Multi-Digest problem

Problem formulation

Given the fragments lengths obtained by complete digestion of an unknown original sequence with each one of k restriction enzymes and the fragments lengths obtained by complete digestion of this sequence by all k enzymes combined; the goal is to reconstruct each enzyme-specific restriction sites map on the original sequence.

k = number of restriction enzymes.

Input: $k + 1$ multisets of fragments lengths. There are k multisets of fragments, each obtained by complete digestion with one of the k enzymes, and one multiset of fragments obtained by complete digestion with all k enzymes combined. Fragments lengths are represented by integers.

Output: k enzyme-specific restriction sites maps. Restriction sites are positions on the original sequence and are represented by integers. They are entirely enzyme-specific. Thus, one site cannot be in multiple restriction sites maps.

Algorithm

Please see attached the algorithm code $multiDigestProblem.py$. Explanations are in the form of comment lines.

Result outputs

Input: k = 2, individual digests {1, 2, 4, 5, 6}, {1, 3, 3, 11} and combined {1, 1, 1, 1, 2, 3, 4, 5}:
Output:

- Enzyme 1 restriction sites: $[1, 5, 10, 12]$
- Enzyme 2 restriction sites: $[11, 14, 15]$

Input: k = 3, individual digests {2, 7, 9}, {4, 4, 5, 5}, {2, 16}, and combined {2, 2, 2, 2, 2, 3, 5}:
Output:

- Enzyme 1 restriction sites: $[2, 11]$
- Enzyme 2 restriction sites: $[16]$
- Enzyme 3 restriction sites: $[4, 9, 13]$