XVIII. A PROGRAM TO EVALUATE THE ITERATIVE SEARCH PROCESS (MORVAL)

Michael Lesk

The various processing methods available in the SMART system[1] can be studied singly or in combination. The use of combined processing methods can produce more accurate retrieval when the methods that are combined complement each other, producing different relevant documents near the top of the rank list. To aid in the evaluation of multiple processing methods, a program named MORVAL has been written which runs as a post-processor of SMART retrieval operations. This program is used to merge the output from different computer runs of the main SMART system, and to compute evaluation measures for the merged output.

MORVAL requires as input the document rank lists produced by each method for each request, and a set of instructions telling which methods are to be merged in this run. For each request and each set of merged methods, a table is produced containing the rank list of the "top" 15 documents produced by the merged method, as well as the ranks of the relevant documents for the merged method, and the SMART evaluation measures for the method.[2] At the end of a MORVAL run, "Cleverdon-type" recall-precision graphs are prepared for each method (averaged over the requests in the run). Also, the averages of the evaluation measures over the requests are printed.

The document rank list for a merged method is prepared by interleaving the rank lists for the individual methods which make up this merged method.

Specifically, the merged list is formed by taking documents from each input method in turn, ignoring documents which have previously been added to the merged list. Thus, the result of merging the three lists A(i), B(i), and C(i), would be the list A(1), B(1), C(1), A(2), B(2), C(2), A(3), ... when duplicate entries are removed. If for example B(2) happens to be the same as A(1), the element of the merged list following A(2) would be C(2) instead. Ties within the input lists are ignored (they are very rare when the cosine correlation is used and the requests are well phrased).

Sample output from the merged methods program is shown in Fig. 1. The request name is listed first, followed by the number of relevant documents. Then the tables for each merged method are printed. The names of the methods merged are printed above the corresponding data. In this example, 28 merges are performed, some of which are merges of single methods, and therefore not "merges" in the literal sense. For each merged method, the top 15 documents in the merged rank list are listed under "TOP 15", while the relevant documents are listed with their rank orders under "RELEVANT." The rank of each relevant document in the merged list is printed immediately to the left of the three-digit document number. This format is patterned after the output of the evaluation link in the SMART retrieval system,[3] but is condensed as much as possible so that the maximum amount of information can be kept on one page. Underneath the tables of documents, the rank recall, log precision, normalized recall, normalized precision, overall, and normalized overall measures for this run are listed.

DIFFERNTL EQ 16 RELEVANT

HARRIS TWO		NULL THES		STAT PHRASE SYNTAX PHR		STAT PHRASE		SYNTAX PHR		HIERARCHY UP		HIER DOWN		HARRIS TWO NULL THES	
TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT
1 384	1 384	1 103	1 103	1 384	1 384	1 384	1 384	1 384	1 384	1 103	1 103	1 384	1 384	1 384	1 384
2 360	2 360	2 384	2 384	2 360	2 360	2 360	2 360	2 360	2 360	2 392	2 392	2 102	2 102	2 103	2 103
3 200	3 200	3 392	3 392	3 386	3 386	3 386	3 386	3 386	3 386	3 384	3 384	3 200	3 200	3 360	3 360
4 392	4 392	4 387	4 387	4 392	4 392	4 392	4 392	4 085	4 085	4 360	4 360	4 103	4 103	4 200	4 200
5 386	5 386	5 386	5 386	5 085	5 085	5 200	5 200	5 392	5 392	5 386	5 386	5 392	5 392	5 392	5 392
6 103	6 103	6 102	6 102	6 200	6 200	6 085	6 085	6 387	6 387	6 200	6 200	6 360	6 360	6 387	6 387
7 085	7 085	7 202	7 202	7 387	7 387	7 387	7 387	7 103	7 103	7 234	9 387	7 169	8 386	7 386	7 386
8 192	9 358	8 360	8 360	8 103	8 103	8 103	8 103	8 358	8 358	8 128	12 202	8 386	10 202	8 102	8 102
9 358	10 102	9 192	10 200	9 358	9 358	9 102	9 102	9 390	9 390	9 387	13 358	9 229	11 085	9 085	9 085
10 102	11 387	10 200	11 358	10 102	10 102	10 390	10 390	10 388	10 388	10 226	16 085	10 202	12 251	10 202	10 202
11 387	12 202	11 358	12 385	11 390	11 390	11 358	11 358	11 202	11 202	11 045	21 385	11 085	13 358	11 192	12 358
12 202	15 251	12 385	13 251	12 388	12 388	12 388	12 388	12 385	12 385	12 202	31 102	12 251	15 387	12 358	13 385
13 229	17 253	13 251	14 085	13 202	13 202	13 202	13 202	13 102	13 102	13 358	53 251	13 358	17 253	13 385	15 251
14 088	23 390	14 085	18 253	14 385	14 385	14 385	14 385	14 169	15 200	14 306	55 253	14 234	22 390	14 229	19 253
15 251	24 388	15 229	66 388	15 169	23 251	15 169	21 251	15 200	20 251	15 192	88 390	15 387	27 388	15 251	29 390
	40 385		113 390		29 253		25 253		30 253		94 388		43 385		30 388
RNK REC=	0.7196	RNK REC=	0.4642	RNK REC=	0.8662	RNK REC=	0.9007	RNK REC=	0.8718	RNK REC=	0.3293	RNK REC=	0.6834	RNK REC=	0.7861
LOG PRE=	0.9169	LOG PRE=	0.8813	LOG PRE=	0.9677	LOG PRE=	0.9751	LOG PRE=	0.9689	LOG PRE=	0.7653	LOG PRE=	0.8995	LOG PRE=	0.9416
NOR REC=	0.9915	NOR REC=	0.9748	NOR REC=	0.9966	NOR REC=	0.9976	NOR REC=	0.9968	NOR REC=	0.9555	NOR REC=	0.9899	NOR REC=	0.9941
NOR PRE=	0.9573	NOR PRE=	0.9365	NOR PRE=	0.9843	NOR PRE=	0.9880	NOR PRE=	0.9849	NOR PRE=	0.8554	NOR PRE=	0.9473	NOR PRE=	0.9707
OVERALL=	1.6365	OVERALL=	1.3455	OVERALL=	1.8340	OVERALL=	1.8758	OVERALL=	1.8407	OVERALL=	1.0946	OVERALL=	1.5829	OVERALL=	1.7277
NOR OVR=	1.9147	NOR OVR=	1.8104	NOR OVR=	1.9674	NOR OVR=	1.9759	NOR OVR=	1.9688	NOR OVR=	1.6329	NOR OVR=	1.8967	NOR OVR=	1.9410

