AD-759 714

NOTES ON A PROBLEM INVOLVING PERMUTATIONS AS SUBSEQUENCES

Malcolm Newey

Stanford University

Prepared for:

National Aeronautics and Space Administration
Advanced Research Projects Agency

March 1973

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

STAN-CS-73-340

AD 759714

NOTES ON A PROBLEM INVOLVING PERMUTATIONS AS SUBSEQUENCES

BY

MALCOLM NEWEY

DDC
MAY 15 1973

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

SUPPORTED BY

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
CONTRACT NSR 05-020-500
AND
ADVANCED RESEARCH PROJECTS AGENCY

ARPA ORDER NO. 457

MARCH 1973

COMPUTER SCIENCE DEPARTMENT
School of Humanities and Sciences
STANFORD UNIVERSITY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

STANFORD ARTIFICIAL INTELLIGENCE LABORATORY MARCH 1973
MEMO AIM-190
COMPUTER SCIENCE DEPARTMENT
REPORT CS-340

NOTES ON A PROBLEM INVOLVING PERMUTATIONS AS SUBSEQUENCES.

by

Malcolm Newey.

ABSTRACT:
The problem (attributed to R. M. Karp by Knuth (see #36 of [1])) is to describe the sequences of minimum length which contain, as subsequences, all the permutations of an alphabet of n symbols. This paper catalogs some of the easy observations on the problem and proves that the minimum lengths for n=5, n=6 & n=7 are 19, 28 and 39 respectively. Also presented is a construction which yields (for n>2) many appropriate sequences of length n^2-2n+4 so giving an upper bound on length of minimum strings which matches exactly all known values.

This research was supported in part by the Advanced Research Projects Agency of the Office of the Secretary of Defence under Contract SD-183 and in part by the National Aeronautics and Space Administration under Contract NSR 05-020-500.

The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency, the National Aeronautics and Space Administration, or the U.S. Government.

Reproduced in the USA. Available from the National Technical Information Service, Springfield, Virginia 22151. Price: full size copy $3.00; microfiche copy $0.95.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Stanford University, Computer Science Department Stanford, California 94305	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

NOTES ON A PROBLEM INVOLVING PERMUTATIONS AS SUBSEQUENCES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

technical, March 1973

5. AUTHOR(S) (First name, middle initial, last name)

Malcolm Newey

6. REPORT DATE March 1973	7a. TOTAL NO. OF PAGES 29	7b. NO. OF REFS 1
8a. CONTRACT OR GRANT NO. SD-183 and NSR 05-020-500	9a. ORIGINATOR'S REPORT NUMBER(S) STAN-CS-73-340	
b. PROJECT NO. c. d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AIM-190	

10. DISTRIBUTION STATEMENT

Distribution Unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The problem (attributed to R. M. Karp by Knuth is to describe the sequences of minimum length which contain, as subsequences, all the permutations of an alphabet of n symbols. This paper catalogs some of the easy observations on the problem and proves that the minimum lengths for n=5, n=6 and n=7 are 19, 28 and 39 respectively. Also presented is a construction which yields (for n>2) many appropriate sequences of length n^2-2n+4 so giving an upper bound on length of minimum strings which matches exactly all known values.

(n sup 2)

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6801

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

DD FORM 1 NOV 65 1473 (BACK)

(PAGE 2)

Security Classification

1 NOTATION.

a) Let S be a sequence of symbols. $|S|$ will be used to denote the total number of symbols in S and so we observe, for example, $|x\ y\ x\ z| = 4$.

b) We say $x \subset y$ in the case where x is a subsequence of y and we say "x is equivalent to y" if x can be obtained from y by a simple change of alphabet; we denote this equivalence by '$\equiv$'.
(e.g. $x\ y \subset x\ y\ y\ x$, $x\ y\ z\ x \equiv 1\ 2\ 3\ 1$)

c) P(A) is used to denote the set of sequences which are permutations of an alphabet A. Cardinality of P(A) will be $(|A|)!$. Also, $P'(A,n)$ is is the set of permutations of all sub-alphabets of A of size n (where $n \le |A|$). Clearly, $P(A) = P'(A,|A|)$.

d) If A is an alphabet then $Q(A) = \{ x \mid x \in A' \wedge \forall y.(y \in P(A) \supset y \subset x)\}$ where A′ is the set of sequences over alphabet A. For example, abcacba $\in$ Q(abc). Also, $Q'(A,n)$ is taken to be the set $\{ x \mid x \in A' \wedge \forall y.(y \in P(A,n) \supset y \subset x)\}$.
So, for example, zyxwxyz $\in$ Q′(wxyz,2) .

e) Now, the LENGTHS of the shortest sequences in Q(A) and Q′(A,n) depend only on the SIZE of the alphabet A. Hence, take M(n) to be the length of the shortest sequence in Q(1 2 3...n) and M′(n,m) to be the length of the shortest sequence in Q′(1 2 3 ... n, m) .
So, for example, $M(1)=1$, $M(2)=3$ and $M'(n,1)=n$.

f) S(n) denotes the n-th symbol of sequence S.
S(n:m) denotes that contiguous subsequence of sequence S which is the symbols from position number n in S to position number m.
#(S,x) denotes the number of ocurrences of the symbol x in sequence S.

g) "CPAF X" is just an abbreviation for "Consider the Permutations of the current Alphabet of the Form X" . The greek letters which appear in X denote arbitrary sequences of symbols.
For example, if the alphabet under discussion were abcde, the command "CPAF bαc" would mean "Consider Permutations of abcde which start with b and end with c".

2 SOME EASY OBSERVATIONS.

2.1 $M(1)=1$.

2.2 $M(2)=3$.

2.3 $M(3)=7$.

2.4 $M'(n,1)=n$.

2.5 $M'(n,2)=(2n-1)$ can be seen as follows:

$M'(n,2) \le 2n-1$ since if A is an alphabet of length n, then the sequence AA(2:2n) is a member of $Q'(A,2)$.
$M'(n.2) \ge 2n-1$ since if A is an alphabet of size n, S is a member of $Q'(A,2)$ and $|S|<2n-1$ then at least two of the symbols of A (x and y, say) only appear once in S; hence 1 of the sequences 'xy' and 'yx' are not subsequences of S.

