A SCRABBLE CROSSWORD GAME PLAYING PROGRAM*

Stuart C. Shapiro
Department of Computer Science
State University of New York at Buffalo
Amherst, New York 14226

Howard R. Smith
Department of Computer Sciences
The University of Texas at Austin
Austin, Texas 78712

Keywords and Phrases: SCRABBLE Crossword Game, game playing, lexicon, dictionary, letter table, trie, data structures, searching.

CR Categories: 3.64, 3.66, 3.74

*The work reported herein was done while both authors were at the Computer Science Department, Indiana University, Bloomington, Indiana 47401.
Abstract

A program has been designed and implemented in SIMULA 67 on a DECSYSTEM-10 to play the SCRABBLE Crossword Game interactively against a human opponent. The heart of the design is the data structure for the lexicon and the algorithm for searching it. The lexicon is represented as a letter table, or trie using a canonical ordering of the letters in the words rather than the original spelling. The algorithm takes the trie and a collection of letters, including blanks, and in a single backtrack search of the trie finds all words that can be formed from any combination and permutation of the letters. Words using the higher valued letters are found before words not using those letters, and words using a collection of letters are found before words using a sub-collection of them. The search procedure detaches after each group of words is found and may be resumed if more words are desired.
1. Introduction

The SCRABBLE Crossword Game is a well known game considered to require a fair amount of intelligence, strategic and tactical skill, facility with words, and luck. Unlike most games in the artificial intelligence literature, it can be played by two or more players and is neither a zero sum game nor a game of perfect information. For these reasons, game tree searching procedures do not seem applicable. When humans play the game, a large easily accessible vocabulary seems to be the most important determiner of victory. One might, therefore, think that it would be easy to write a program that plays the SCRABBLE Crossword Game at the championship level. However, several issues are not so clear: How should the lexicon be organized for maximum usefulness?; How should a program decide where to play?; How can a program take advantage of the small literature on the strategy and tactics of the game (e.g. [1])?; What is the relative importance of a good memory for words vs. skillful decisions about what letters to use or not to use and where to play?

Our interest has focused on the organization of the lexicon. This lexicon differs from those usually used in natural language processing programs because of the use to which it is to be put. Usually one is confronted with a possible word. One must determine if it is a word, and, if so, segment it into affixes and stem, and retrieve lexical information associated with the stem. One must allow for the possibility that the word is lexically ambiguous and that the segmentation can be done in several different ways. The problem for the SCRABBLE Crossword Game lexicon is, given a set of letters, find all the words that can be made from any combination and permutation of them. This is a very different problem, but similar to the information retrieval problem of,
given a set of keys find all records that contain any subset of them, or the pattern recognition problem of, given a set of features, some of which may be spurious, find all objects that might have given rise to them.

We have designed a lexicon and an algorithm for searching it. Around them, we have designed and implemented a program that plays the SCRABBLE Crossword Game interactively at a competitive, human level. We chose SIMULA 67 [2,3,4] for the programming language for reasons to be discussed below. In the sections below, we first briefly describe the program organization and the board evaluation technique, then we discuss the lexicon and the lexicon search algorithm in more detail.

2. Program Organization

Because our major concern was with the lexicon and associated algorithms, the rest of the program is the minimal required to play a reasonable game. We will describe the program as it is, and in the final section of this paper indicate desirable enhancements.

The program is written to play a two person game - the program player against one human opponent. Figure 1 shows the overall organization of the major program modules, which we shall describe below.

2.1 The Game Manager

The game manager handles interaction with the human opponent and keeps track of all game information. It deals out tiles to the human opponent and the program player at the beginning of the game and after each play. Dealing is done randomly (without replacement) from a list initialized correctly at the beginning of the game. The seed of the pseudo-random number generator is chosen by the human at the start of
FIGURE 1: Organization of Major Program Modules
the game so that a game can be repeated to see the effect of learning new words. Whether the program or the human plays first is decided by randomly drawing a tile for each, according to the rules. The game manager's basic cycle is as follows:

1. Print out the current board.

2. Print out the human's rack and request an instruction. (The human's options at this point are described in Section 2.3).

3. If the human specifies a word to play, pass it to the referee for checking and scoring.

4. If the referee says that the play is illegal (see Section 2.5), print an appropriate message and go to step (2).

5. If the move was legal, update the board representation and the human's score, deal the human new tiles to replace the ones just used, print the score for the move just made and the new rack so the human can start thinking about the next move.

6. Pass the human's move to the board analyzer for updating the list of possible plays.

7. If the inspect made is on, print the program player's rack.

8. Get the program's move, update the board and the program's score, print the move and the score of the move.

9. Pass the program's move to the board analyzer.

10. Go to step (1).

The game ends when the human so specifies in step (2). Each player's total score is decreased by the value of the tiles in his/her rack,
and the final scores are printed.

2.2 The Board Representation

The board is represented as a 15 x 15 array of SQUAREs. In interaction with the human, the rows are numbered 1-15 and the columns A-O, in the official manner [1,pp33-4]. Each SQUARE has a value as shown below:

<table>
<thead>
<tr>
<th>Kind of SQUARE</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>normal</td>
<td>0</td>
</tr>
<tr>
<td>double letter</td>
<td>1</td>
</tr>
<tr>
<td>triple letter</td>
<td>2</td>
</tr>
<tr>
<td>double word</td>
<td>3</td>
</tr>
<tr>
<td>triple word</td>
<td>4</td>
</tr>
</tbody>
</table>

If a SQUARE is occupied by a tile, it also contains the tile, an indication of whether the word using that tile is horizontal or vertical, and the value of a play using that tile as assigned by the board analyzer. Figure 2 shows how the board is printed before any play has been made.

2.3 The Human Opponent

When it is the human's turn to move, the following options are available:

0. terminate the game
1. play a word
2. pass without exchanging tiles
3. pass and exchange some tiles
4. have the total scores of both players printed
### FIGURE 2: How the board looks to the human before any words are played.
have the board and human's rack printed

access the lexicon manager (see Section 2.7)

general instructions for interacting with the program

general this list of options.

When the human plays a word, a notation close to the official notation of the SCRABBLE Crossword Game PLAYERS [1,pp33-4] is used. The word is typed in with each letter already on the board preceded by a "$". When a blank tile is played, a "#" is typed followed by the letter it is being used as. For example, the play, $F#IRM, means that the player is using a blank, an R and an M from the rack and an F from the board, is using the blank as an I, and is spelling the word, "firm".

The location of the word is indicated by the official notation. For example, to place a word horizontally on row 12 from column K to column N, the notation is 12K-N. To place a word vertically on column C from row 13 to row 15, the notation is C13-15.

2.4 The Program Player

There are two basic ways to choose a play: 1) pick a place to play, then find a word that will fit; 2) pick a word to make, then find a place to put it. In either case, one might assign high priority to using or to not using certain tiles in the rack in order to keep the rack balanced or to build toward being able to make a bonus word (using all seven tiles on the rack). The second method is easier to use if the word chosen is made entirely from tiles in the rack, than if tiles from the board are also to be used (see Appendix, moves 3, 17 and 29). This is most appropriate when a bonus word can be made entirely from the rack. We decided to use the first method.
We adopted the following simplified approach. Possible places to play are tiles already on the board. If a tile is already being used in a horizontal word, only a vertical play will be considered and vice versa. This means that often valuable moves like 3, 17, 19, 21, 33, 35 and 39 in the Appendix will not be considered by this version of the program.

In certain situations, choice of a tile to play on puts obvious (to humans) limitations on possible words. For example, if we decide to play on a K in column N only words with a K as last letter or second to last letter are possible. One might decide to build such limitations into the lexicon search algorithm. However, this complicates situations where we decide to play on a centrally located tile. We decided to ignore such limitations in the lexicon lookup and use the referee to check each word found to see if it really fits where the player wants to put it.

The program player uses three parameters, MAXLOC, MAXWDS, and MINPT. It considers the best MAXLOC positions from the ordered list maintained by the board analyzer. For each position, it passes the tiles in its rack plus the tile at the position to the lexicon search algorithm, and considers the first MAXWDS words found. Each word is passed to the referee along with its proposed location. The referee determines if the word really fits there, and if so, returns a score for the play. As soon as a play worth at least MINPT points is found, it is played. If no such play is found, the highest scoring play is made. If none of the MAXLOC x MAXWDS words are playable, the player exchanges all the tiles on its rack for new ones. Of course, if fewer than seven tiles are left to draw from, this exchange is illegal and the program just passes. In the game recorded in the Appendix, MAXLOC
was 6, MAXWDS was 5 and MINPT was 14. Of the program's 25 moves, 6 were worth MINPT or more, 10 were worth fewer than MINPT, the program exchanged its rack 7 times, and passed twice. The average score for the 16 words actually played was 13.125. The average score for the human's 25 words was 13.88.

When the inspect mode is on, all positions and words the program player considers are printed on the terminal.

2.5 The Referee

The referee is called by the program player with a proposed move and by the game manager with the human's move. In either case, a word and a location are passed to the referee, which checks the following (not necessarily in the order shown):

1. The location and the length of the word are consistent.

2. The word will not extend off the board.

3. All letters except those preceeded by "$" are on the player's rack.

4. The word contains at least one tile from the board or is adjacent to a tile on the board in at least one place.

5. The location places each letter preceeded by a "$" where such a tile is already on the board and every other tile on an empty square.

6. The two ends of the word are not adjacent (in the direction of the word) to non-empty squares.

If the referee is judging the human's move, it passes the word and all cross-words formed by the play to the lexicon search routine. If one of these words is not in the lexicon, the human is asked if it
really is a word. If the human responds "yes", the word is given to
the lexicon manager for immediate insertion into the lexicon. If the
response is "no", the current version of the program allows the play
anyway.

If the play is legal, the referee computes its score, including
the main word and all newly formed cross-words, and returns this to
the caller.

2.6 The Board Analyzer

The board analyzer is called by the game manager after each move
to update the values of the playable tiles and the ordered list of these
positions. The analyzer begins by setting to 0 the value of each tile
already on the board that has now been used in both directions. It
then computes a value for each tile newly placed on the board by the
play. If the play was vertical (horizontal), the value is the sum of
the value of the square and the value of each unoccupied square in the
same row (column) except that no square is included that is more than
seven squares away from the tile being evaluated or that is beyond the
closest occupied square. Those occupied squares are then reevaluated and
the ordered list of playable positions is updated for the program
player's next move. The values calculated by the board analyzer are
stored in each evaluated SQUARE as mentioned in Section 2.2. For exam-
ple, Figure 3 shows a section of the board after placing REA$L at 12A-D.
The value of the L at 12D is changed to 0, the values of the R, E, and A
in row 12 are set to 9, 8, and 4 respectively, and the W at 9C is re-
evaluated from 11 to 8.

We realize that the board evaluation scheme needs improvement, but
it does provide a rough approximation to an adequate ordering of the
FIGURE 3: A section of the board after playing REA$L at 12A-D
playable positions and we were more concerned with the lexicon, to which we now turn.

2.7 The Lexicon Manager

The lexicon manager can perform the following operations:

1. Insert a word into the lexicon

2. Delete a word from the lexicon

3. Check if a particular word is in the lexicon

4. Find words that can be formed from any permutation and combination of a set of tiles.

The referee uses operation 3 to check the human's plays and operation 1 when the human insists that a word that isn't in the lexicon should be. The referee also uses operation 3 to check cross-words formed by plays proposed by the program player. The program player uses operation 4 to find possible words to play. The human can interact with the lexicon manager by choosing option 6 (see Section 2.3). He/She can then repeatedly use options 1, 2 or 4 before returning to the game. This option has been used by the authors for early editing of the lexicon. For example, we entered all legal two letter words [1,p.156] this way.

Option 4 takes advantage of the SIMULA 67 detach facility by taking a set of tiles and a number, n. It returns approximately n words (see below), and then can either search the lexicon using a different set of tiles or can resume the search for more words using the original set of tiles. The availability of detach was the main reason for choosing SIMULA 67 as the programming language.
3. The Lexicon Data Structure

The lexicon was designed for the particular problem, "given a collection of letters, find all words that can be formed from any combination and permutation of the letters, and find these words in approximately the order of their value in the SCRAMBLE Crossword Game." Two key ideas were combined in the design of the data structure - letter tables [5, 6, 8], or tries [7], and canonical ordering.

A letter table is a binary tree in which each node has the following four fields: 1) a letter; 2) information depending on the application; 3) a success link to a subtree; 4) a fail link to a subtree. To look up a word in a letter table one starts at the root and follows fail links until a node is found whose letter is the first letter in the word. Then one looks up the remainder of the word in the success subtree of that node. When a node is found for the last letter in the word, the information field contains information about that word, e.g. lexical features. Note that as a word is looked up, all words formed from initial sub-sequences of the word are also found.

Consider the problem of designing a human readable anagram lexicon to find all words that can be formed from permutations of a given collection of letters. Take any word to be listed and reorder its letters according to any canonical ordering, e.g. alphabetical. List these re-spellings lexicographically based on the same ordering, and under each re-spelling list the original word. For example, to find anagrams of "live", look up "eily", and find "evil", "live", and "vile".

Combining these two ideas, the data structure of our lexicon is a letter table using a canonical ordering of letters instead of the original spelling order. The information field of each node contains a list of the words formed from permutations of the letters which access
the node. So that words can be found in approximately the order of their values in the SCRABBLE Crossword Game, the canonical ordering used is the value of the letters in that game. Among letters with the same value, the one with the least frequency in English is sorted first to increase the bushiness of the tree. The ordering used is:

ZQJXKVWYFBPMCGDULSIRNOATE

Figure 4 shows a small lexicon using this data structure. The fields are shown as indicated below.

<table>
<thead>
<tr>
<th>letter</th>
<th>words</th>
</tr>
</thead>
<tbody>
<tr>
<td>fail</td>
<td>succ</td>
</tr>
</tbody>
</table>

Each node is numbered for reference in Section 4.

The lexicon is maintained in a sequential file on disk, ordered in the same way it is searched: letter field, words field, success subtree, fail subtree. This file is scanned as the tree is searched and subtrees that are not needed are not stored in main memory.

4. The Lexicon Search Algorithm

The search algorithm takes the lexicon tree and a list of letters in the canonical order. The list of letters may contain one or more occurrences of the blank, "#", which may match any letter and are initially placed at the front of the list. The search algorithm finds all words in the tree that are formed from any permutation and combination of the letters in the list. It finds words using the higher valued letters before words not using them. It also finds words using a collection of letters before words using a proper sub-collection of those letters. For example, if a word can be formed using all the letters
FIGURE 4: An example of the lexicon data structure
in the list, this will be the first word found.

In the algorithm presented below, the procedure \texttt{detach} is to be understood as causing control to be returned to the original caller of the procedure in such a way that the current value of the local variables are available to the original caller, and so that the procedure may be resumed at the point right after the \texttt{detach}. Other variables and functions used are:

- \texttt{tree}: a node of the lexicon tree
- \texttt{letters}: a list of letters ordered as discussed above
- \texttt{nil}: the empty list and the empty tree
- \texttt{letter(tree)}
- \texttt{words(tree)}: the four fields of tree
- \texttt{succ(tree)}
- \texttt{fail(tree)}
- \texttt{hd(letters)}: the first letter on the list
- \texttt{tl(letters)}: the list without the first letter
- \texttt{next-let(letters)}: the first non-\# letter on the list, or \texttt{nil} if there is none
- \texttt{move-to-front(letter, letters)}: equal to \texttt{letters}, but with the first occurrence of \texttt{letter} moved to the front of the list
- \texttt{before(letter1, letter2)}: true if \texttt{letter1} is before \texttt{letter2} in the canonical ordering or if \texttt{letter2} is \texttt{nil}, false otherwise
- \texttt{remove-next-let(letters)}: equal to \texttt{letters} with the first none-\# letter removed.
The search algorithm is:

```
procedure findwords(tree, letters);
    while tree ≠ nil and letters ≠ nil do
        if letter(tree) = next-let(letters) then
            begin {lexicon node matches first non-# letter}
                letters ← move-to-front(next-let(letters), letters);
                findwords(succ(tree), tl(letters));
                if words(tree) ≠ nil then detach;
                letters ← tl(letters); {now ignore this letter and}
                tree ← fail(tree) {continue searching}
            end
        else if hd(letters) = '#' then
            {lexicon node doesn't match first non-# letter, but it can match the #}
            if before(letter(tree), next-let(letters)) then
                begin {allow the # to match the node}
                    findwords(succ(tree), tl(letters));
                    if words(tree) ≠ nil then detach;
                    tree ← fail(tree) {the # can match something else}
                end
            else letters ← remove-next-let(letters)
                {we are past this letter, so ignore it}
            else if before(hd(letters), letter(tree)) then
                letters ← tl(letters) {we're past this letter, so ignore it}
            else tree ← fail(tree) {keep looking}
```

Table 1 shows a trace of this procedure on the lexicon of Figure 4 and the list of letters, #VSIE. Note that success links are followed recur-
<table>
<thead>
<tr>
<th>Recursion Level</th>
<th>Node of Tree</th>
<th>Letter of Tree</th>
<th>Letters</th>
<th>Test Results</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Q</td>
<td>#VSIE</td>
<td>Q≠V, hd(LETTERS)≠#, before(Q,V) # matches Q, follow succ(1)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>U</td>
<td>VSIE</td>
<td>U≠V, hd(LETTERS)≠#, before(V,U) ignore V</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>U</td>
<td>SIE</td>
<td>U≠S, hd(LETTERS)≠#, ¬before(S,U) follow fail(2)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>nil</td>
<td>SIE</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Q</td>
<td>#VSIE</td>
<td>words(1) = nil follow fail(1)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>K</td>
<td>#VSIE</td>
<td>K≠V, hd(LETTERS)≠#, before(K,V) # matches K, follow succ(5)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>L</td>
<td>VSIE</td>
<td>L≠V, hd(LETTERS)≠#, before(V,L) ignore V</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>L</td>
<td>SIE</td>
<td>L≠S, hd(LETTERS)≠#, ¬before(S,L) follow fail(6)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>nil</td>
<td>SIE</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>K</td>
<td>#VSIE</td>
<td>words(5) = nil follow fail(5)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>V</td>
<td>#VSIE</td>
<td>V=V move V to front, follow succ(11)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>L</td>
<td>#SIE</td>
<td>L≠S, hd(LETTERS)≠#, before(L,S) # matches L, follow succ(12)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>S</td>
<td>SIE</td>
<td>S=S follow succ(13)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>I</td>
<td>IE</td>
<td>I=I follow succ(14)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>E</td>
<td>E</td>
<td>E=E follow succ(15)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>nil</td>
<td>nil</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>E</td>
<td>E</td>
<td>words(15) = LIVES, EVILS Detach</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>E</td>
<td></td>
<td></td>
<td>ignore E and follow fail(15)</td>
</tr>
<tr>
<td>5</td>
<td>nil</td>
<td>nil</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>I</td>
<td>IE</td>
<td>words(14) = nil ignore I and follow fail(14)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>nil</td>
<td>E</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>S</td>
<td>SIE</td>
<td>words(13) = nil ignore S and follow fail(13)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>I</td>
<td>IE</td>
<td>I=I follow succ(16)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>E</td>
<td>E</td>
<td>E=E follow succ(17)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>nil</td>
<td>nil</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>Recursion Level</td>
<td>Node of Tree</td>
<td>Letter of Tree</td>
<td>Letters</td>
<td>Test Results</td>
<td>Action</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------</td>
<td>-----------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>E</td>
<td>E</td>
<td>words(17) = LIVE, EVIL, VILE</td>
<td>Detach</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>E</td>
<td>E</td>
<td></td>
<td>ignore E and follow fail(17)</td>
</tr>
<tr>
<td>4</td>
<td>nil</td>
<td>#SIE</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>I</td>
<td>IE</td>
<td>words(16) = nil</td>
<td>ignore I, follow fail(16)</td>
</tr>
<tr>
<td>3</td>
<td>nil</td>
<td>#SIE</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>L</td>
<td>#SIE</td>
<td>words(12) = nil</td>
<td>follow fail(12)</td>
</tr>
<tr>
<td>2</td>
<td>nil</td>
<td>#SIE</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>V</td>
<td>V#SIE</td>
<td>words(11) = nil</td>
<td>ignore V, follow fail(11)</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>N</td>
<td>#SIE</td>
<td>N#S, hd(LETTERS)=#, ¬before(N,S)</td>
<td>ignore S</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>N</td>
<td>#IE</td>
<td>N#I, hd(LETTERS)=#, ¬before(N,I)</td>
<td>ignore I</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>N</td>
<td>#E</td>
<td>N#E, hd(LETTERS)#, before(N,E)</td>
<td># matches N, follow succ(18)</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>O</td>
<td>E</td>
<td>O#E, hd(LETTERS)#, ¬before(E,O)</td>
<td>follow fail(19)</td>
</tr>
<tr>
<td>2</td>
<td>nil</td>
<td>E</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>N</td>
<td>#E</td>
<td>words(18) = nil</td>
<td>follow fail(18)</td>
</tr>
<tr>
<td>1</td>
<td>nil</td>
<td>#E</td>
<td></td>
<td></td>
<td>Return</td>
</tr>
</tbody>
</table>


sively since we will later back up along these links, but fail links are followed iteratively since once following a fail link from a node we are no longer interested in that node or its success subtree. This is the method by which only those pieces of the tree in which we are interested are retained in main memory. Also note that since words are found in groups, if \( n \) words were wanted, at least \( n \) words, but possibly a few more, will be found.

One deficiency of this search algorithm is that when the program player picks a board position to play on and passes the tile at that position along with the tiles on its rack to the search algorithm, words that do not use the tile on the board may be returned. We want to be able to mark certain letters in the list as required to be used in the retrieved words. We will define the procedure \texttt{findwords2} that takes a node of the tree and an ordered list of nodes, \( n \) with the following fields:

\begin{align*}
\text{letter}(n) & : \text{a letter corresponding to an element of the second argument of findwords} \\
\text{req}(n) & : \text{true if letter}(n) \text{ is required to be in the retrieved words, false otherwise} \\
\text{done}(n) & : \text{true if no node further on in the list is required, false otherwise}
\end{align*}

Given such a list with the letter and req fields set, it is easy to define the procedure \texttt{setdone} which returns the list with the done fields set properly. In the procedure below, \texttt{letters} is such a list after \texttt{setdone} has been applied, \texttt{next-let(letters)} returns the letter field of the first node in \texttt{letters} whose letter field is not \#,

\texttt{next-req(letters)} returns the req field of the first node in \texttt{letters} whose
letter field is not #, remove-next-let(letters) removes from letters
the first node whose letter field is not #, and move-to-front(letter,
letters) moves to the front of letters the first node whose letter
field is letter and if that node has a true req field returns setdone
of the changed list. Lines in findwords2 that are different from
corresponding lines in findword are marked by a "*" in the left margin.

procedure findwords2(tree, letters);
    while tree ≠ nil and letters ≠ nil do
        if letter(tree) = next-let(letters) then
            begin {lexicon node matches first non-# letter}
                letters + move-to-front(next-let(letters), letters);
                findwords(succ(tree), tl(letters));
                if words(tree) ≠ nil and done(hd(letters)) then
                    detach;
                if req(hd(letters)) then tree + nil
                else begin letters + tl(letters);
                    tree fail(tree)
                end
            end
        else if hd(letter) = '#' then
            {lexicon node doesn't match first non-# letter, but
            but it can match the #}
            if before(letter(tree), next-let(letters)) then
                begin {allow the # to match the node}
                    findwords(succ(tree), tl(letters));
                    if words(tree) ≠ nil and done(hd(letters))
                        then detach
                    tree fail(tree) {the # can match something else}
                end
            else if next-req(letters) then letters + nil
            else letters + remove-next-let(letters)
        else if before(hd(letters), letter(tree)) then {we've passed
            this letter}
            if req(hd(letters)) then letters + nil
        else letters + tl(letters)
    else tree fail(tree) {keep looking}
Notice that the req field keeps us from ignoring required letters, and the done field prevents words from being returned unless all required letters are included.

5. Conclusions

The lexicon data structure and search algorithm seem successful. The program player quickly makes use of high valued tiles on the board or on its rack. There is still much room for improvement besides increasing the program's vocabulary. Implementing findwords2 would make the program player's search for words to play more efficient and more likely to yield playable words without changing the MAXLOC and MAXWDS parameters. The board evaluation scheme is still rather primitive. For example, the value of double letter squares should be twice the average value of the (still unplayed) tiles instead of 1. The program player should be modified to allow for playing words parallel to adjacent words and for extending existing words. Strategies such as maintaining a balanced rack and retaining high frequency poly-graphs to increase the future probability of a bonus word might be added. Even without these improvements, the program plays a competitive game against a human opponent.

Acknowledgements

The authors appreciate the work of Ben Shneiderman, Margaret Ambrose and Barbara Rasche who along with the senior author, worked on a preliminary version of this program. Computer services were provided by IUPUI Computing Facilities as part of the Indiana University Computing Network.
References


Appendix

Table 2 provides a record of a complete game played between the program player and a human, DPF. DPF had previously played several rushed rounds against the program and consistently scored lower than the program. However, in the game presented here, he took advantage of his knowledge of the program's weaknesses. The game was played with the inspect mode on and took 17 minutes, 28.56 seconds of CPU time on a DECSYSTEM-10 and 2 hours, 50 minutes, 39.05 seconds of real time. At the end of the game, there were 1488 words in the lexicon.

In Table 2, a letter already on the board is underlined and blanks are represented as a square. Figure 5 shows the board at the end of the game.

```
   A B C D E F G H I J K L M N O
  1  4  1  4  E  Q U E 4  1
  2  3  2  E V E  L O 2
  3  3  1  I W I N K 3
  4  1  3  I L O I N 1  4
  5  3  S N O W 5
  6  2  2  D E R E 6
  7  1  T H V A N 7
  8  4  1 C H O O F 1 A P T 8
  9  1 B U S Y 1 I F N R 9
 10  2  L A R A G E 2 E 10
 11  T I M I Y E 11
 12  1  3  D I M E L A Z E S 12
 13  J C O G 1 O N  I T 13
 14  A 3 A 2 R A D A R P I T 14
 15  R E B U D G E T S 1 0 15
   A B C D E F G H I J K L M N O
```

**FIGURE 5:** The board at the end of the game of Table 2
<table>
<thead>
<tr>
<th>Rack</th>
<th>Word</th>
<th>Location</th>
<th>DPF Play</th>
<th>DPF Total</th>
<th>Score Play</th>
<th>Score Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>IFOHIOT</td>
<td>HOOF</td>
<td>8G-J</td>
<td>20</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>2.</td>
<td>GSMHR[ ]B</td>
<td>TRIM</td>
<td>J8-11</td>
<td>28</td>
<td>48</td>
<td>9</td>
</tr>
<tr>
<td>3.</td>
<td>LATIFIT</td>
<td>FAIL</td>
<td>K9-12</td>
<td>66</td>
<td>68</td>
<td>12</td>
</tr>
<tr>
<td>4.</td>
<td>EZIPGBH</td>
<td>HO</td>
<td>I7-8</td>
<td>14</td>
<td>62</td>
<td>9</td>
</tr>
<tr>
<td>5.</td>
<td>ITYTTIT</td>
<td>THY</td>
<td>G7-9</td>
<td>12</td>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>6.</td>
<td>AEZIGSB</td>
<td>LAZE</td>
<td>12K-N</td>
<td>12</td>
<td>80</td>
<td>12</td>
</tr>
<tr>
<td>7.</td>
<td>YCITIT</td>
<td>YET</td>
<td>N11-13</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>8.</td>
<td>UNOIGSB</td>
<td>BUSY</td>
<td>9D-G</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>9.</td>
<td>DLCITIT</td>
<td>CULT</td>
<td>E8-11</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>10.</td>
<td>PEDNOIG</td>
<td>ZIP</td>
<td>M12-14</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>11.</td>
<td>ONADIIT</td>
<td>SAID</td>
<td>F9-12</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>12.</td>
<td>RPEDNOG</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>AREONIT</td>
<td>PI</td>
<td>14M-N</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>14.</td>
<td>GMRDNEI</td>
<td>DIME</td>
<td>12F-I</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>15.</td>
<td>AAREONT</td>
<td>MOAT</td>
<td>H12-15</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>16.</td>
<td>ANRGRDN</td>
<td>RADAR</td>
<td>14G-K</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>17.</td>
<td>TEEAREN</td>
<td>ENTREE</td>
<td>06-11</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>18.</td>
<td>UAJTNGN</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>ACISEAE</td>
<td>LAZES</td>
<td>12K-O</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>20.</td>
<td>PUBLSGD</td>
<td>BUDGET</td>
<td>15C-H</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>21.</td>
<td>RACIEAE</td>
<td>REBUDGET</td>
<td>15A-H</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>22.</td>
<td>INQIVPS</td>
<td>ON</td>
<td>13H-I</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>23.</td>
<td>UACIEAE</td>
<td>CAB</td>
<td>C13-15</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>24.</td>
<td>EIQIVPS</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>VNUIEAE</td>
<td>VAN</td>
<td>7M-0</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>26.</td>
<td>ROGLNIGI</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>ANUIEAE</td>
<td>VANE</td>
<td>M7-10</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>28.</td>
<td>EQVLUNI</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.</td>
<td>ITOUIAE</td>
<td>TO</td>
<td>O14-15</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>30.</td>
<td>DEEOOKW</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31.</td>
<td>JNUIIAE</td>
<td>JAR</td>
<td>A13-15</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>32.</td>
<td>PGGRLVQ</td>
<td>RAP</td>
<td>N6-8</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>33.</td>
<td>ISNIUIE</td>
<td>ENDS</td>
<td>I12-15</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>34.</td>
<td>EOGGLVQ</td>
<td>exchange rack</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35.</td>
<td>WINIUIE</td>
<td>WRAP</td>
<td>N5-8</td>
<td>14</td>
<td>62</td>
<td>14</td>
</tr>
<tr>
<td>Rack</td>
<td>Word</td>
<td>Location</td>
<td>DPF</td>
<td>Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>----------</td>
<td>-----</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36.</td>
<td>EDOWNSE</td>
<td>5J-N</td>
<td>14</td>
<td>153</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37.</td>
<td>OINIUIE</td>
<td>L3-5</td>
<td>6</td>
<td>297</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38.</td>
<td>ONKEDWE</td>
<td>3K-N</td>
<td>22</td>
<td>175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39.</td>
<td>QGOIUUE</td>
<td>10J-M</td>
<td>5</td>
<td>302</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40.</td>
<td>LUIOEDE</td>
<td>N1-3</td>
<td>14</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41.</td>
<td>LQOIUIE</td>
<td>1L-N</td>
<td>22</td>
<td>324</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42.</td>
<td>EGUDIOE</td>
<td>pass</td>
<td></td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43.</td>
<td>XOLOIE</td>
<td>K3-6</td>
<td>9</td>
<td>333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44.</td>
<td>EGUDIOE</td>
<td>6J-K</td>
<td>7</td>
<td>196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45.</td>
<td>EVXOLOI</td>
<td>4I-L</td>
<td>4</td>
<td>337</td>
<td></td>
<td></td>
</tr>
<tr>
<td>46.</td>
<td>EGUIOE</td>
<td>13C-E</td>
<td>6</td>
<td>202</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47.</td>
<td>EVXOI</td>
<td>I1-4</td>
<td>8</td>
<td>345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48.</td>
<td>EUIE</td>
<td>2H-J</td>
<td>8</td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49.</td>
<td>XO</td>
<td>2N-O</td>
<td>2</td>
<td>347</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50.</td>
<td>UI</td>
<td>pass</td>
<td></td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51.</td>
<td>X</td>
<td>pass</td>
<td>347</td>
<td>339</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>208</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>