



Bilkent University

Department of Computer Engineering

CS 461 - Artificial Intelligence

Term Project Report

CROSSWIND

Section: 1

Contributing Members:

Melisa Taşpınar	21803668
Yiğit Gürses	21702746
Ömer Ünlüsoy	21702136
Zeynep Berfin Gökalp	21601956

Table of Contents

1 Introduction	2
2 Description of the Implementation	3
3 Conclusion	6
4 Screenshots	8
Bibliography	15
Appendix	16

1 Introduction

Crossword is a type of puzzle, offering clues with word plays for humans to come up with answers that fills a grid. To have any chance of success with such a puzzle, the solver is required to have some knowledge about popular culture, history, and language (a.k.a. common sense). Moreover, for a given clue, the reasoning ability to select a single reasonable answer from many possible options that fits into the grid is crucial [1]. If a set of candidate words are given, a computer program can easily perform such a task using the constraints of the puzzle. However, due to the tricky nature of crossword clues, it is very hard for a program to come up with a reasonable set of candidates. These features of crossword puzzles not only make them engaging and fun for humans, but also make them a great challenge to test the prowess of artificial intelligence techniques [2]. “Information retrieval and set elimination”, “database search”, and “machine learning” are several approaches for solving crossword puzzles [1]. However, using a combination of them to get a better overall performance should also be possible.

An approach that utilizes NLP could be considered more true to the spirit of crossword puzzles than a brute force approach based on only constraint satisfaction. Advancements in the field of NLP enable new and smarter approaches for candidate generation. In [3], Jobin et. al. uses Lesk score and path similarity to rank all words in wordnet in order to get a set of candidate words. According to their results, over half the time, the 1st ranked word was the correct answer. This shows that such an approach can be very successful. However, they stopped their work at the candidate generation stage and did not make a complete crossword solver.

In this project, the aim was to create an AI based program that solves the Newyork Times mini crossword puzzle designed by Joel Fagliano. This is a 5x5 puzzle that has 5 across and 5 down clues and their corresponding entries. The entries usually have many crossings which creates lots of constraints. This makes the puzzle more approachable to AI based programs.

Our program searches several sources on the internet to retrieve possible answers for clues and selects a set of answers from them. The online sources used were merriam-webster's dictionary and thesaurus, and Wikipedia. We also downloaded Wordnet and used it as an offline resource. Constraint satisfaction and depth first search were used in the solution process. In the following sections, the approaches used throughout the project will be described and the results will be evaluated.

2 Description of the Implementation

As stated earlier, the main aim of this project is to come up with answers for Newyork Times mini crossword puzzle and maximize the number of entries guessed correctly. To achieve this, there are several stages that need to be completed successfully. First of all, the puzzle data needs to be scraped from the NYT website. Then, using the puzzle data, candidate words for each clue need to be generated. These candidates should be the required length, be relevant to the clue and hopefully include the correct answer. Then, the number of candidates must be reduced using constraint satisfaction in order to have a reasonable time complexity and reduce unnecessary computations. After the reduction is complete, possible solutions from the remaining candidates need to be generated. This was accomplished using DFS and constraint satisfaction as will be explained later on. Finally, after the possible solutions are generated, the optimal solution should be selected and displayed on the screen.

Before the solution stage which includes most of the AI aspects of the program, candidate words need to be generated in the first stage. First step of candidate generation is scraping words from several sources using the given clues. In this stage, we used a custom method of scoring candidates. These scores were later used in the solution stage. The sources we used for this stage were merriam-webster, wikipedia and wordnet. The clues are searched as a whole in wikipedia, however for merriam-webster and wordnet we split the clues and search individual words. Otherwise, it is not possible to get reasonable sets of candidates from them except for

very short clues. When scraping words from these sources, we also give them a score based on two factors. For all sources, the closer the word is to the top of the page, the higher the score it has. We assumed that more relevant words come up before irrelevant words in general. For example, in a wikipedia page, the word we are looking for is most likely contained in the summary section at the top. For wikipedia, the page we found the word in also affects its score. We check several pages that are relevant to the clue, and here we assumed that the pages that were recommended at the top were the most relevant ones. All scrapers score the words they generate individually according to the following formula: $\text{score} = (\text{word_count} - \text{index}) / \text{word_count}$ where word_count is the total number of words scraped and index is the index of the given word starting from the top of the page.