2.6 $M'(n,m) \ge (m.(2n-m+1)/2)$ ($n \ge m$, of course)

This result is more easily remembered as

$$M'(n,m) \ge n + n-1 + n-2 + \ldots\ldots + n-m+1 .$$

Suppose A is an alphabet of size n and S is a sequence from $Q'(A,m)$ of minimum length (i.e. $|S|=M'(n,m)$). It is noted in (2.4) that $M'(n,1)=n$ so take $m \ge 2$. Segment S as TxU where the sequences T,U and the symbol x are chosen so that x does not appear in T but all the other symbols of A do. Clearly, $|T| \ge (n-1)$. Now note that all permutations of subalphabets of A of size m which start with x are subsequences of xU. Hence all permutations of subalphabets of A\x of size (m-1) are subsequences of U (A\x is A without x and $|A\backslash x|=(n-1)$). $|U| \ge M'(n-1,m-1)$, therefore, and so $M'(n,m)$ (which is simply $|S|$) is at least $(n-1) + 1 + M'(n-1,m-1)$. This recurrence relation is readily solved to give the result.

2.7 $M(n) \ge (n.(n+1)/2)$.

Simple corollary of 2.6 using $M(n)=M'(n,n)$.

2.8 $M'(n,m) \le (m.(n-1)+1)$

Given an alphabet, A, of size n , the following construction gives an element of Q'(A,m) of length $m*(n-1)+1$:-
Generate m permutations of the alphabet A1, A2, A3, ... Am such that A1(n)=A2(1), A2(n)=A2(1) etc. Now, B = A1 A2(2:n) A3(2:n)...Am(2:n) is in Q'(A,m) since if C is any permutation of any subalphabet of A of size m, C(j) is either in the j-th component of B or IS the last symbol of the (j-1)th component (for j>1).

2.9 $M(n) \le (n.n-n+1)$

A simple corollary of 2.8.

2.10 $M'(n,3) = (3n-2)$ $(n \ge 3)$.

From 2.6 we get $M'(n,3) \ge (3n-3)$.
From 2.8 we get $M'(n,3) \le (3n-2)$.
Suppose the lower value is obtained for an alphabet A ($|A|=n$) and S is a sequence of length 3n-3 which is in Q'(n,3). Now no symbol can appear only once in S for then we would have $|S| \ge (2.M(n-1,2)+1) = (4n-5)$ which is a contradiction for $n \ge 3$. Hence there must be at least 3 symbols which occur just 2 times each for a total of 6 times. However M(3)=7 so there must be some permutation of these three symbols which is not a subsequence of S. This contradiction gives us the result.

2.11 Members of Q(1 2 3) of Length 7.

The following is an exhaustive list of minimum solutions for a 3 symbol alphabet. We consider, of course, only equivalence classes (with respect to the operator ■).

1 2 3 1 2 1 3	1 2 3 1 2 3 1	1 2 3 1 3 2 1
1 2 3 2 1 2 3	1 2 3 2 1 3 2	
1 2 1 3 1 2 1	1 2 1 3 2 1 2	

2.12 $\forall S \in Q(A).\ \exists a \in A.\ \#(S,a) \ge |A|$.

Use induction on the alphabet size. The case $|A|=1$ is trivial so suppose the result holds for all alphabets of size less than n, $|A|=n$ and $S \in Q(A)$. Segment S as TxU where sequences T,U and symbol x are chosen so that x does not appear in T but every other symbol of A does. Use A\x to denote A minus symbol x, and we get $U \in Q(A \backslash x)$. Now $|A \backslash x| = n-1$ and so we can find y such that $\#(U,y) \ge (n-1)$. Clearly $\#(S,y) \ge n$.

2.13 $\forall S \epsilon Q'(A,m).\ Card(\{ a \mid a \epsilon A \wedge \#(S,a) \geq m \}) \geq (n-m+1)$

Let A be any alphabet, m be any integer such that $|A| \geq m$ and S be some member of $Q'(A,m)$. Select sequence B - a permutation of A such that the symbols of B are in order of decreasing frequency in S.
Now take sequence S′ to be the sequence formed by deleting those symbols from S which are in B(1:n-m). S′ is a member of Q(B(n-m+1:n)) and so some symbol must appear at least m times in S′ and hence in S.
Therefore, $\#(S,B(1)) \geq \#(S,B(2)) \geq \ldots\ldots \geq \#(S,B(n-m+1)) \geq m$ which gives the quoted result.

2.14 $M'(n,m) \geq m(n-m)+M(m)$

A corollary of 2.13 .

2.15 $M(4)=12$.

Take A to be the alphabet (sequence) 1 2 3 4 .

1 2 3 4 1 2 3 1 4 2 1 3 ϵ Q(A) and so $M(4) \leq 12$.

Suppose $S \epsilon Q(A)$ and $|S|<12$.
Compute the least integer j such that S(1:j) contains each symbol of A. Note $j \geq 4$ and S(j) is not in S(1:j-1).
Considering permutations of A which start with S(j), we get that $|S| \geq 3 + \#(S,S(j)) + M(3) = 10 + \#(S,S(j))$.
Using $|S|<12$ we get $j=4$ and $\#(S,S(j))=1$.
Therefore, S(4) appears only at position 4 of S. Now consider the permutations of A that end with S(4) and get that $4 \geq M(3)$ which is a contradiction.

From this contradiction we see that $M(4) \geq 12$.

2.16 $\forall A.\ \forall x \epsilon A.\ \exists S \epsilon Q(A).\ \#(S,x)=1$

Suppose we are given an alphabet A and x is some symbol of A. We take the subalphabet A\x and find some member T from Q(A/x).
Clearly $T x T \epsilon Q(A)$ and also $\#(TxT,x)=1$.
This is quite a useful result to keep in mind when pondering what properties members of Q(A) might have.

