

NAG Toolbox for MATLAB

m01dc

1 Purpose

m01dc ranks a vector of character data in ASCII or reverse ASCII order of a specified substring.

2 Syntax

```
[irank, ifail] = m01dc(ch, m1, l1, l2, order, 'm2', m2)
```

3 Description

m01dc uses a variant of list-merging, as described by Knuth 1973, pp 165-166. The function takes advantage of natural ordering in the data, and uses a simple list insertion in a preparatory pass to generate ordered lists of length at least 10. The ranking is stable: equal elements preserve their ordering in the input data.

Only the substring (**l1:l2**) of each element of the array **ch** is used to determine the rank order.

4 References

Knuth D E 1973 *The Art of Computer Programming (Volume 3)* (2nd Edition) Addison–Wesley

5 Parameters

5.1 Compulsory Input Parameters

1: **ch(m2)** – string array

Elements **m1** to **m2** of **ch** must contain character data to be ranked.

Constraint: the length of each element of **ch** must not exceed 255.

2: **m1** – int32 scalar

The index of the first element of **ch** to be ranked.

Constraint: **m1** > 0.

3: **l1** – int32 scalar

4: **l2** – int32 scalar

Only the substring (**l1:l2**) of each element of **ch** is to be used in determining the rank order.

Constraint: $0 < \mathbf{l1} \leq \mathbf{l2} \leq \text{LEN}(\mathbf{ch}(1))$.

5: **order** – string

If **order** = 'A', the values will be ranked in ASCII order.

If **order** = 'R', in reverse ASCII order.

Constraint: **order** = 'A' or 'R'.

5.2 Optional Input Parameters

1: **m2** – int32 scalar

Default: The dimension of the arrays **irank**, **ch**. (An error is raised if these dimensions are not equal.)

the index of the last element of **ch** to be ranked.

Constraint: $\mathbf{m2} \geq \mathbf{m1}$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **irank(m2)** – int32 array

Elements **m1** to **m2** of **irank** contain the ranks of the corresponding elements of **ch**. Note that the ranks are in the range **m1** to **m2**: thus, if **ch**(*i*) is the first element in the rank order, **irank**(*i*) is set to **m1**.

2: **ifail** – int32 scalar

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

On entry, **m2** < 1,
or **m1** < 1,
or **m1** > **m2**,
or **l2** < 1,
or **l1** < 1,
or **l1** > **l2**,
or **l2** > LEN(**ch**(1)).

ifail = 2

On entry, **order** is not 'A' or 'R'.

ifail = 3

On entry, the length of each element of **ch** exceeds 255.

7 Accuracy

Not applicable.

8 Further Comments

The average time taken by the function is approximately proportional to $n \times \log n$, where $n = \mathbf{m2} - \mathbf{m1} + 1$.

The function relies on the Fortran 77 intrinsic functions LLT and LGT to order characters according to the ASCII collating sequence.

9 Example

```
ch = {'A02AAF' 289'; 'A02ABF' 523'; 'A02ACF' 531'; 'C02ADF' 169';  
      'C02AEF' 599'; 'C05ADF' 1351'; 'C05AGF' 240'; 'C05AJF' 136'; 'C05AVF'  
      211'; 'C05AXF' 183'; 'C05AZF' 2181'};  
m1 = int32(1);
```

```
l1 = int32(7);  
l2 = int32(12);  
order = 'Reverse ASCII';  
[irank, ifail] = m01dc(ch, m1, l1, l2, order)
```

```
irank =  
      6  
      5  
      4  
     10  
      3  
      2  
      7  
     11  
      8  
      9  
      1  
  
ifail =  
      0
```