After the scraping process is complete, we process the acquired words to obtain our final candidate sets. First of all, all non-alphabetic characters are removed, stopwords are eliminated and all characters are made lowercase. Then, for each word we find its synsets in wordnet and add the synonyms to the candidate set with a score penalty (score of original word * 0.5). Then, all words that are not the required length for the given entry are eliminated. The remaining words give us the finalized candidate sets. We take the union of these finalized sets from different sources and if a word was scraped from multiple sources, we add the scores given by those sources. Here, the idea is that if a word is scraped by multiple sources, it is more likely to be the actual answer.

After the candidate generation stage is completed, we pass the final candidates to the solver and start the solution process. A depth first search tree will be used in order to generate possible solutions. However, before starting the search, an elimination of candidates based on constraint satisfaction is done in order to reduce time complexity. For each crossing/constraint in the puzzle, the pair of candidate sets that are related to the given constraint is compared. During this comparison, if in one candidate set, there exists a word such that none of the words in the other set can satisfy the constraint, we count that as a violation and add 1 to that word's violation count. After checking all

constraints and obtaining all violation counts, we remove all words that have a violation count above or equal to “k” from the candidate sets. We repeat this process until the iteration where no words are removed. In the final implementation, we selected “k” as 2 since setting it to 1 made the program remove the correct answers most of the time. Setting it higher than 2 makes the constraint reduction virtually useless as most of the words will remain after reduction. After this reduction, to further reduce the time complexity, we sort the remaining candidates based on their scores obtained from the scrapers and select the best “n” candidates for each clue. Here, “n” can grow depending on how fast the computer is, how the candidate sets are and how efficient the program is. In our implementation, we used 30.

Now that the unnecessary candidates have been eliminated, the root node is constructed and the search process begins. The states of the nodes in the tree are represented as “selections”. A selection is either empty or a single word, and each Node has a corresponding selection for each clue. When generating a child, for each selection that has not been made in the parent (empty selection), each candidate available for that selection corresponds to a new child. We iterate through these candidates and set them as selections for the new children. However, there is a very important condition that eliminates some of the candidates from being used for generation of new children. After expanding the root node, we only select candidates that satisfy at least some constraint. This prevents the tree from generating branches that have all across selections and no down selections, or vice versa. Note the depth of a Node in this tree corresponds to the number of selections made in that node. The root node with a depth of 0 has no selections. Whenever a new child is constructed, a reduction of candidates based on its current selections is made. This is a similar process to the previously mentioned constraint satisfaction method, however the constraints are only checked between a selection and a candidate set instead of two candidate sets. This allows the program to consider more options as it progresses in the tree. If the previous reduction method was used here, many branches - possibly including the ones that lead to the correct solution - could be lost.

Using DFS, we search through the whole tree and keep a list of leaves encountered. In the implementation of the DFS tree, a queue structure was used. We also keep a dictionary of previously seen nodes so that we do not expand the same node twice which would be unnecessary computation. First, the root node is added to the queue. Then, while the queue is not empty the following steps are taken in each iteration. One node is popped from the end of the queue, and its children are constructed. Non-leaf children which were not seen before (checked from the dictionary) are pushed to the beginning of the queue and added to the seen dictionary. Leaves are held in another dictionary to be used in the next stage.