3 M(5)=19.

Take A to be the alphabet (sequence) 1 2 3 4 5 .

i) 1 2 3 4 5 1 2 3 4 1 5 2 3 1 4 5 2 1 3 ϵ Q(A)
so we have $M(5)\le 19$.

ii) Suppose $S\epsilon Q(A)$ and $|S|<19$.
Break up S as $T\,y\,U$ (where T and U are segments of S and y is a single symbol) such that Ty is the shortest initial segment of S which is in $Q'(A,2)$ so $|Ty|\ge M'(5,2)=9$.
Choose x in T such that xy is not a subsequence of T (this is possible otherwise S was not segmented as prescribed).

Considering members of P(A) starting with xy, get

$$|S| \ge 9 + M(3) + \#(U,x) + \#(U,y) = 16 + \#(U,x) + \#(U,y).$$

Now, supposing x does not appear in U, consider subsequences of S that end with x and derive the contradiction

$$|S|\ge M(4)+2+M(3)=21.$$

Conclude $\#(U,x)\ge 1$ (and similarly $\#(U,y)\ge 1$).

Reconciling inequalities, we get $\#(U,x)=1$, $\#(U,y)=1$, $|T|=8$, $|U|=9$ and $|S|=18$.

In U, x and y appear just once each and so one sequence of xy and yx , call it Z, is not a subsequence of U.
Consider, then, permutations of A of the form αZ and get

$$|T| \ge M(3) + \#(T,x) + \#(T,y) \ge 9$$ -- a contradiction!

We therefore conclude that $M(5)\ge 19$.

iii) From i) and ii) deduce $M(5)=19$.

4 M(6)=28 and M(7)=39.

i) Take A to be the alphabet (sequence) 1 2 3 4 5 6 .

1 2 3 4 5 6 1 2 3 4 5 1 6 2 3 4 1 5 6 2 3 1 4 5 6 2 1 3
is in Q(A) so we have M(6)≤28 .

The proof of M(6)≥28 is given as Appendix 1 because it is long and uninformative.

These two facts give the result M(6)=28.

ii) Take A to be the alphabet 1 2 3 4 5 6 7.

1 2 3 4 5 6 7 1 2 3 4 5 6 1 7 2 3 4 5
1 6 7 2 3 4 1 5 6 7 2 3 1 4 5 6 7 2 1 3
is in Q(A) so we have M(7) ≤ 39 .

M(7)≥39 (proved as appendix 2) and so we have M(7)=39.

5 Minimum Length Solutions for Alphabets of Size 4.

Let A be the alphabet a b c d .
We wish to enumerate the equivalence classes in Q(A) of the minimum length (ie 12). Suppose $S \in Q(A)$ and $|S|=12$.

Lemma: $\forall p \in A.\ \#(S,p) \geq 2$
$p \in A \wedge \#(S,p)=0$ is absurd.
Suppose $p \in A \wedge \#(S,p)=1$ We have that S has the form UpV.
CPAF αp to get $|U| \geq M(3)=7$; CPAF $p\alpha$ to get $|V| \geq M(3)=7$.
We immediately have the contradiction $|S| = |UpV| \geq 15$.

Lemma: $\exists p.\ \#(S,p)=2$
Suppose not. In view of above lemma, $\forall p \in A.\ \#(S,p) \geq 3$ which is a violation of the result 2.12 (page 3).

Supposing $\#(S,p)=2$, choose T,U,V such that S = TpUpV.
CPAF $p\alpha$ to get $|UV| \geq 7$; CPAF αp to get $|TU| \geq 7$.
Now $|U| = |U|+(|S|-12) = (|U|+|T|+|U|+|V|+2)-12 \geq 4$.
Also $|T| = |S|-2-|U|-|V| \leq 3$ and similarly $|V| \leq 3$.

Suppose $|T|<3$. Thus $\exists x \in A.\ \neg(x \in T) \wedge \neg(x=p)$.
CPAF $xp\alpha$ to give $|V| \geq M(2)+\#(V,x)=3+\#(V,x)$. So $\#(V,x)=0$.
CPAF αpx to give the contradiction $|T| \geq M(2)=3$.
Hence $|T|=3$ and similarly $|V|=3$ giving $|U|=4$.

Suppose $q \in A$ and $\neg(q=p) \wedge \#(T,q)=0$.
CPAF $qp\alpha$ to get $\#(V,q)=0$. Hence by a lemma above, $\#(U,q) \geq 2$.
CPAF $q\alpha p$ to get the contradiction $|U| \geq M(2)+\#(U,q) \geq 5$.
Hence $\forall q.\ q \in A \supset (q=p \vee \#(T,q)=\#(V,q)=1)$.

From this discussion we get that there are representatives of all the equivalence classes of the form
a b c d U d V where $|U|=4$, $|V|=3$, $a \in V$, $b \in V$, $c \in V$.

CPAF αd we get abcU is in Q(a b c) and is of min. length.
Using result (2.11) we get 5 possibilities for U; namely:
(1) abac (2) abca (3) acba (4) babc (5) bacb.

Similarly UV is in Q(a b c) and is of minimum length.
Performing a small amount of hand checking and using 2.11 again we get that there are exactly 9 equivalence classes:-

abcd abca dbac	abcd acba dbca	abcd bacb dabc
abcd abca dbca	abcd acba dcab	abcd bacb dacb
abcd abca dcba	abcd acba dcba	abcd bacb dcab

6. An $n^2 -2n +4$ Construction for Alphabet of size n.

Given an alphabet sequence, A, of length at least three, it is asserted that the following recipe gives a sequence in Q(A).

```
Set the sequence variable  B ← A(2:n);

Write(A);
DO (n-2) TIMES { Write(A(1)); Write( B(1:n-2) );
                 B ← (B(n-1) B(1:n-2));  };
Write(A(1));  Write(B(1));
```

The total number of symbols written = n+ (n-2)*(1+n-2) +2
= n^2-2n+4 .

We now verify that the sequence produced is indeed in Q(A).

First note that the operation " B ← B(n-1)B(1:n-2) " simply rotates the sequence of n-1 symbols in B.

Next note that the first symbol of A (we will call it a) is written exactly n times. Letting C be the result of the above construction, we segment C as follows:
C = aJaKaLa...aYaZab where the (n-1) sequences J,K,L,...Y,Z do not contain the symbol a.
For convenience we will use call J,K,L,.....Y,Z units and will refer to them as U[1], U[2], ... U[n-1].

