1993 Paper 2 Question 9

A *multiset* is an unordered collection of elements. An element may occur zero or more times. Two multisets are equal if they contain the same number of occurrences of each element. For example, the multisets $\{a, a, b\}_m$ and $\{a, b, a\}_m$ are equal; they differ from $\{a, b\}_m$.

Let M and N stand for multisets. Write $\texttt{occs}(x, M)$ for the number of occurrences of x in M. Write $\texttt{mplus}(M, N)$ for the sum of M and N, defined by

$$\texttt{occs}(x, \texttt{mplus}(M, N)) = \texttt{occs}(x, M) + \texttt{occs}(x, N)$$

for all x. Write $\texttt{mequal}(M, N)$ for the equality test for M and N.

The question concerns how to represent multisets of strings within ML. For each of the given data representations (a), (b) and (c) describe how you would implement `occs`, `mplus` and `mequal`. If possible, incorporate simple efficiency improvements. In each case state the approximate running time of $\texttt{occs}(x, M)$, $\texttt{mplus}(M, N)$ and $\texttt{mequal}(M, N)$ using O-notation.

(a) Represent multisets using lists of strings, for example $\{a, b, a\}_m$ by $[a, b, a]$.

(b) Represent multisets by lists of pairs of the form (x, k), for example $\{a, b, a\}_m$ by $[(a, 2), (b, 1)]$.

(c) Represent multisets using binary trees.

Compare the representations (a), (b) and (c), stating the advantages and drawbacks of each.

[Answers need not contain ML code. You may refer to algorithms and data structures from the notes *Problem Solving in ML*.]