Once all leaves are found and there is no node left to expand, we progress to the final stage. In this stage, we find the leaf with the highest score and return it as the proposed solution. The score of a leaf is calculated by summing up the scores of its selected words (the scores were obtained during the scraping stage as mentioned). Note that this score is highly correlated with the depth of a node but lower depth nodes can still have higher scores than higher depth nodes. After the proposed solution is acquired, the results are displayed on the screen.

3 Conclusion

In this project, we implemented an end-to-end program that scrapes the NYT mini-crossword puzzle, scrapes candidate answers from various sources and then proposes a solution to the puzzle. One of the most challenging aspects of this task was to generate a reasonable number of candidate words that is also likely to contain the correct answer for a given clue. Increasing the number of candidates we scraped made it more likely for the candidate sets to include the correct answers. However, as a side effect multiple solutions that did not include the correct answers but satisfied more constraints than the desired solution exist in the results of our search algorithm. This made it very hard to select the desired solution among the many options. To combat this, we tried scoring the words as explained previously, and using those scores for

elimination of candidates and solutions. However, the results were still not satisfactory for us.

Our approach to scoring words was not the most sophisticated but it did improve the results significantly. An NLP approach such as the one in [2] could have been used to further improve the scoring process.

Overall, this project gave us an idea of what is possible to accomplish using simple AI techniques. This opens up new opportunities especially combined with the machine learning techniques we have been seeing in other courses. We learned different approaches for solving puzzles using AI and cemented the knowledge we gained from the course. This whole process was challenging but also quite informative.

4 Screenshots

The proposed program's outputs are presented here, Note that the date on the screenshot is the time the screenshot was taken, not the date of the puzzle.

The screenshot displays the CROSSWIND application interface. On the left, the 'OFFICIAL SOLUTION' crossword grid is shown with the following text:

	1	2	3	4
	T	I	N	A
5	C	A	M	E
6	A	B	O	V
7	L	O	U	I
8	L	O	T	S

Below the grid, the text '06/05/2021 | 17:45:39 | CROSSWIND' is visible. In the center, the 'ACROSS' and 'DOWN' clues are listed:

ACROSS

- 1 Amy's co-host at four Golden Globes ceremonies
- 5 Animal that can drink over 30 gallons of water at a time
- 6 Upwards of
- 7 Vuitton of fashion
- 8 Tons and tons

DOWN

- 1 Not allowed
- 2 "See ya later!"
- 3 St. Kitts and ___ (island nation)
- 4 Brewpub orders
- 5 Get-together on Zoom or Hangouts

On the right, the 'PROPOSED SOLUTION' crossword grid is shown with the following text:

	1	2	3	4
	F	R	E	E
5	C	E	A	S
6	T	O	D	A
7	P	R	I	O
8	B	O	O	T

Figure 1: Sample Output

The screenshot displays the CROSSWIND application interface. On the left, the 'OFFICIAL SOLUTION' crossword grid is shown with the following text:

1	2	3	4	
S	T	A	G	
5	T	E	S	6
				A
7	U	P	T	O
8	F	I	R	S
	9	D	O	S
				E

Below the grid, the text '06/05/2021 | 0:26:25 | CROSSWIND' is visible. In the center, the 'ACROSS' and 'DOWN' clues are listed:

ACROSS

- 1 Antlered animal
- 5 Auto company headquartered in Silicon Valley
- 7 Model/actress Kate
- 8 With 9-Across, part one of the Moderna and Pfizer vaccines
- 9 See 8-Across

DOWN

- 1 Double ___ Oreos
- 2 Showing little enthusiasm
- 3 Houston baseball player
- 4 Luster for your lips
- 6 Poker payment

On the right, the 'PROPOSED SOLUTION' crossword grid is shown with the following text:

1	2	3	4	
S	E	T	S	
5	C	O	N	6
				I
7	S	A	R	A
8	A	F	O	R
	9	S	I	T
				E

Figure 2: Sample Output

CROSSWIND

OFFICIAL SOLUTION

	1	P	E	P	
4	B	O	G	E	5
6	A	L	G	A	E
7	D	O	O	R	S
	8	S	N	L	

09/05/2021 | 18:31:10 | CROSSWIND

ACROSS

1 Energy
4 Five on a par four, e.g.
6 Bottom of a pond food chain
7 "Light My Fire" band, with "the"
8 Show on which Kate McKinnon impersonates Hillary Clinton