Now J contains all symbols A(2:n) but K,L,...Y,Z each contain just n-2 of the symbols of A(2:n). However the symbol of A(2:n) that does not appear in some unit U[k] is both the last symbol of U[k-1] and follows the a that follows U[k] in C.

Let P be a permutation of A. We will show that P must be a subsequence of C.

Suppose a appears in the jth position of P. We first show that the string P(1:j) (simply a if j=1) can be matched to the the head of C aJaKaL...U[j-1]a . Trivially true if j=1.
If j>1 then P(1) is in J, clearly. Also if j>k>1 then P(k) can be matched to U[k] if it is in that unit or else the last symbol of U[k-1].
Similarly the n-j symbols of P(j+1:n) can be matched to U[j]aU[j+1]a...aU[n-1]ab . If j<k≤n then P[k] will either match something in U[k-1] or the symbol which follows the a which follows U[k-1].

7. A More General n^2-2n+4 Construction.

It is asserted that the following algorithm, regardless of which internal choices are made, also produces a member of Q(A) of length n^2-2n+4. The proof of membership in Q(A) follows by the same method used in proving the validity of the simpler 'program'. It is also readily seen that the previous construction is a special case of this more general one.

```
SUBROUTINE SR1:
        Write the symbol [x];
        Write the symbol [y];

SUBROUTINE SR2:
        SR1;
        Write in any order the [n-3] symbols of A which do not include
        [x] or [y] or [z].
        DO   y←z   AND  set z to the last symbol written.

SUBROUTINE SR3:
        DO SR2 [n-2] TIMES;
        SR1;

SUBROUTINE SR4:
        DO SR2 [n-3] TIMES;
        SR1;
        Write in any order the [n-2] symbols of A which are not [x],[y];
        Write the symbol [x];

MAIN ROUTINE:
        Write down the alphabet (A);
        DO EITHER  { x←A(1); y ← any symbol of A(2:n-1); z←A(n); }
        OR  { x ← A(2); y ← A(1); z ← A(n);  };
        DO EITHER  SR3  OR  SR4;
```

SYMBOL COUNT.

If M symbols are written each time a certain routine is obeyed then we say that the SYMBOL COUNT for that routine is M.

Symbol Count for SR1 = 2 ;

Symbol Count for SR2 = n-1 ;

Symbol Count for SR3 = $(n-2)*(n-1)+2 = n^2 -3n +4$;

Symbol Count for SR4 = $(n-3)*(n-1)+(n+1) = n^2 -3n +4$.

Hence Symbol Count for total algorithm = $n^2 - 2n + 4$.

Note that no distinct sequences produced by this algorithm are equivalent since all such begin with a copy of the alphabet.

Note also that every sequence so produced ends with some permutation of the alphabet.

Given an alphabet A, the reversal of any sequence which is a member of Q(A) is also a member of Q(A). It should be noted that the the reverse of any sequence generated according to this construction is equivalent to some other sequence given by the construction.

8. Constructing Elements of $Q'(A,m)$.

Section 6 contained a simple construction for generating elements of $Q(A)$ (for given alphabet A of size $n>2$) which were of length n^2-2n+4 . This algorithm is now modified to generate members of $Q'(A,m)$ (where $2<m\le n$) of length $mn-2m+4$.

```
Set the sequence variable   B ← A(n-m+2:n);
Write(A);
DO  m-2  TIMES  Write( A(1:n-m+1) );
                Write( B(1:m-2) );
                B ← B(m-1)B(1:m-2);
Write( A(1:n-m+1) );
Write( B(1) );
```

The total number of symbols written is easily seen to be

$$n + (m-2)(n-m+1 + m-2) + (n-m+1) + 1 = mn-2m+4 .$$

Just as this algorithm is a modification of the one in section 6, the proof of the correctness of the construction is an extension of the previous proof.

This construction gives an upper bound on $M'(n,m)$ for $n\ge m>2$ of $mn-2m+4$ and so using this knowledge, the proposition 2.14 and the various values of $M(4)$, $M(5)$, $M(6)$ & $M(7)$ we already know, we compute the new results:-

$$M'(n,4) = 4n-4$$
$$M'(n,5) = 5n-6$$
$$M'(n,6) = 6n-8$$
$$M'(n,7) = 7n-10$$

9. Discussion.

The construction of section 7 gives many sequences of the desired length. It gives all nine equivalence classes of sequences in Q(a b c d) of length 12, 128 classes in Q(a b c d e) which may or may not be all of them, and 32,400 classes from Q(a b c d e f). It does NOT get all the sequences of Q(a b c d e f) since all the ones produced start with one copy of the alphabet however the following sequences from Q(a b c d e f):

abcdebfdcabedcfbadecbdfacebd

abcdeafdcbaedcfabdecafdbcead

(among others known) DO NOT! In fact, the second of these examples does not even end with a permutation of the alphabet.
An easy to derive lower bound on the number of classes is ((n-3)!)↑(n-1).

We now tabulate the known values of the functions M & M′ .

m	M(m)	m^2-2m+4	M′(n,m)
1	1	3	n
2	3	4	2n-1
3	7	7	3n-2
4	12	12	4n-4
5	19	19	5n-6
6	28	28	6n-8
7	39	39	7n-10

The fact that the actual values of M(n) exactly match the n^2-2n+4 figure for $2 < n \le 7$ make the construction relatively important. It also suggests the obvious conjecture that M(n) is exactly n^2-2n+4 for all $n>2$. However, there is another competing conjecture which gives exact fit at n=1,2 as well as the other known values of M(n) but is more complicated:-

$$M(n) = \begin{cases} n^2 & \text{for } n=1 \\ n^2-n+1 & \text{for } 2\le n\le 3 \\ n^2-2n+4 & \text{for } 4\le n\le 7 \\ n^2-3n+11 & \text{for } 8\le n\le 15 \\ \cdots\cdots & \\ n^2-m*n+F(m) & \text{for } 2^m \le n\le 2.2^m-1 \end{cases}$$

where $F(0)=0 \wedge F(n)=n+2*F(n-1)$.

Of course, knowing whether the value for M(8) is 51 or 52 would help by eliminating one of these postulates.

It is surprising that the best lower bound we have on $M(n)$ is $n^2/2$ since it would appear that it is of order n^2. This conjecture is readily stated formally as:-

$$\forall k.\ k<1 \supset \exists N.\ n>N \supset (M(n) > k*n^2)$$

It should be noted that just the mechanical checking of the membership of a sequence (over alphabet A) in Q(A) is quite time- consuming. A program is available in ALGOL but (although it includes some means for pruning the tree of permutations) takes a long time to check that all permutations of the alphabet are subsequences of the given sequence. The actual times on a PDP10 are 3, 17 and 60 seconds for alphabets of sizes 8, 9 & 10 respectively.

REFERENCE:

1. Chvatal,V., Klarner,D.A., Knuth,D.E., "Selected Combinatorial Research Problems", Report CS 292, Computer Science Department, Stanford University, June 1972.

APPENDIX 1. Proof of $M(6) \geq 28$.

Take A to be an alphabet of size 6 ($|A|=6$).
Moreover, suppose $S \in Q(A)$ and $|S|<28$.

Now choose sequences T, V and symbols x,y such that

a) Tx is the shortest head of S that is in $Q'(A,2)$;

b) yV is the shortest tail of S that is in $Q'(A,1)$;

Choose $w \in T$ such that $w \neq x \wedge \neg(wx \subset T)$.
We have immediately that $|T| \geq 10$, $|V| \geq 5$ and from consideration of the elements of P(A) of the forms $wx\alpha$ & αy get $|S| \geq |T|+1+M(4)$, $|S| \geq |V|+1+M(5)$, $|T| \leq 14$, $|V| \leq 7$, $|S| \geq 25$.
Hence we can segment S as the sequence TxUyV and note
$10 \leq |T| \leq 14$, $2 \leq |U| \leq 10$, $5 \leq |V| \leq 7$, $25 \leq |S| \leq 27$.

Again CPAF $wx\alpha$ and get $|UyV| \geq M(4)+2=14$. Hence (using $|S| \leq 27$) $|T| \leq 12$ and (using $|V| \leq 7$) $|U| \geq 6$. Also CPAF αy again to deduce $|TxU| \geq M(5)+1=20$. Therefore, $|S| \geq 20+1+|V| \geq 26$ and (using $|T| \leq 12$) $|U| \geq 7$. Lastly (using $|S| \leq 27$ and $|TxU| \geq 20$), $|V| \leq 6$.

Suppose $\#(U,w)=0$. Since $|yV| \leq 7$ but contains all of A, there must be 5 symbols of yV which appear just once. Therefore we choose p,q such that p,q,x,w are distinct, $\neg(pq \subset yV)$ and p,q both appear twice in T. We can do this since only one symbol of Tx can appear only once. Now CPAF αwpq to get $|T| \geq M(3) + \#(T,w) + \#(T,p) + \#(T,q) \geq 12$.
So $|T|=12$ and $\#(T,w)=1$. Segment S as LwMxUyV noting that since LwMx is in $P(A,2)$ and $\#(L,w)=0$, $|M| \geq 4$. This gives that $|L| \leq 7$ and $\#(MxU,w)=0$. $M(5,2)=9$ so we pick p,q such that $\neg(pq \subset L)$ and p,q,w distinct. Now CPAF $pqw\alpha$ to get $|yV| \geq M(3)+\#(yV,w) \geq 8$.
This contradiction gives $\#(U,w) \geq 1$.

Again CPAF $wx\alpha$ and get $|UyV| \geq M(4)+\#(UyV,w)+\#(yV,x) \geq 15$.
Use $|S| \leq 27$ to get $|T| \leq 11$ and use $|V| \leq 6$ to get $|U| \geq 8$.

Now let $t \in A$ be such that $\#(U,t)=0$. As above we choose p,q so that t,p,q are distinct, $\neg(pq \subset yV)$ and p,q both appear at least twice in T. CPAF αtpq to deduce the contradiction

$$|Tx| \geq M(3) + \#(Tx,t) + \#(Tx,p) + \#(Tx,q) \geq 12 \quad !!$$

Hence all symbols appear at least once in U.

Yet again CPAF $wx\alpha$ to get $|UyV| \geq M(4)+\#(UyV)+\#(UyV) \geq 16$.
As before deduce $|T| \leq 10$ and $|U| \geq 9$. Also CPAF αy to give $|TxU| \geq M(5)+\#(TxU,y) \geq 21$ and then $|S|=27$, $|V|=5$
We also have $|T|=10$, $|U|=10$ and $\forall t.\ t \in A \supset t \in U$.

The proof is concluded by deriving contradictions in the various possible cases of equality among w,x,y.

CASE 1. $x=y$. and so $S = TxUxV$.
We know $\#(T,x) \geq 1$ and $\#(U,x) \geq 1$ so CPAF αx and get the contradiction $21 = |TxU| \geq M(5) + \#(TxU,x) \geq 22$.

CASE 2. $x \neq y$.

CASE 2a. $w \neq y$ (i.e. w, x, y all distinct).

CPAF $wx\alpha y$ to get $|U| \geq M(3)+\#(U,w)+\#(U,x)+\#(U,y) \geq 10$

Therefore $\#(U,w) = \#(U,x) = \#(U,y) = 1$.

Now this gives that one of wx or xw, call it Z, is such that $\neg(Z \subset U)$. CPAF αZy and get $|T| \geq M(3)+\#(T,w)+\#(T,x)+\#(T,y)$

But $\#(T,w)+\#(T,y) \geq 3$ and so $|T| \geq 11$ -- contradiction!!

CASE 2b. $w=y$.

Find the first symbol of V which is not x ; call it z.

Note that since $yV \in P(A) \wedge |yV|=|A|$, z appears just once in V.

CPAF $yx\alpha z$ to deduce $|U| \geq M(3)+\#(U,y)+\#(U,x)+\#(U,z) \geq 10$.

