

Test for Java Software Developer position.

Expectations:

- Write clean, self-describing production ready code with OOP in mind.
- Cover your implementation with unit tests
- If you have questions, ask :)

Demultiplexer

Create a simplified sequencing data demultiplexer. The tool receives a list of DNA sequences ([ACTG]*) which should be split into groups based on the configuration defined in the sample.conf file.

The tool receives the following input data:

- path to the sequencing data file: text file, contains one sequence per line,
- path to the configuration file: text file, each line contains the group name, sequence prefix and postfix separated by ws,
- path prefix to the output files which serves as a base to generate output filenames for each group.
- name of the alignment we want to run (endsAlignment or midAlignment or bestAlignment)

Behaviour of alignments:

endsAlignment: if a sequence has matching prefix and postfix by the selected alignment it should go into the corresponding group.

midAlignment: : if a sequence has matching infix by the selected alignment it should go into the corresponding group.

bestAlignment: get the sequences that match the most with the provided infix. Matching most means that if you fit the infix to any substring of the sequence the number of the identical characters on the same position will be the highest.

Sequences not matching any group should go into the 'unmatched' group.

Example

sample.seq:

ACTCACGACCACTAACTAGCAATACGATCG

CAGTAAGCGATCAGACAGTACAGACGTACA

ACTCACGACCACTAACTGGCAATACGATCG

AGACAACATCAGATCGCAAGACGACAGATA

sample.conf:

```
{
  "endsAlignment" : {
    "group1" : {
      "prefix" : "ACTCACG",
      "postfix" : "ACGATCG"
    },
    "group2" : {
      "prefix" : "CAGTAAG",
      "postfix" : "ACGTACA"
    }
  },
  "midAlignment" : {
    "group1" : {
      "infix" : CACTAACT
    },
    "group2" : {
      "infix" : CAGACAGT
    }
  },
  "bestAlignment" : {
    "group1" : {
      "infix" : CTATCTAGCAAT
    }
  }
}
```

Expected outputs:

1. For *endsAlignment*:

group1.seq:

ACTCACGACCACTAACTAGCAATACGATCG ACTCACGACCACTAACTGGCAATACGATCG

group2.seq:

CAGTAAGCGATCAGACAGTACAGACGTACA

unmatched.seq:

AGACAACATCAGATCGCAAGACGACAGATA

2. For *midAlignment*:

group1.seq:

ACTCACGACCACTAACTAGCAATACGATCG ACTCACGACCACTAACTGGCAATACGATCG

group2.seq:

CAGTAAGCGATCAGACAGTACAGACGTACA

unmatched.seq:

AGACAACATCAGATCGCAAGACGACAGATA

3. For *bestAlignment*:

group1.seq:

ACTCACGACCACTAACTAGCAATACGATCG

unmatched.seq:

AGACAACATCAGATCGCAAGACGACAGATA CAGTAAGCGATCAGACAGTACAGACGTACA
ACTCACGACCACTAACTGGCAATACGATCG