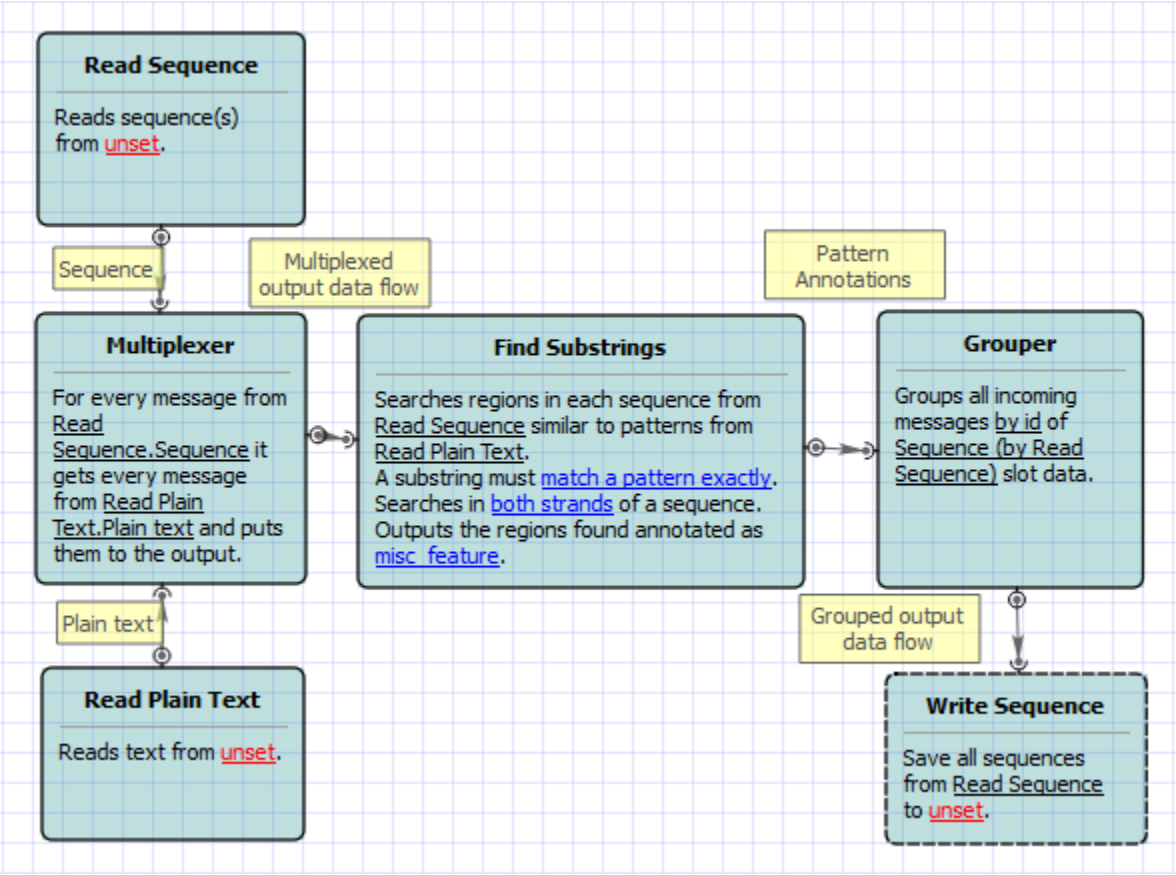
Find Substrings at Sequences

This scheme describes how to find substrings in sequences and group these sequences by different parameters.

First, the workflow reads sequences and text strings (patterns) from files. Then, these data sets are multiplexed using this rule: every sequence is united with every pattern. After multiplexing these united data sets are transported to the find patterns element. The results of patterns searching are grouped by id of a sequence: original and find patterns annotations are united into two new grouped annotations sets. And finally, the grouped data are written into file, specified by a user.

By default, sequence multiplexed using the rule “1 to 1”. You can configure this value in the *Multiplexer* element parameters. Also, you can configure the *Pattern* element parameters and *Grouper* element parameters.

To try out this sample, add the input files to the *Read Sequence* and *Read Plain Text* elements, select the name and location of the output files in the *Write Sequence* element and run the workflow.



Also, if required, you can change parameters. Use the workflow wizard to guide you through the parameters setup process. The first wizard page will appear when you click on the Show wizard button on the Workflow Designer toolbar:

