

Subsequences in Context

The nucleotide type takes on one of the values A, C, G, T, or the special value ε . Given a stream function

```
nucleotide next(stream *s);
```

which produces the next value in a stream of nucleotides (or which produces ε if there are no more elements in the stream), implement an online algorithm to find sequences of values and to print them along with some surrounding context. Specifically, given a target sequence T and two numbers x and y , print the x preceding stream values, the target, then the y succeeding stream values. Do this for each occurrence of the target.

Be aware that:

- Either of x or y (or both) may be zero, indicating that no preceding or succeeding context should be printed.
- T will not be empty.
- The end-of-stream marker ε will not appear in T .
- If the stream produces fewer than x nucleotides before T , or fewer than y nucleotides after T , print as many as there are.
- Targets may overlap in the stream, and each should be treated as a distinct match.
- The stream has no maximum size: it may be larger than your program's available memory.

Deliverable

Please deliver the source code to a working program which

- reads standard input as the nucleotide stream and
- takes x , T , and y as command-line arguments.

Please indicate any special requirements your program has for compilation. Your program will be run with valid input.

Example

Suppose the stream produces

AAGTACGTGCAGTGAGTAGTAGACCTGACGTAGACCGATATAAGTAGCTA ε

Your program, when run with $x = 5$, $T = \text{AGTA}$, and $y = 7$, should find the first target sequence T with context as shown here:

$\underbrace{\text{A}}_x \underbrace{\text{AGTA}}_T \underbrace{\text{CGTGCAG}}_y \text{TGAGTAGTAGACCTGACGTAGACCGATATAAGTAGCTA}\varepsilon$

The program should print the following lines (whose targets we've marked):

A $\underbrace{\text{AGTA}}_T$ CGTGCAG (1)

CAGTG $\underbrace{\text{AGTA}}_T$ GTAGACC (2)

TGAGT $\underbrace{\text{AGTA}}_T$ GACCTGA (3)

ATATA $\underbrace{\text{AGTA}}_T$ GCTA (4)

Notice that lines (2) and (3) display overlapping targets and that (1) and (4) show fewer than x and y elements of context, respectively.