HARRIS TWO STAT PHRASE		HARRIS TWO SYNTAX PHR		HARRIS TWO HIERARCHY UP		HARRIS TWO HIER DOWN		STAT PHRASE HIERARCHY UP		STAT PHRASE HIER DOWN		SYNTAX PHR HIERARCHY UP		SYNTAX PHR HIER DOWN	
TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT
1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384
2 360	2 360	2 360	2 360	2 103	2 103	2 360	2 360	2 103	2 103	2 360	2 360	2 103	2 103	2 360	2 360
3 200	3 200	3 200	3 200	3 360	3 360	3 102	3 102	3 360	3 360	3 102	3 102	3 360	3 360	3 102	3 102
4 386	4 386	4 386	4 386	4 392	4 392	4 200	4 200	4 392	4 392	4 386	4 386	4 392	4 392	4 386	4 386
5 392	5 392	5 392	5 392	5 200	5 200	5 392	5 392	5 386	5 386	5 200	5 200	5 386	5 386	5 200	5 200
6 103	6 103	6 085	6 085	6 386	6 386	6 103	6 103	6 200	6 200	6 392	6 392	6 085	6 085	6 085	6 085
7 085	7 085	7 103	7 103	7 085	7 085	7 386	7 386	7 085	7 085	7 103	7 103	7 387	7 387	7 103	7 103
8 387	8 387	8 387	8 387	8 234	11 358	8 085	8 085	8 387	8 387	8 085	8 085	8 200	8 200	8 392	8 392
9 192	10 358	9 192	10 358	9 192	12 387	9 169	11 358	9 234	11 102	9 387	9 387	9 234	10 358	9 387	9 387
10 358	11 102	10 358	11 390	10 128	13 102	10 192	13 202	10 128	12 390	10 169	12 390	10 358	12 390	10 169	11 358
11 102	12 390	11 390	12 102	11 358	16 202	11 358	14 387	11 102	14 358	11 229	13 202	11 128	13 388	11 358	12 390
12 390	13 202	12 102	13 388	12 387	20 251	12 229	15 251	12 390	16 388	12 390	14 358	12 390	15 202	12 390	14 388
13 202	14 388	13 388	14 202	13 102	21 253	13 202	19 253	13 226	17 202	13 202	15 388	13 388	17 385	13 229	15 202
14 388	17 385	14 202	15 385	14 226	28 385	14 387	27 390	14 358	18 385	14 358	16 251	14 226	18 102	14 388	16 385
15 229	18 251	15 385	19 251	15 045	29 390	15 251	28 388	15 045	30 251	15 388	17 385	15 202	28 251	15 202	17 251
	21 253		21 253		31 388		52 385		34 253		21 253		43 253		21 253
RNK REC=	0.8947	RNK REC=	0.9007	RNK REC=	0.6507	RNK REC=	0.6326	RNK REC=	0.7234	RNK REC=	0.8889	RNK REC=	0.7083	RNK REC=	0.9007
LOG PRE=	0.9657	LOG PRE=	0.9678	LOG PRE=	0.8784	LOG PRE=	0.8906	LOG PRE=	0.9142	LOG PRE=	0.9630	LOG PRE=	0.9161	LOG PRE=	0.9681
NOR REC=	0.9974	NOR REC=	0.9976	NOR REC=	0.9883	NOR REC=	0.9873	NOR REC=	0.9916	NOR REC=	0.9973	NOR REC=	0.9910	NOR REC=	0.9976
NOR PRE=	0.9832	NOR PRE=	0.9843	NOR PRE=	0.9347	NOR PRE=	0.9421	NOR PRE=	0.9558	NOR PRE=	0.9819	NOR PRE=	0.9568	NOR PRE=	0.9845
OVERALL=	1.8604	OVERALL=	1.8685	OVERALL=	1.5291	OVERALL=	1.5232	OVERALL=	1.6376	OVERALL=	1.8519	OVERALL=	1.6244	OVERALL=	1.8688
NOR OVR=	1.9704	NOR OVR=	1.9723	NOR OVR=	1.8761	NOR OVR=	1.8786	NOR OVR=	1.9140	NOR OVR=	1.9682	NOR OVR=	1.9118	NOR OVR=	1.9724