Immediately we see $\#(U,x) = \#(U,z) = 1$ and so one of xz, zx (call it Z) is not a subsequence of U.

CPAF αZy to get $|T| \geq M(3)+\#(T,x)+\#(T,y)+\#(T,z)$.

Use $\#(T,y)+\#(T,z) \geq 3$ for the contradiction $|T| \geq 11$.

APPENDIX 2. Proof of $M(7) \geq 39$.

Take A to be an alphabet of size 7 ($|A|=7$).
Moreover, suppose $S \in Q(A)$ and $|S|<39$.

Choose sequences T,U,W and symbols a,b,c such that
- a) Ta is the shortest head of S that is in $Q'(A,1)$
- b) cW is the shortest tail of S that is in $Q'(A,1)$
- c) TaUb is the shortest head of S that is in $Q'(A,2)$

We segment S as TaUbVcW and readily prove:
$6 \leq |T| \leq 8$, $5 \leq |U| \leq 9$, $8 \leq |V| \leq 18$, $6 \leq |W| \leq 8$, $36 \leq |S| \leq 38$;
as well as $|T|+|U| \leq 15$,

Suppose for some p in A, $\#(V,p)=0$.
If p is the symbol b, $M'(6,3)+\#(TaUb,p) \geq 18 > |TaUb|$ so we can choose q,r,s such that $\mathrm{distinct}(p,q,r,s) \wedge \neg(qrs \subset TaUb)$ so that $\neg(qrsp \subset TaUbV)$. CPAF $qrsp\alpha$ we get a contradiction $|cV| \geq 4+M(3)$.
Otherwise p,b are distinct and $M'(6,3)+\#(TaU) \geq 17 \geq |TaU|$ so we rechoose q,r,s such that $\mathrm{distinct}(p,q,r,s) \wedge \neg(qrs \subset TaU)$ which means $\neg(qrsp \subset TaUbV)$. As before get a contradiction.

Lemma 1: $\forall x \in A.\ \#(V,x) \geq 1$ follows from these contradictions.

Suppose $p \in A \wedge \mathrm{distinct}(a,p)$. We know $\#(T,p) \geq 1$ and $\#(Ub,p) \geq 1$ and $\#(V,p) \geq 1$ and $\#(cW,p) \geq 1$ so conclude $\#(S,p) \geq 4$. Also we have $\#(V,a) \geq 1$ and $\#(cW,a) \geq 1$ so that $\#(S,a) \geq 3$.
We sharpen our inequalities now. CPAF $a\alpha$ to get $|T| \leq 7$, $|S| \geq 37$; CPAF $ab\alpha$ to get $|T|+|U| \leq 13$; CPAF αb to get $|W| \leq 7$. Hence
$6 \leq |T| \leq 7$, $5 \leq |U| \leq 7$, $13 \leq |V| \leq 18$, $6 \leq |W| \leq 7$, $37 \leq |S| \leq 38$.

Suppose, in fact, $\#(S,a)=3$.
We re-segment S as TaJaKaL where $\#(TJKL,a)=0$ and $L \subset W$.
There is at most one repeated symbol in T since $|Ta| \leq |A|+1$.
Let z denote this symbol if it exists else any symbol of T.
Choose p,q such that $\mathrm{distinct}(p,q,a,z) \wedge \neg(pq \subset T)$.
CPAF $pqza\alpha$ to deduce that some subsequence G of KaL belongs to Q(A1) where A1 is obtained from A by deleting p,q,a,z.
$|G| \geq M(3)=7$ so some symbol of G appears at least 3 times.
So we choose y to be such a symbol and note
$\mathrm{distinct}(a,y) \wedge \#(T,y)=1 \wedge \#(KaL,y) \geq 3$.
Now one of py and yp (call it Z) is not a subsequence of T.
CPAF $Zza\alpha$ to show we can choose x with the properties
$\mathrm{distinct}(x,y,a) \wedge \#(T,x)=1 \wedge \#(KaL) \geq 3$.

Now, one of the sequences xy and yx is not a subsequence of T (call it Y) and CFAP Yaα to get

$|KaL| \geq M(4) + \#(KaL,a) + \#(KaL,x) + \#(KaL,y) \geq 19.$

By symmetry $|TaJ| \geq 19$ to give the contradiction $|S| \geq 19+19+1$.

Lemma 2: $\forall x \in A.\ \#(S,x) \geq 4$ is immediate.

Again CPAF aα to get $|T|=6$, $|S|=38$, $\#(S,a)=4$;
Also CPAF αc to derive $|W|=6$, $|U|+|V|=23$, $\#(S,c)=4$.
Then CPAF abα to get $|VcW| \geq M(5)+\#(VcW,a)+\#(VcW,b) \geq 23$
which leads to $16 \leq |V| \leq 18$ and $5 \leq |U| \leq 7$.

Suppose that p,q are such that $\neg(pq \subset V)$. We have that $\#(TaUb,p)+\#(TaUb,q) \geq 3$. Now $|TaUb| \leq 15$ and so

$|TaUb| < M'(5,3)+\#(TaUb,p)+\#(TaUb,q)$. Hence we

choose j,k,l such that $distinct(j,k,l,p,q) \wedge \neg(jkl \subset TaUb)$.
CPAF jklpqα so $|cW| \geq M(2)+5 = 8 > |cW|$ -- a contradiction!
Thus $\forall p \in A.\ \forall q \in A.\ \#(V,p)+\#(V,q) \geq 3$.
In particular, letting z be the first symbol of cW which is not one of a,b, $\#(V,a)+\#(V,b)+\#(V,z) \geq 5$.
CPAF acαz to get $|V| \geq M(4)+\#(V,a)+\#(V,b)+\#(V,z) \geq 17$
Thus we have new bounds for U,V:- $5 \leq |U| \leq 6$, $17 \leq |V| \leq 18$.

We now choose sequence H and symbol d such that
dHcW is the shortest tail of S in Q(A).
By symmetry with the results for U we have that $5 \leq |H| \leq 6$ and so we re-segment S as TaUbGdHcW where
$|T|=6$, $5 \leq |U| \leq 6$, $10 \leq |G| \leq 12$, $5 \leq |H| \leq 6$, $|W|=6$, $|S|=38$, $\#(S,a)=4$, $\#(S,c)=4$.

