę jak mur dla każdego mężczyz
ny i każdej kobiety z tego .
< |127| Prawda, dlatego to kobiety, jako najsłabsze, tulą się zawsze do muru. Ja odtrącę od muru ludzi Montekic
h, a kobiety Montekich przyprę do muru.
> |127| Prawda, dlatego to kobiety, jako najsłabsze, tulą się zawsze do muru. Ja też odtrącę od muru ludzi Mont
ekich, a kobiety Montekich do muru.
< |152| Tym lepiej, że się liczysz do zwierząt; bo gdybyś się liczył do ryb, to byłbyś pewnie sztokfiszem. Weź
no się za instrument, bo oto nadchodzi dwóch domowników Montekiego.
> |152| Tym lepiej, że się liczysz do zwierząt; bo gdybyś się liczył do , to byłbyś pewnie sztokfiszem. Weź no
się za instrument, bo oto nadchodzi domowników .
< |159| Mój giwer już dobyty: zaczep ich, ja stanę z tyłu.
> |159| Mój już dobyty: zaczep ich, ja stanę z tyłu.
< |174| Ja bym się miał bać z twojej przyczyny!
> |174| Ja bym się miał z twojej przyczyny!
< |189| Nie chcą, ale jak śmią. Ja im gębę wykrzywię; hańba im, jeśli to ścierpią.
> |189| Nie jak chcą, ale jak śmią. Ja im gębę ; hańba im, jeśli to ścierpią.
< |200|
> |202| ABRAHAM
< |204| Czy na nas się skrzywiłeś, mości panie?
> |204| Czy na nas się skrzywiłeś, panie?
< |211| Będziemy-ż mieli prawo sobą, jak powiem: tak jest?
> |211| Będziemy-ż mieli prawo za sobą, jak powiem: tak jest?
< |216| .
> |216| Nie.
< |221| Nie, mości panie; nie skrzywiłem na was, tylko skrzywiłem się tak sobie.
> |221| Nie, mości panie; nie skrzywiłem się na was, tylko skrzywiłem się tak sobie.
< |228| waść szukasz?
> |228| Zaczepki szukasz?
< |233| Zaczepki? nie.
> |233| Zaczepkinie.
< |238| Jeżeli jej szukasz, jestem na waścine usługi. Mój pan tak dobry jak i wasz.
> |238| Jeżeli jej szukasz, to jestem na waścine usługi. Mój pan tak dobry jak i wasz.
< |257| Powiedz: lepszy. Oto nadchodzi z krewnych mego pana.
> |257| Powiedz: lepszy. Oto nadchodzi jeden z krewnych mego pana.
< |272| Dobądźcie mieczów, macie serca. Grzegorzu, pamiętaj o swoim pchnięciu.
> |272| Dobądźcie mieczów, jeśli macie serca. Grzegorzu, pamiętaj o swoim pchnięciu.
< |277| Odstąpcie, głupcy; schowajcie miecze . Sami nie wiecie, robicie.
> |277| Odstąpcie, głupcy; schowajcie miecze do pochew. nie wiecie, co robicie.
< |281| / Wchodzi Tybalt. /
> |281| Wchodzi Tybalt. /
< |290| BENWOLIO
> |290|
< |299| Tego wyrazu, tak jak nienawidzę
> |299| wyrazu, tak jak nienawidzę
< |300| Szatana, wszystkich Montekich i ciebie.
< |301| Broń się, nikczemny tchórzu.
< |302|
< |303| / Walczą. Nadchodzi kilku przyjaciół obu partii mieszają się do zwady; wkrótce potem wchodzą mieszczani
e z pałkami.
> |300| Szatanawszystkich Montekich i ciebie.
> |301| Broń sięnikczemny tchórzu.
> |302|
> |303| Walczą. Nadchodzi kilku przyjaciół obu partii i mieszają się zwady; wkrótce potem wchodzą mieszczanie z
pałkami. /
< |308| Hola! berdyszów! pałek! Dalej po !
> |308| Hola! berdyszów! pałek! Dalej po nich!
< |317| Mój mieczhej!
> |317| Mój miecz! hej!
< |351| Zapamiętali niesforni poddani,
> |351| niesforni poddani,
< |356| Pod karą tortur wypuśćcie natychmiast
> |356| karą tortur wypuśćcie natychmiast
< |357| Z dłoni skrwawionych tę broń buntowniczą
< |358| I posłuchajcie tego, co niniejszym
> |357| Z dłoni skrwawionych tę broń buntowniczą
> |358| I tego, co niniejszym
< |362| Trzykroć już spokój miasta zakłóciły,
> |362| Trzykroć spokój miasta zakłóciły,
< |363| Tak że poważni wiekiem i zasługą
< |364| Obywatele werońscy musieli
> |363| Tak że poważni wiekiem i zasługą
> |364| Obywatele musieli
< |365| Porzucić swoje wygodne przybory
> |365| Porzucić wygodne przybory
< |368| Niechęci wasze przecinać. Jeżeli
> |368| Niechęci wasze przecinaćJeżeli
< |375| Dalsza ma wola oznajmiona będzie.
> |375| Dalsza ma wola oznajmiona będzie
< |376| Jeszcze raz wzywam tu obecnych
< |377| Pod karą śmierci, się rozeszli.
< |378|
< |379| / Książę z orszakiem wychodzi. Podobnież Kapulet, Pani Kapulet, Tybalt, obywatele i słudzy/
> |376| Jeszcze raz wzywam wszystkich tu obecnych
> |377| Pod karą śmierci, aby się rozeszli.