Sample MORVAL Output for One Request

Figure 1

HIERARCHY UP HIER DOWN		NULL THES STAT PHRASE		NULL THES SYNTAX PHR		HARRIS TWO NULL THES STAT PHRASE		HARRIS TWO NULL THES SYNTAX PHR		HARRIS TWO NULL THES HIERARCHY UP		HARRIS TWO NULL THES HIER DOWN		HARRIS TWO STAT PHRASE HIERARCHY UP	
TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT
1 103	1 103	1 103	1 103	1 103	1 103	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384
2 384	2 384	2 384	2 384	2 384	2 384	2 103	2 103	2 103	2 103	2 103	2 103	2 103	2 103	2 103	2 103
3 392	3 392	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360	3 360
4 102	4 102	4 392	4 392	4 392	4 392	4 200	4 200	4 200	4 200	4 392	4 392	4 102	4 102	4 392	4 392
5 200	5 200	5 386	5 386	5 386	5 386	5 392	5 392	5 392	5 392	5 200	5 200	5 200	5 200	5 200	5 200
6 360	6 360	6 387	6 387	6 387	6 387	6 386	6 386	6 386	6 386	6 387	6 387	6 392	6 392	6 386	6 386
7 386	7 386	7 200	7 200	7 085	7 085	7 387	7 387	7 387	7 387	7 386	7 386	7 387	7 387	7 085	7 085
8 234	11 387	8 102	8 102	8 102	8 102	8 102	8 102	8 085	9 085	8 102	8 102	8 386	8 386	8 387	8 387
9 169	14 202	9 085	9 085	9 202	9 202	9 085	9 085	9 102	9 102	9 085	9 085	9 085	9 085	9 234	12 358
10 128	16 085	10 202	10 202	10 358	10 358	10 202	10 202	10 202	10 202	10 202	10 202	10 202	10 202	10 192	13 102
11 387	17 251	11 192	12 390	11 192	12 390	11 192	12 358	11 192	12 358	11 234	14 358	11 169	13 358	11 128	14 390
12 229	18 358	12 390	13 358	12 390	13 200	12 358	13 390	12 358	13 390	12 192	17 385	12 192	15 395	12 358	17 202
13 226	23 253	13 358	14 385	13 200	14 388	13 390	14 385	13 390	14 388	13 128	19 251	13 358	16 251	13 102	18 388
14 202	29 385	14 385	15 388	14 385	15 385	14 385	15 388	14 388	15 385	14 358	23 253	14 229	21 253	14 390	21 385
15 045	31 390	15 388	16 251	15 385	16 251	15 388	17 251	15 385	17 251	15 226	34 390	15 385	32 390	15 226	23 251
	40 388		21 253		21 253		22 253		22 253		36 388		33 388		25 253
RNK REC=	0.5991	RNK REC=	0.9315	RNK REC=	0.9315	RNK REC=	0.9189	RNK REC=	0.9189	RNK REC=	0.6869	RNK REC=	0.7351	RNK REC=	0.7598
LOG PRE=	0.8595	LOG PRE=	0.9794	LOG PRE=	0.9794	LOG PRE=	0.9760	LOG PRE=	0.9760	LOG PRE=	0.9084	LOG PRE=	0.9251	LOG PRE=	0.9179
NOR REC=	0.9854	NOR REC=	0.9984	NOR REC=	0.9984	NOR REC=	0.9981	NOR REC=	0.9981	NOR REC=	0.9900	NOR REC=	0.9921	NOR REC=	0.9931
NOR PRE=	0.9230	NOR PRE=	0.9901	NOR PRE=	0.9901	NOR PRE=	0.9884	NOR PRE=	0.9884	NOR PRE=	0.9524	NOR PRE=	0.9618	NOR PRE=	0.9578
OVERALL=	1.4587	OVERALL=	1.9109	OVERALL=	1.9109	OVERALL=	1.8949	OVERALL=	1.8949	OVERALL=	1.5952	OVERALL=	1.6602	OVERALL=	1.6777
NOR OVR=	1.8498	NOR OVR=	1.9820	NOR OVR=	1.9820	NOR OVR=	1.9788	NOR OVR=	1.9788	NOR OVR=	1.9026	NOR OVR=	1.9224	NOR OVR=	1.9233

HARRIS TWO STAT PHRASE HIER DOWN		HARRIS TWO SYNTAX PHR HIERARCHY UP		HARRIS TWO SYNTAX PHR HIER DOWN		HARRIS TWO HIERARCHY UP HIER DOWN	
TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT	TOP 15	RELEVANT
1 384	1 384	1 384	1 384	1 384	1 384	1 384	1 384
2 360	2 360	2 103	2 103	2 360	2 360	2 103	2 103
3 102	3 102	3 360	3 360	3 102	3 102	3 360	3 360
4 200	4 200	4 392	4 392	4 200	4 200	4 392	4 392
5 386	5 386	5 200	5 200	5 386	5 386	5 102	5 102
6 392	6 392	6 386	6 386	6 392	6 392	6 200	6 200
7 103	7 103	7 085	7 085	7 085	7 085	7 386	7 386
8 C85	8 085	8 387	8 387	8 103	8 103	8 C85	8 085
9 387	9 387	9 234	11 358	9 387	9 387	9 234	13 358
10 169	12 358	10 192	13 390	10 169	12 358	10 169	14 387
11 192	14 390	11 358	14 102	11 192	13 390	11 192	17 202
12 358	15 202	12 128	15 388	12 358	15 388	12 128	19 251
13 229	16 388	13 390	17 202	13 390	16 202	13 358	22 253
14 390	17 251	14 102	19 385	14 229	17 385	14 387	31 385
15 202	19 385	15 388	24 251	15 388	18 251	15 229	33 390
	22 253		25 253		22 253		35 388
RNK REC=	0.8500	RNK REC=	0.7816	RNK REC=	0.8608	RNK REC=	0.6154
LOG PRE=	0.9503	LOG PRE=	0.9269	LOG PRE=	0.9541	LOG PRE=	0.8712
NOR REC=	0.9961	NOR REC=	0.9939	NOR REC=	0.9965	NOR REC=	0.9863
NOR PRE=	0.9754	NOR PRE=	0.9628	NOR PRE=	0.9773	NOR PRE=	0.9303
OVERALL=	1.8003	OVERALL=	1.7086	OVERALL=	1.8149	OVERALL=	1.4866
NOR OVR=	1.9561	NOR OVR=	1.9323	NOR OVR=	1.9597	NOR OVR=	1.8620

Figure 1 (Continued)

Figure 2 shows the summary output at the end of a MORVAL run. Again, the names of the runs merged are listed first. Directly beneath them is the recall-precision table. The recall is listed in the left column at intervals of 0.1; the corresponding precision is printed in the right column. These tables are averages over the method in question for all requests. Recall-precision graphs for individual requests are too erratic and depend too heavily on the number of relevant documents to be of much use. The averages of the first four evaluation measures are listed below the tables.

The input deck required by subroutine MORVAL is as follows:

(1) Monitor control cards and program deck;

(2) * DATA control card for FMS;

(3) An optional control card with SIZE in columns 1-4 and a decimal integer in columns 6-8. This specifies the total number of documents in the collection. If omitted, the collection is assumed to contain 405 documents;

(4) An optional control card with ORDER in columns 1-5 and an integer in column 8. This card specifies that all merges of a certain type are to be performed. The possible values for the integer in column 8 are interpreted according to Table 1. If both a SIZE and an ORDER card are used, they may be in either order;

(5) Any number (including zero) of cards specifying precise merges. These cards begin with MERGE in columns 1-5 and a blank in column 6. The names (twelve characters long) of the runs to be merged

```
AVERAGES AND QUASI-CLEVERDON GRAPHS
  HARRIS TWO          NULL THES          STAT PHRASE          STAT PHRASE          SYNTAX PHR          HIERARCHY UP          HIER DOWN          HARRIS TWO
                                         SYNTAX PHR                                                                                             NULL THES

    0.1 0.9167          0.1 0.9350          0.1 0.9552          0.1 0.9558          0.1 0.9384          0.1 0.8724          0.1 0.8983          0.1 0.9471
    0.2 0.8913          0.2 0.8799          0.2 0.8709          0.2 0.8900          0.2 0.8379          0.2 0.7542          0.2 0.8648          0.2 0.8961
    0.3 0.8757          0.3 0.8295          0.3 0.8070          0.3 0.8345          0.3 0.7765          0.3 0.7137          0.3 0.8513          0.3 0.8665
    0.4 0.8428          0.4 0.7878          0.4 0.7697          0.4 0.7981          0.4 0.7666          0.4 0.6755          0.4 0.8224          0.4 0.8425
    0.5 0.7779          0.5 0.7592          0.5 0.7423          0.5 0.7692          0.5 0.7352          0.5 0.6538          0.5 0.7792          0.5 0.7935
    0.6 0.7096          0.6 0.6665          0.6 0.7211          0.6 0.7417          0.6 0.6790          0.6 0.5868          0.6 0.6748          0.6 0.7405
    0.7 0.6002          0.7 0.5995          0.7 0.6805          0.7 0.7060          0.7 0.6253          0.7 0.5178          0.7 0.6070          0.7 0.6531
    0.8 0.5345          0.8 0.5503          0.8 0.6311          0.8 0.6519          0.8 0.5645          0.8 0.4451          0.8 0.5378          0.8 0.5429
    0.9 0.4628          0.9 0.4793          0.9 0.5292          0.9 0.5457          0.9 0.4667          0.9 0.3537          0.9 0.4390          0.9 0.4853
    1.0 0.4059          1.0 0.3411          1.0 0.4272          1.0 0.4412          1.0 0.3853          1.0 0.2979          1.0 0.3189          1.0 0.4151
RNK REC= 0.5167 RNK REC= 0.4666 RNK REC= 0.5605 RNK REC= 0.5748 RNK REC= 0.4959 RNK REC= 0.4225 RNK REC= 0.4700 RNK REC= 0.5318
LOG PRE= 0.7418 LOG PRE= 0.7260 LOG PRE= 0.7714 LOG PRE= 0.7820 LOG PRE= 0.7375 LOG PRE= 0.6376 LOG PRE= 0.7125 LOG PRE= 0.7523
NOR REC= 0.9699 NOR REC= 0.9592 NOR REC= 0.9730 NOR REC= 0.9740 NOR REC= 0.9594 NOR REC= 0.9457 NOR REC= 0.9661 NOR REC= 0.9693
NOR PRE= 0.8757 NOR PRE= 0.8580 NOR PRE= 0.8818 NOR PRE= 0.8904 NOR PRE= 0.8533 NCR PRE= 0.7981 NOR PRE= 0.8632 NOR PRE= 0.8786

  HARRIS TWO          HARRIS TWO          HARRIS TWO          HARRIS TWO          STAT PHRASE          STAT PHRASE          SYNTAX PHR          SYNTAX PHR
  STAT PHRASE         SYNTAX PHR          HIERARCHY UP        HIER DOWN           HIERARCHY UP         HIER DOWN            HIERARCHY UP        HIER DOWN

    0.1 0.9431          0.1 0.9529          0.1 0.9124          0.1 0.9230          0.1 0.9399          0.1 0.9299          0.1 0.9456          0.1 0.9174
    0.2 0.8989          0.2 0.8871          0.2 0.8816          0.2 0.8863          0.2 0.8882          0.2 0.8803          0.2 0.8642          0.2 0.8556
    0.3 0.8678          0.3 0.8271          0.3 0.8195          0.3 0.8429          0.3 0.8161          0.3 0.8240          0.3 0.7708          0.3 0.7957
    0.4 0.8227          0.4 0.8034          0.4 0.7785          0.4 0.8186          0.4 0.7530          0.4 0.7928          0.4 0.7387          0.4 0.7779
    0.5 0.7944          0.5 0.7868          0.5 0.7366          0.5 0.7715          0.5 0.7237          0.5 0.7622          0.5 0.7101          0.5 0.7629
    0.6 0.7515          0.6 0.7401          0.6 0.6894          0.6 0.6885          0.6 0.6762          0.6 0.7444          0.6 0.6698          0.6 0.7304
    0.7 0.7134          0.7 0.6964          0.7 0.6534          0.7 0.6037          0.7 0.6320          0.7 0.6987          0.7 0.6201          0.7 0.6875
    0.8 0.6235          0.8 0.5917          0.8 0.5659          0.8 0.5457          0.8 0.5920          0.8 0.6378          0.8 0.5595          0.8 0.6317
    0.9 0.5288          0.9 0.4943          0.9 0.4348          0.9 0.4626          0.9 0.5338          0.9 0.5537          0.9 0.4931          0.9 0.5291
    1.0 0.4466          1.0 0.4188          1.0 0.3850          1.0 0.3846          1.0 0.4434          1.0 0.4544          1.0 0.3900          1.0 0.4159
RNK REC= 0.5732 RNK REC= 0.5325 RNK REC= 0.4905 RNK REC= 0.5140 RNK REC= 0.5592 RNK REC= 0.5824 RNK REC= 0.5141 RNK REC= 0.5471
LOG PRE= 0.7809 LOG PRE= 0.7572 LOG PRE= 0.7244 LOG PRE= 0.7403 LOG PRE= 0.7621 LOG PRE= 0.7805 LOG PRE= 0.7413 LOG PRE= 0.7641
NOR REC= 0.9751 NOR REC= 0.9711 NOR REC= 0.9658 NOR REC= 0.9694 NOR REC= 0.9726 NOR REC= 0.9740 NOR REC= 0.9682 NOR REC= 0.9713
NOR PRE= 0.8963 NOR PRE= 0.8842 NOR PRE= 0.8608 NOR PRE= 0.8740 NOR PRE= 0.8838 NOR PRE= 0.8931 NOR PRE= 0.8687 NOR PRE= 0.8806
```

Sample MORVAL Output of Averages over Requests

Figure 2

HIERARCHY UP HIER DOWN	NULL THES STAT PHRASE	NULL THES SYNTAX PHR	HARRIS TWO NULL THES STAT PHRASE	HARRIS TWO NULL THES SYNTAX PHR	HARRIS TWO NULL THES HIERARCHY UP	HARRIS TWO NULL THES HIER DOWN	HARRIS TWO STAT PHRASE HIERARCHY UP
0.1 0.8764	0.1 0.9812	0.1 0.9812	0.1 0.9583	0.1 0.9583	0.1 0.9458	0.1 0.9583	0.1 0.9431
0.2 0.7926	0.2 0.9333	0.2 0.9062	0.2 0.9250	0.2 0.9162	0.2 0.8938	0.2 0.9108	0.2 0.8917
0.3 0.7677	0.3 0.9070	0.3 0.8750	0.3 0.8879	0.3 0.8688	0.3 0.8350	0.3 0.8590	0.3 0.8348
0.4 0.7599	0.4 0.8502	0.4 0.8370	0.4 0.8552	0.4 0.8305	0.4 0.8137	0.4 0.8330	0.4 0.7767
0.5 0.7179	0.5 0.7907	0.5 0.7847	0.5 0.8202	0.5 0.8074	0.5 0.7901	0.5 0.7876	0.5 0.7651
0.6 0.6665	0.6 0.7338	0.6 0.7232	0.6 0.7784	0.6 0.7769	0.6 0.7274	0.6 0.7387	0.6 0.7237
0.7 0.5945	0.7 0.6696	0.7 0.6466	0.7 0.7037	0.7 0.6940	0.7 0.6215	0.7 0.6491	0.7 0.6639
0.8 0.4873	0.8 0.6123	0.8 0.6006	0.8 0.6033	0.8 0.5984	0.8 0.5371	0.8 0.5499	0.8 0.6016
0.9 0.3868	0.9 0.5401	0.9 0.5140	0.9 0.5312	0.9 0.5167	0.9 0.4698	0.9 0.4837	0.9 0.5176
1.0 0.2994	1.0 0.4696	1.0 0.4346	1.0 0.4567	1.0 0.4376	1.0 0.4060	1.0 0.4152	1.0 0.4364
RNK REC= 0.4323	RNK REC= 0.5878	RNK REC= 0.5537	RNK REC= 0.5784	RNK REC= 0.5545	RNK REC= 0.5173	RNK REC= 0.5323	RNK REC= 0.5568
LOG PRE= 0.6699	LOG PRE= 0.7877	LOG PRE= 0.7673	LOG PRE= 0.7832	LOG PRE= 0.7699	LOG PRE= 0.7460	LOG PRE= 0.7545	LOG PRE= 0.7683
NOR REC= 0.9658	NOR REC= 0.9735	NOR REC= 0.9708	NOR REC= 0.9743	NOR REC= 0.9714	NOR REC= 0.9668	NOR REC= 0.9705	NOR REC= 0.9736
NOR PRE= 0.8448	NOR PRE= 0.9009	NOR PRE= 0.8891	NOR PRE= 0.9004	NOR PRE= 0.8911	NOR PRE= 0.8712	NOR PRE= 0.8815	NOR PRE= 0.8897

HARRIS TWO STAT PHRASE HIER DOWN	HARRIS TWO SYNTAX PHR HIERARCHY UP	HARRIS TWO SYNTAX PHR HIER DOWN	HARRIS TWO HIERARCHY UP HIER DOWN
0.1 0.9343	0.1 0.9608	0.1 0.9529	0.1 0.9152
0.2 0.8816	0.2 0.8826	0.2 0.8676	0.2 0.8713
0.3 0.8444	0.3 0.8020	0.3 0.8094	0.3 0.7957
0.4 0.8016	0.4 0.7614	0.4 0.7858	0.4 0.7717
0.5 0.7776	0.5 0.7456	0.5 0.7745	0.5 0.7496
0.6 0.7466	0.6 0.7166	0.6 0.7380	0.6 0.6811
0.7 0.7088	0.7 0.6548	0.7 0.6915	0.7 0.5982
0.8 0.6161	0.8 0.5763	0.8 0.5997	0.8 0.5156
0.9 0.5303	0.9 0.4907	0.9 0.5048	0.9 0.4344
1.0 0.4382	1.0 0.4074	1.0 0.4057	1.0 0.3742
RNK REC= 0.5696	RNK REC= 0.5186	RNK REC= 0.5335	RNK REC= 0.4941
LOG PRE= 0.7759	LOG PRE= 0.7463	LOG PRE= 0.7550	LOG PRE= 0.7256
NOR REC= 0.9739	NOR REC= 0.9695	NOR REC= 0.9712	NOR REC= 0.9679
NOR PRE= 0.8931	NOR PRE= 0.8758	NOR PRE= 0.8813	NOR PRE= 0.8637

Figure 2 (Continued)

Integer	Interpretation
0	Ignore this card.
1	Perform all one-way merges (single methods).
2	Perform all two-way merges (pairs of methods).
3	Perform all three-way merges (triplets of methods).
4	Perform all one and two-way merges.
5	Perform all one and three-way merges.
6	Perform all two and three-way merges.
7	Perform all one, two, and three-way merges.
8	Perform the highest order possible merge; that is, merge all runs.
9	Perform all merges of all orders.

Interpretation of the ORDER Card

TABLE 1

are listed five to a card, using the leftmost columns first. If more than five runs are being merged, the list of runs is continued on cards with PLUS in columns 1-4.

(6) A card naming a processing run. It contains a * in column 1 and the twelve character name of the run in columns 2-13. This name is used to identify the run on the printed output and to interpret the MERGE control cards;

(7) The data deck for this processing method. This deck is submitted exactly as punched by the evaluation link of SMART. It contains for each request the document rank list, the request name, and the list of relevant documents. Specifically, the document rank list is punched in four-column fields, 18 per card. The document identifiers are three characters long and occupy the rightmost three characters of each field. The list terminates with a document named with three blanks or three zeros. The next card contains the number of relevant documents in column 1-3 and the name of the request in columns 6-17. The list of relevant documents follows in the same format as the document rank list except that no termination sentinel is needed since the number of relevant documents is known;

(8) Steps 6 and 7 are repeated for every run. The runs may be in any order and the requests may be in any order within each run. All the requests for one run must be together, however; and

(9) The deck ends with a card containing *FINISH in columns 1-7.

Current size limits for MORVAL are:

(1) Maximum of 50 requests and 50 relevant documents for each request;

(2) Maximum of ten runs;

(3) Maximum of 31 merges. The number of merges generated by each of the options on an ORDER card is shown in Table 2;

(4) Maximum of one thousand relevant documents for all requests; and

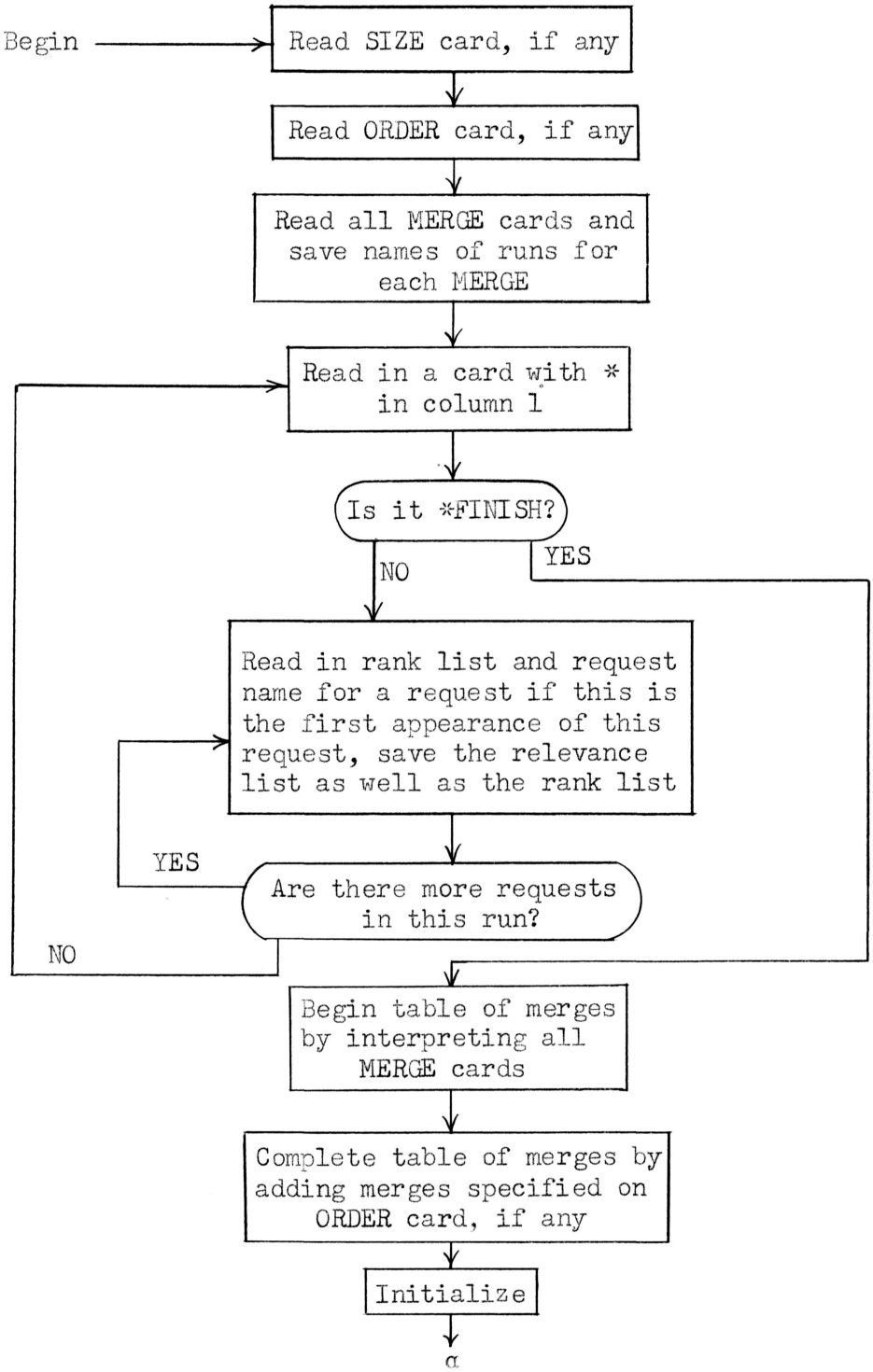
(5) Maximum of 20 thousand occurrences of documents in input rank lists (this represents about one box of data cards).

Order Number	Number of Runs									
	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10
2	.	1	3	6	10	15	21	28	36	45
3	.	.	1	4	10	20	35	56	84	120
4	.	3	6	10	15	21	28	36	45	55
5	.	.	4	8	15	26	42	64	93	130
6	.	.	4	10	20	35	56	84	120	165
7	.	.	7	14	25	41	63	92	129	175
8	1	1	1	1	1	1	1	1	1	1
9	1	3	7	15	31	63	127	255	511	1023

Number of Merges Specified by ORDER Cards
(31 is the Maximum Number of Merges Allowed)

TABLE 2

A flowchart for MORVAL is shown in Fig. 3. MORVAL is written entirely in FORTRAN II.



Flowchart for MORVAL

Figure 3

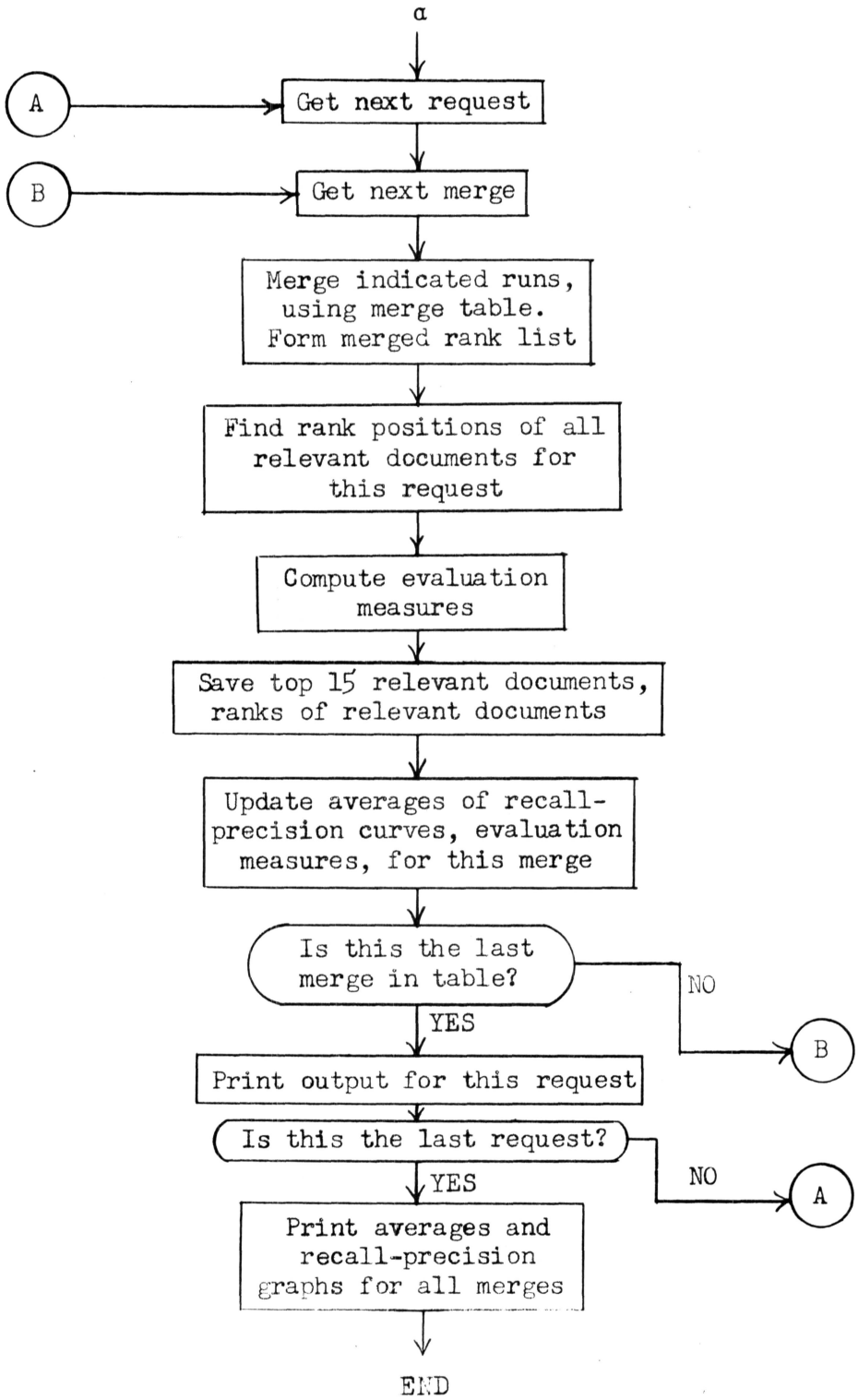


Figure 3 (continued)

REFERENCES

1. Lesk, Michael., "SMART Operating Instructions," Information Storage and Retrieval, Report No. ISR-8, Sec. II, The Computation Laboratory of Harvard University (December 1964).

2. Salton, Gerard., "The Evaluation of Automatic Retrieval Procedures," Information Storage and Retrieval, Report No. ISR-8, Sec. IV, The Computation Laboratory of Harvard University (December 1964).

3. Lesk, Michael., "The SMART System – Typical Processing Sequence," Information Storage and Retrieval, Report No. ISR-8, Sec. I, The Computation Laboratory of Harvard University (December 1964).