CROSSWORD CONSTRUCTION BY COMPUTER

A. ROSS ECKLER
Morristown, New Jersey

In the November 1980 Word Ways, I briefly described a computer program designed by Lawrence J. Mazlack to place letters one at a time in a specified crossword grid until all vertical and horizontal slots are filled by words in a stored dictionary. Even though his test grids were very simple (the largest one, a 13-by-13, contained 30 per cent blank squares, 7 per cent unchecked letters, and no words of more than four letters), the program did not always succeed in fully filling them with words. More dismally, Mazlack’s program completely failed to find a double word square of size 4-by-4, although this may have been due to a limited vocabulary size (2000 words).

I recently learned of two other researchers who have tackled the crossword construction problem with considerably greater success: P.D. Smith and S.Y. Steen published "A Prototype Crossword Compiler" in The Computer Journal (1981), and O. Feger published "Ein Programm zur Konstruktion von Kreuzworträtseln" in Angewandte Informatik (1975). Their success is apparently due to two factors: (1) larger vocabulary stored in the computer (Smith and Steen 7869 words of lengths 2 - 17, Feger about 10000 words of lengths 3 - 8), (2) a construction algorithm which fills in words one at a time rather than letters one at a time. The ingenuity of their programs lies in the order in which successive word slots (starting with one hand-filled slot to prime the pump) are selected to be filled. Briefly, the strategy is to tackle the more difficult word slots (those that have the fewest alternatives in the dictionary) first. Usually, this means that longer slots are filled first, but if a slot is partially filled in with letters from crossing words, its alternatives may be so reduced that it is promoted to the head of the queue. If, in a sequence of such decisions, the computer is unable to fill a word slot from the dictionary, it backs up and tries another word in an earlier slot.

In the table below, I summarize the results of a number of computer runs by both authors.

<table>
<thead>
<tr>
<th>Feger</th>
<th>Per Cent of Squares that are</th>
<th>Time (secs.)</th>
<th>Longest Word</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Blank</td>
<td>Unchecked</td>
<td>Checked</td>
</tr>
<tr>
<td>8 x 10</td>
<td>15</td>
<td>10</td>
<td>75</td>
</tr>
<tr>
<td>9 x 11</td>
<td>27</td>
<td>12</td>
<td>61</td>
</tr>
<tr>
<td>10 x 12</td>
<td>14</td>
<td>8</td>
<td>78</td>
</tr>
</tbody>
</table>

How clues magazines work with their crossword puzzles is not explained by the techniques described here. However, some of the average of blank spaces is indicated by the use of satisfactory lesser words such as clamos, or by techniques for unchecking words. In some restriction, in addition to what happens to the solution of the puzzle, the program may be able to provide a solution if the solution is not unique.
ITER

computer

one at a time

e than four

failed to

have been

sent blank

with

failed to

have been

sent blank

them with

failed to

have been

sent blank

The first three Smith and Steen puzzles were based on crossword

grids actually published in British newspapers of May 1980 (Daily Re­

cord, Daily Telegraph, Guardian); a typical puzzle is given below.

How close are computers to producing crossword puzzles found in

magazines and newspapers? If it is possible for cryptic clue-weavers to

work with any words presented to them, then they are very close indeed.

However, I suspect that this is not the case, and that the higher percent­

ages of blank squares and unchecked letters in British puzzles are neces­

sitated by the fact that many words don't lend themselves to clever crypt­

ic clues. In American puzzles, any word can be clued; howeve­

rs one at a which suc­

ome the pump)

the more dif­

the dictionary)

, but if a slot

lternatives

ue. If, in a

word slot

an earlier slot.

of computer

Least

Word

8

8

7

Smith and Steen

12 x 12 29 43 28 73 8

15 x 15 31 44 25 535 11

15 x 15 37 45 18 6 11

4 x 4 0 0 100 1500,15,304

The first three Smith and Steen puzzles were based on crossword

grids actually published in British newspapers of May 1980 (Daily Re­

cord, Daily Telegraph, Guardian); a typical puzzle is given below.

How close are computers to producing crossword puzzles found in

magazines and newspapers? If it is possible for cryptic clue-weavers to

work with any words presented to them, then they are very close indeed.

However, I suspect that this is not the case, and that the higher percent­

ages of blank squares and unchecked letters in British puzzles are neces­

sitated by the fact that many words don't lend themselves to clever crypt­

ic clues. In American puzzles, any word can be clued; howeve­

rs one at a which suc­

ome the pump)

the more dif­

the dictionary)

, but if a slot

lternatives

ue. If, in a

word slot

an earlier slot.

of computer

Least

Word

8

8

7

The first three Smith and Steen puzzles were based on crossword

grids actually published in British newspapers of May 1980 (Daily Re­

cord, Daily Telegraph, Guardian); a typical puzzle is given below.

How close are computers to producing crossword puzzles found in

magazines and newspapers? If it is possible for cryptic clue-weavers to

work with any words presented to them, then they are very close indeed.

However, I suspect that this is not the case, and that the higher percent­

ages of blank squares and unchecked letters in British puzzles are neces­

sitated by the fact that many words don't lend themselves to clever crypt­

ic clues. In American puzzles, any word can be clued; howeve­

rs one at a which suc­

ome the pump)

the more dif­

the dictionary)

, but if a slot

lternatives

ue. If, in a

word slot

an earlier slot.