Suppose x is such that $x \neq a \wedge x \neq c \wedge \neg(e \in G)$.
If $x \neq b$ then CPAF abeα to get
$|dHcW| \geq M(4)+(\#(dHcW,a)+\#(dHcW,b))+\#(dHcW,e) \geq 12+3+2$
- a contradiction.
If $x \neq d$ then CPAF αedc to get
$|TaUb| \geq M(4)+(\#(TaUb,c)+\#(TaUb,d))+\#(TaUb,e) \geq 12+3+2$
- also a contradiction.
The remaining case is $x=b=d$. Lemma 1 (with $\#(S,c)=4$) gives that $\#(TaUb,c) \leq 2$ and since there is at most one symbol in TaUb appearing 3 times, we choose p,q (not c or b) so that $\#(TaUb,p) \leq 2$ and $\#(TaUb,q) \leq 2$. Since $M(3)=7$ there is some permutation Z of c,p,q that is not a subsequence of TaUb. CPAF Zbα to get
$|HcW| \geq M(3)+\#(HcW,b)+\#(HcW,c)+\#(HcW,p)+\#(HcW,q) \geq 7+1+2+2+2 = 14$.
- a contradiction.
From these 3 contradictions we get $(x \in A \wedge x \neq a \wedge x \neq c) \supset \#(G,x) \geq 1$.
Now suppose $\neg(a \in G)$. Choose p,q,r so that $distinct(a,p,q,r)$ and $\neg(pqr \subset dHcW)$. CPAF αapqr. Clearly $a \in U$ [else $|T| \geq M(4)$] and so $\#(TaUb,a) \geq 2$. Hence
$|TaUb| \geq M(3)+\#(TaUb,a)+\ldots+\#(TaUb,r) \geq 7+2+2+2+2 = 15$
From this contradiction we get $\#(G,a) \geq 1$ and by symmetry $\#(G,c) \geq 1$.

Lemma 3: $\forall x \in A.\ \#(G,x) \geq 1$ follows.

Suppose $x \in A \wedge x \neq a \wedge x \neq c$. $\#(T,x)=\#(W,x)=1$, $\#(Ub,x)\geq 1$, $\#(dH,x)\geq 1$ and $\#(G,x)\geq 1$ to yield

Lemma 4: $\forall x \in A.\ (x \neq a \wedge x \neq c) \supset \#(S,x)\geq 5$.

Suppose distinct(a,b,c).
We first choose z to be the first symbol of W which is not a,b.
$b\neq a \wedge b\neq c$ so we have $b\in G$, $b\in dH$ giving $\#(GdH,b)\geq 2$.
$z\neq a \wedge z\neq c$ so we have $z\in G$, $z\in dH$ giving $\#(GdH,z)\geq 2$.
Also $a\neq c$ so $a\in dH$ and we have $a\in G$ giving $\#(GdH,a)\geq 2$.
CPAF $ab\alpha z$ to derive $|GdH| \geq M(4)+\#(GdH,a)+\#(GdH,b)+\#(GdH,z) \geq 18$.
We get from this that $|U|=5$ and also $\#(GdH,b) = 2 = \#(GdH,z)$.
This then gives that $\#(S,z)=5$ and $\#(S,b)=5$.

Let p,q,r be the 3 symbols of the A which are not a,b,c,z.
$\#(S,a) + \#(S,b) + \#(S,c) + \#(S,z) = 4+4+5+5 = 18$
so $\#(S,p) + \#(S,q) + \#(S,r) = 20$.
Since no symbol appears twice in TaUb, can choose a permutation Z of pqr so that $\neg(Z\subset TaUb)$.
CPAF $Z\alpha$ to get $25=|GdHcW|\geq M(4)+(20-6)=26$ - a contradiction.
Similarly 'distinct(a,d,c)' gives a contradiction.

Lemma 5: $\neg$distinct(a,b,c) $\wedge$ $\neg$distinct(a,d,c).

In view of lemma 5, two important cases are $a=c$ and $\neg(a=c)$.

CASE 1. $a=c$.

Suppose first that $a\in U$. Clearly $|U|=6$ and $|TaUb|=14$.
Letting z be the first symbol of W not a,b CPAF $ab\alpha z$ to get $|GdH| \geq 12+\#(GdH,a)+\#(GdH,b)+\#(GdH,z) \geq 17$.
But $|GdH|=17$ so we see $\#(GdH,b) = 2 = \#(GdH,z)$.
Thus $\#(S,a)+\#(S,b)+\#(S,z)=14$.
Now choose p,q,r,s such that pqrsabz is a permutation of A and $\#(S,p) \geq \#(S,q) \geq \#(S,r) \geq \#(S,s)$. Now since some symbol appears at least 7 times in S, $\#(S,p)\geq 7$ and $\#(S,q)+\#(S,r)+\#(S,s)\leq 17$.
Hence $\#(S,s)\leq 5$ and so $\#(S,p)+\#(S,q)+\#(S,r)\geq 19$.
Now each of p,q,r appears exactly twice in TaUb and so

i) $\#(GdHaW,p)+\#(GdHaW,q)+\#(GdHaW,r) \geq 13$

ii) since $M(3)=7$ there is a permutation of pqr (call it Z) such that $\neg(Z \subset TaUb)$.

CPAF $Z\alpha$ to get $24 =|GdHaW| \geq M(4)+13 = 25$.
This contradiction gives us $\#(U,a)=0$.

Again letting z be the first symbol of W not a,b we have $\#(GdH,a)\geq 2$, $\#(GdH,b)\geq 2$, $\#(Gdh,z)\geq 2$ so CPAF $ab\alpha z$ to deduce $|GdH|\geq 18$ and hence $|U|=5$ and $\#(S,b)=\#(S,z)=5$
Similarly, $\#(S,d)=5$ and $|H|=5$.

$|G|=12$ and $\#(G,a) = \#(G,b)=2$ so the other 5 symbols appear a total of 8 times in G. Hence choose p,q so that $\neg(pq\subset G)$ and distinct(a,b,p,q). $\neg(abpq \subset TaUbG)$ so CPAF $abpq\alpha$ to derive a contradiction $|dHaW| \geq 7 + 3*2 + 1 = 14$.