> |378|
> |379| / Książę z orszakiem wychodzi. Podobnież Kapulet, Pani Kapulet, Tybalt, obywatele i słudzy. /
< |384| Kto wszczął tę nową zwadę? Mów, synowcze,
> |384| Kto wszczął tę nową zwadę? , synowcze,
< |394| I harde zionąc w uszy wyzwanie,
> |394| I harde zionąc mi w uszy wyzwanie,
< |395| Jął się wywijać nim i siec powietrze,
> |395| Jął się nim i siec powietrze,
< |398| Cięcia i pchnięcia zamieniali, zbiegł się
> |398| Cięcia i pchnięcia zamienializbiegł się
< |416| Tam, już tak , syn wasz się przechadzał.
> |416| Tam, już tak rano, syn wasz się przechadzał.
< |417| Ledwiem go ujrzał, pobiegłem ku niemu;
< |418| Lecz on, spostrzegłszy , skrył natychmiast
< |419| I w najciemniejszej ukrył się gęstwinie.
< |420| Pociąg jego do
> |417| Ledwiem go ujrzał, pobiegłem niemu;
> |418| Lecz on, spostrzegłszy mię, skrył się natychmiast
> |419| I w najciemniejszej ukrył się gęstwinie.
> |420| Pociąg ten do odosobnienia
< |421| Mierząc mym własnym (serce nasze bowiem
< |422| Jest najczynniejsze, kiedyśmy samotni),
> |421| Mierząc mym własnym (serce nasze bowiem
> |422| Jest najczynniejsze, kiedyśmy samotni,
< |425| Stroniąc od tego, co rad mnie unikał.
> |425| od tego, co rad mnie unikał.
< |434| słońce sprzed łoża
> |434| Wesołe słońce sprzed łoża Aurory
< |435| Zaczęło ściągać cienistą kotarę,
< |436| On, uciekając od widoku światła,
> |435| Zaczęło ściągać cienistą kotarę,
> |436| Onuciekając widoku światła,
< |441| Jeśli się na to lekarstwo nie .
> |441| Jeśli się na lekarstwo nie znajdzie.
< |462| Lecz on jedyny powiernik swych smutków.
> |462| Lecz on powiernik swych smutków.
< |463| Tak im jest wierny, tak zamknięty w ,
> |463| Tak im jest wierny, tak zamknięty w sobie,
< |469| zbrakłoby nam zaradczego środka.
> |469| Nie zbrakłoby nam zaradczego środka.
< |482| Obyś w tej sprawie, co nam serce ,
> |482| Obyś w tej sprawie, co nam serce rani
< |488| BENWOLIO
> |488|
< |505|                          Jak nudnie
> |505| Jak nudnie
< |506| Wloką się chwile. Moi-ż to rodzice
< |507| Tak spiesznie w tamtą ulicę?
> |506| Wloką się chwile. Moi-ż to rodzice
> |507| Tak spiesznie w tamtą zboczyli ulicę?
< |522| Miłość więc?
> |522| Miłość ?
< |548| Niestety! Czemuż, z zasłoną na skroni,
> |548| Niestety! Czemuż, z na skroni,
< |551| Jakiś spór? Nie mów mi o nim, wiem wszystko.
> |551| Jakiś spór? Nie mów mi o , wiem wszystko.
< |552| W grze tu nienawiść wielka, lecz i miłość.
< |553| O! wy sprzeczności niepojęte !
> |552| W grze tu nienawiść wielka, lecz i miłość.
> |553| O! wy sprzeczności niepojęte dziwa!
< |557| Szpetny chaosie wdzięków! Ciężki puchu!
> |557| Szpetny wdzięków! Ciężki puchu!
< |558| Jasna mgło! Zimny żarzeMartwy ruchu!
> |558| Jasna mgło! Zimny żarze! Martwy ruchu!
< |561| Czy się nie śmiejesz?
> |561| Czy się śmiejesz?
< |571| Nad czym, poczciwa duszo?
> |571| Nad czympoczciwa duszo?
< |582|                          A więc strzała
> |582| A więc strzała
< |583| Miłości nawet przez odbitkę działa?
< |584| Dość mi już ciężył mój smutek, jego
> |583| Miłości nawet przez odbitkę działa?
> |584| Dość mi już ciężył mój smutek, ty jego
< |587| Nie ulgą, ale nowym jest kamieniem
> |587| ulgą, nowym jest kamieniem
< |588| Dla mego serca. Miłość, przyjacielu,
< |589| To dym, co z westchnień się unosi;
> |588| Dla mego serca. Miłość, przyjacielu,
> |589| To dym, co z parą westchnień się unosi;
< |592| Czymże jest więcej? Istnym amalgamem,
> |592| jest więcejIstnym amalgamem,
< |614|
> |616| ROMEO
< |619| Mam-że wraz jęczyć mówić?
> |619| Mam-że wraz jęczyć i mówić?
< |626| Kogóż to kochasz? Powiedz.
> |626| Kogóż to kochaszPowiedz.
< |632| Pisać testament: będzie-ż to wezwanie
> |632| Pisać : będzie-ż to wezwanie
< |640| Gdym to , nimeś mi powierzył.
> |640| Gdym to pomyślał, nimeś mi powierzył.
< |649| BENWOLIO
> |649|
< |661| Odpiera szturmy spojrzeń napastniczych
> |661| Odpiera szturmy spojrzeń napastniczych;
< |681| Temu skazanym - wieczne cierpieć męki.
> |681| Temu skazanym - wieczne cierpieć
< |686| Jest na to radaprzestań myśleć o niej.
> |686| Jest na to rada: przestań myśleć o niej.
```

In [ ]: