Zadanie 1.

Napisz metodę, która będzie sprawdzała poprawność wyrażenia matematycznego tylko w kontekście używania zwykłych nawiasów (), przyjmującą postać:

```
boolean validate(String expression) {}
```

Założenia: wyrażenie nie może przyjmować wartości `null`.

Przykłady:

```
(expression = result)
"(2 + 4) * 4" = true
"(4 * 4 - (log20 / 5)" = false
"3 - (5 * 2))" = false
"3 + 2x * 5" = true
```

Zadanie 2.

Napisz metodę, która będzie dzieliła tekst na segmenty zawierające się w znacznikach "<ABC>" i </ABC>" od pozostałych fragmentów tekstu, przyjmującą postać:

```
List<Segment> separate(String text) {}
```

Założenia: teksty nie zawsze będą zwierać znaczniki oraz nie mogą przyjmować wartości `null`.

```
class Segment{
   String text;
   boolean inTag;
}
```

Przykład:

```
"<ABC>Ala</ABC> ma <ABC> kota<ABC> a</ABC> kot </ABC> ma Ale"
```

Wynik:

```
{"Segment{text="<ABC>Ala</ABC>", inTag=true}";
"Segment{text=" ma ", inTag=false}";
"Segment{text="<ABC> kota<ABC> a</ABC> kot </ABC>", inTag=true}";
"Segment{text=" ma Ale", inTag=false}";
}
```

Zadanie 3.

Zbuduj klasyfikator, który przewiduje typ dowolnej strony internetowej na podstawie jej struktury i treści. Klasyfikator powinien działać na poziomie najwyższym (lvl0).

Klasyfikator powinien klasyfikować poszczególne strony na poziomie URL z wykorzystaniem następujących typów:

- HIGH_QUALITY_CONTENT – strona zawierająca co najmniej 3 zdania, z wysokiej jakości treściami, które stanowią wartość z punktu widzenia ekstrakcji informacji: artykułami, raportami, itp.
- FORUM – strona zawierająca jedynie komentarze internautów na dany temat, na temat

danego artykułu, produktów czy usług
- OTHER_TEXT_CONTENT – strona zawierająca treść ale nie stanowiąca wielką wartość z punktu widzenia ekstrakcji treści (strony informacyjne, dane tabelaryczne, itp.)
- OTHER_MULTIMEDIA_CONTENT -> strona zawierająca jedynie treść multimedialne (javascript, aplikacje, galerie zdjęć i video, itp.)
- NO_CONTENT – strona nie zawierająca wartościowej treści w postaci tekstu lub multimediów (strony do logowania, formularze, itp)
- UNKNOWN – klasyfikator nie jest w stanie określić typu strony

Zbiór uczący train.tar.gz, udostępniony jest tu:

https://proxy.applica.pl/train.tar.gz

Struktura udostępnionego pliku wygląda następująco: każdy z plików zawiera informację o pojedynczej stronie internetowej w postaci:
Line 1: PageTypelvl0 PageTypelvl1
Line 2: URL
Lines 3-...: HTML code
Rozwiązania prosimy przesłać na adres: adam.dancewicz@applica.pl