CASE 2. $\neg(a=c)$.

We have $a \neq b$ and $c \neq d$ so Lemma 5 gives both $b=c$ and $d=c$. Hence S looks like TaUbGaHbW with $|T|=6$, $5 \le |U| \le 6$, $10 \le |G| \le 12$, $5 \le |H| \le 6$, $|W|=6$, $\#(G,a)=\#(G,b)=1$, $\#(T,b)=\#(W,a)=1$. Clearly $\#(TUH,a) = 0 = \#(UHW,b)$.

We can write the alphabet in order of decreasing frequency in S as pqrstab where all except a,b occur at least 5 times and $\#(S,p) \ge 7$. Hence, as p,q,r,s,t appear a total of 30 times $\#(S,t)=5$ and $\#(S,s) \le 6$ and $\#(S,p)+\#(S,q)+\#(S,r) \ge 19$.

CASE 2a: $|U|=5$.

Some permutation, Z, of pqr will not be a subsequence of TaUb so CPAF $Z\alpha$ to get $|GaHbW| \ge 12+19-6 = 25$.
This gives us that $\#(S,p)+\#(S,q)+\#(S,r) = 19$ and $\#(S,s)=6$.
We then deduce $\#(S,p)=7$, $\#(S,q) = \#(S,r) = 6$.

Now if z denotes the last symbol of T then CPAF $z\alpha$ to get $32 = |aUbGaHbW| \ge M(6) + \#(S,z) - 1$ or $\#(S,z) \le 5$
But $z \neq a$ so $\#(S,z) \ge 5$ so we deduce $z=t$.
Similarly the first symbol of W is t.

Recall that $\neg(Z \subset TaUb)$, $\#(G,a)=\#(G,b)=1$ and note $\#(G,t)=1$.
CPAF $Zab\alpha$ to deduce that $ab \subset G$.
CPAF $Ztb\alpha$ to deduce that $tb \subset G$.
Similarly deduce that $at \subset G$.
i.e. a precedes t precedes b (in G).

Suppose t is not the last symbol of U. We find y,z such that $\neg(yzt \subset TaUb)$ and so $\neg(yztab \subset TaUbGaH)$. CPAF yztab for the contradiction by which we can conclude $U(5)=t$.

We have that S has the form $T'taU'ftbGaHbtW'$ where $T't=T$, $U'ft=U$ and $tW'=W$ (this defines T', U', f, W').
Clearly $f \neq a$, $f \neq b$, $f \neq t$ and so $\#(S,f) \ge 6$.
Now $\neg(tf \subset TaUb)$ so CPAF $tf\alpha ab$ to get $|G| \ge 7+3+\#(G,f)$.

Suppose $\#(G,f)=1$. From $\#(S,f) \ge 6$ deduce $\#(H,f)=2$.
Now one of tf, ft is not in G - call it Z.
CPAF $abZ\alpha$ to get $|aHbW| \ge 7+1+2+2+3=15$ - a contradiction.

Hence we have $\#(G,f)=2$ and $|G|=12$ so $|H|=5$.

Now let the last symbol of T' be g and suppose $b \neq g$.
$\neg(gb \subset TaU)$ and $\neg(ta \subset G)$ so $\neg(gbta \subset TaUbG)$.
CPAF $gbta\alpha$ to get a contradiction.
Hence the last symbol of T' is b.

Now $\neg(bf \subset T'taU')$ but we have $\neg(ta \subset bG)$ so $\neg(bfta \subset TaUbG)$.
CPAF $bfta\alpha$ to get $12 = |HbW| \ge 7+1+1+2+2 = 13$.
This last contradiction dispenses with CASE 2a.

CASE 2b: $|H|=5$.

The elimination of this case is similar to CASE 2a.

CASE 2c: $|U|\neq 5 \wedge |H|\neq 5$.

We have so far that $S = TaUbGaHbW$ with $|T|=|U|=|H|=|W|=6$
$|G|=10$, $\#(G,a)=\#(G,b)=1$, $\#(TUH,a) = \#(UHW,b) = 0$.

Suppose first that $\#(S,s)=5$.
Without loss of generality suppose s precedes t in G.
$\neg(abts \subset TaUbGa)$. Moreover if any p,q or r precedes s in H
then CPAF $abts\alpha$ to get $|HbW|>7+1+1+4=13$ - a contradiction.
Hence only t may precede s in H.
Similarly only s may follow t in U.
Now CPAF $at\alpha sb$ to get $|G|\geq M(7)+\#(G,a)+\#(G,b)+\#(G,s)+\#(G,t)=11$.
The contradiction serves to give us $\#(S,s)\neq 5$.
Hence $\#(S,s)=6$ and $\#(S,p)=7$, $\#(S,q)=\#(S,r)=6$.

Letting x be the duplicated symbol in U and y the duplicated symbol in H, $\#(U,x)=2$, $\#(H,y)=2$.

If $x=y$ then $\#(S,x)\geq 7$ so $x=p$ and thus $\#(G,x)=1$.
- One of yt, ty (call it Z) is not a subsequence of G.
- CPAF $abZ\alpha$ to get $|HbW|\geq 7+1+1+2+3=14$ - contradiction.

Else if $y\neq p$ then $\#(S,y)=6$ (note $y\neq a$, $y\neq b$, $y\neq t$) and $\#(G,y)=1$
- One of yt, ty (call it Z) is not a subsequence of G.
- CPAF $abZ\alpha$ to get $|HbW|\geq 7+1+1+2+3=14$ - contradiction.

Else $x\neq y \wedge y=p$ so $x\neq p$ and $\#(S,x)=6$.
- One of xt, tx (call it Z) is not a subsequence of G.
- CPAF αZab to get $|TaU|\geq 7+1+1+2+3=14$ - contradiction.

This trio of contradictions completely eliminates CASE 2c.

CASES 2a, 2b, 2c all provided contradictions as did CASE 1 so the assumption that $|S|<39$ is proved impossible.

Q.E.D.