CLAIMS

1. A system for segmenting an input data stream using vector processing, comprising:

a processor adapted to repeat the following steps throughout an input data stream to create a segmented data stream consisting a plurality of segments:

apply a rolling sequence over a sequence of consecutive data items of an input data stream, the rolling sequence includes a subset of consecutive data items of the sequence;

calculate, concurrently, a plurality of partial hash values, each by one of a plurality of processing pipelines of the processor, each for a respective one of a plurality of partial rolling sequences each including evenly spaced data items of the subset;

determine compliance of each of the plurality of partial hash values with at least one respective partial segmentation criterion; and

designate the sequence as a variable size segment when at least some of the plurality of partial hash values comply with the respective at least one partial segmentation criterion.

2. The system according to claim 1, wherein the processor is a single-instruction-multiple data, SIMD, processor.

3. The system according to claim 1 or 2, wherein the processor is adapted to calculate each of the plurality of partial hash values as a partial rolling hash value using a respective partial hash value of a respective previous partial rolling sequence, an omitted data item which is omitted from the respective partial rolling sequence and an added data item which is added to the respective partial rolling sequence.

4. The system according to any of claims 1 to 3, wherein the processor is adapted to designate the sequence as the variable size segment when a number of consecutive complying partial hash values calculated for successive subsets of

the sequence which comply with the respective at least one partial segmentation criterion exceeds the number of the plurality of partial rolling sequences.

5. The system according to any of claims 1 to 4, wherein the sequence includes a pre-defined minimum number of the consecutive data items.

6. The system according to any of claims 1 to 5, wherein the processor is adapted to designate at least one large sequence of the plurality of sequences as the variable size segment when a size of the large sequence exceeds a pre-defined maximum value before detecting compliance of the partial hash values with the respective at least one partial segmentation criterion.

7. A method for segmenting an input data stream using vector processing, comprising:

using a processor adapted to repeat the following steps throughout an input data stream to create a segmented data stream consisting a plurality of segments:

apply a rolling sequence over a sequence of consecutive data items of an input data stream, the rolling sequence includes a subset of consecutive data items of the sequence;

calculate, concurrently, a plurality of partial hash values, each by one of a plurality of processing pipelines of processor, each for a respective one of a plurality of partial rolling sequences each including evenly spaced data items of the subset;

determine compliance of each of the plurality of partial hash values with at least one respective partial segmentation criterion; and

designate the sequence as a variable size segment when at least some of the plurality of partial hash values comply with the respective at least one partial segmentation criterion.

8. The method according to claim 7, wherein each of the plurality of partial hash values is calculated as a partial rolling hash value using a respective partial hash value of a respective previous partial rolling sequence, an omitted data

item which is omitted from the respective partial rolling sequence and an added data item which is added to the respective partial rolling sequence.

9. The method according to claim 7 or 8, wherein the sequence is designated as the variable size segment when a number of consecutive complying partial hash values calculated for successive subsets of the sequence which comply with the respective at least one partial segmentation criterion exceeds the number of the plurality of partial rolling sequences.

10. The method according to any of claims 7 to 9, wherein the sequence includes a pre-defined minimum number of the consecutive data items.

11. The method according to any of claims 7 to 10, wherein at least one large sequence of the plurality of sequences is designated as the variable size segment when a size of the large sequence exceeds a pre-defined maximum value before detecting compliance of the partial hash values with the respective at least one partial segmentation criterion.