

## Question 4

(Note: I use the convention that  $\mathbb{N}$  does not include 0)

We will use numbers and sequences to represent letters and words, this is to make notation easier to work with. So for example, 'a' will be represented by 0, 'b' with 1, and 'z' with 25. The word *abdc* will be represented by the sequence  $(0, 1, 3, 2)$ .

We define a factorial number system, where the  $i$ th digit has base  $i$ , and the place value of the  $i$ th digit is  $(i-1)!$ . A number in the factorial system is written with pipes separating the digits and subscripted with '!'. For example,  $3|0|2|1|0! = 3 \times 4! + 0 \times 3! + 2 \times 2! + 1 \times 1! + 0 \times 0! = 77$ . The highest digit allowed in the  $i$ th digit is  $i-1$ , hence the first digit in the factorial number system is always zero.

Fix some  $n \in \mathbb{N}$ . A sequence  $(a_0, \dots, a_{n-1})$  of natural numbers is a *good* sequence just in case each number  $0, 1, \dots, n-1$  appears exactly once in the sequence. Let  $A$  be the set of good sequences. Note that  $|A| = n!$ . We may define a total order on  $A$  by way of lexicographic order, it is easy to see that the words will be published in lexicographic order. Let  $B = \{0, 1, 2, \dots, n! - 1\}$  we define a map  $f : B \rightarrow A$  given by

$$f(x_0|x_1|\dots|x_{n-1}!) = (L(S_0)_{x_0}, L(S_1)_{x_1}, \dots, L(S_{n-1})_{x_{n-1}})$$

Where  $L(S)$  takes a finite set of natural numbers and returns the sequence of natural numbers in  $S$  in strictly increasing order,  $L(S)_i$  denotes the  $i$ th element in the sequence (starting from zero), and  $S_0 = \{0, \dots, n-1\}$ ,  $S_1 = S_0 \setminus \{L(S_0)_{x_0}\}$ ,  $S_2 = S_1 \setminus \{L(S_1)_{x_1}\}$  etc...

An example will clear things up, take  $n = 4$  and consider  $f(3) = (a_0, a_1, a_2, a_3)$ . Note that  $3 = 0|1|1|0!$ .

- $S_0 = \{0, 1, 2, 3\}$ , hence  $L(S_0) = (0, 1, 2, 3)$ , so  $a_0 = L(S_0)_0 = (0, 1, 2, 3)_0 = 0$ .
- $S_1 = S_0 \setminus a_0 = \{1, 2, 3\}$ , hence,  $L(S_1) = (1, 2, 3)$ , so  $a_1 = L(S_1)_1 = (1, 2, 3)_1 = 2$ .
- $S_2 = S_1 \setminus a_1 = \{1, 3\}$ , hence  $L(S_2) = (1, 3)$ , so  $a_2 = L(S_2)_1 = (1, 3)_1 = 3$ .
- $S_3 = S_2 \setminus a_2 = \{1\}$ , hence  $L(S_3) = (3)$ , so  $a_3 = L(S_3)_0 = (1)_0 = 1$ .

Hence  $f(3) = (0, 2, 3, 1)$

It remains to show that for every  $x_0|x_1|\dots|x_{n-1}! \in B$ ,  $f(x_0|x_1|\dots|x_{n-1}!) = (a_0, \dots, a_n)$  does give a good sequence, and indexing the sequence  $L(S_i)$  with index  $x_i$  does not go out of bounds. This is the case since at each  $i$ ,  $S_i = S_0 \setminus \{a_0, \dots, a_{i-1}\}$ , hence  $a_i$  cannot be equal to any of  $a_0, \dots, a_{i-1}$ . The indexing does not go out of bounds since  $x_i$  is at the  $n-i$ th digit and by definition of the factorial number system,  $x_i < n-i$ , in addition,  $|S_i| = n-i$ , so  $L(S_i)_{x_i}$  is well defined.

Claim:  $f$  is strictly increasing

Let  $x = x_0|x_1|\dots|x_{n-1}!$  and  $y = y_0|y_1|\dots|y_{n-1}!$  such that  $x < y$ . Then there exist an  $i$  such that  $x_0 = y_0, x_1 = y_1, \dots, x_{i-1} = y_{i-1}$  and  $x_i < y_i$ . Let  $f(x) = (a_0, \dots, a_{n-1})$  and  $f(y) = (b_0, \dots, b_{n-1})$  then it follows that  $a_0 = b_0, a_1 = b_1, \dots, a_{i-1} = b_{i-1}$  (since the last  $i$  digits of  $x$  and  $y$  are the same). Hence  $S_i = S_0 \setminus \{a_0, \dots, a_{i-1}\} = S_0 \setminus \{b_0, \dots, b_{i-1}\}$ , now  $a_i = L(S_i)_{x_i}$  and  $b_i = L(S_i)_{y_i}$ , since  $L(S_i)$  is an increasing sequence, and  $x_i < y_i$ , it follows that  $a_i < b_i$ . And so  $(a_0, \dots, a_{n-1}) < (b_0, \dots, b_{n-1})$  (lexicographically)

Since  $f$  is a strictly increasing function between two finite totally ordered sets of the same size, it follows that  $f$  is a bijection.

With the relevant theory established, we return to the question at hand. There are  $26!/21$  words in a volume, we wish to find the last word of the first volume, i.e. the  $26!/21$ th word,

which corresponds to  $f(26!/21 - 1)$ . We will calculate  $26!/21$  in the factorial number system then subtract 1.