DOWN

1 Preppy shirts
2 Goad
3 Find in an oyster
4 Scolding word to a dog
5 "Sure thing!"

PROPOSED SOLUTION

	1	G	A	S	
4	C			C	5
6	A	L	G	A	E
7	D	O	O	R	S
	8	S	E	E	

Figure 3: Sample Output

CROSSWIND

OFFICIAL SOLUTION

		1	A	D	3	O
	4	S	P	E	C	
5	M	A	R	C	H	
6	A	L	I	A	S	
7	Y	E	L	L		

09/05/2021 | 18:59:58 | CROSSWIND

ACROSS

1 "Much ___ About Nothing"
4 Design detail, for short
5 Martin Luther King Jr. led one on Washington
6 Spy's assumed name
7 Holler

DOWN

1 Ron's assistant on "Parks and Recreation"
2 Sticker on a car's back windshield
3 Arthur ___ Sulzberger Jr., publisher of The New York Times
4 50% off event
5 Possibly will

PROPOSED SOLUTION

		1	A	D	3	O
	4			E	C	
5	M	A	R	C	H	
6	A	L	I	A	S	
7	Y	O	W	L		

Figure 4: Sample Output

CROSSWIND

OFFICIAL SOLUTION

		1	V	2	A	3	C
	4	E	I	R	E		
5	E	D	S	E	L		
6	R	I	O	T	S		
7	G	E	R	E			

09/05/2021 | 19:05:30 | CROSSWIND

ACROSS

1 Carpet cleaner, for short
4 Dublin's land
5 Ford failure of the late 50's
6 Takes to the streets
7 "Chicago" actor Richard

DOWN

1 Golfer's headgear
2 Glacial ridge
3 Cartoon frames
4 Actress Falco of "Nurse Jackie"
5 Fraction of a joule

PROPOSED SOLUTION

		1	W	2	E	3	E
	4	P		S	N		
5	X	A		K	D		
6	F	R	E	E	S		
7	S	T	A	R			

Figure 5: Sample Output

CROSSWIND

OFFICIAL SOLUTION

1	W	2	A	3	L	4	D	5	O
6	O	R	C	A	S				
7	M	E	D	I	A				
8	E	N	T	R	Y				
9	N	A	V	Y					

09/05/2021 | 19:11:48 | CROSSWIND

ACROSS

1 Hard-to-find guy in a crowd
6 Killer whales
7 Social ____
8 Contest submission
9 Dark blue shade

DOWN

1 Bathroom door sign
2 Basketball venue
3 Samsung purchase (warning: no vowels)
4 What the lactose intolerant avoid
5 "____, can you see ..."

PROPOSED SOLUTION

1	W	2	I	3	G	4	H	5	T
6	A	L	I	V	E				
7	T	W	O	E	S				
8	E	V	E	N	T				
9	R	A	C	Y					

Figure 6: Sample Output

CROSSWIND

OFFICIAL SOLUTION

	1	2	3	
	T	H	Y	
4	S	H	E	5
6	H	E	L	S
7	I	S	L	S
8	M	E	A	N
				T

09/05/2021 | 19:34:31 | CROSSWIND

ACROSS

1 Your, in the Bible
4 *They're counted at bedtime
6 *Positions of leadership
7 *Capri and Wight
8 Was serious about

DOWN

1 "One of ___ days ..."
2 Very, slangily
3 Country south of Saudi Arabia
4 Carpenter's thin wedge
5 Sound made by a cheater during a test

PROPOSED SOLUTION

	1	2	3	
	S	V	R	
4	A	O	A	5
6	L	L	L	E
7	S	A	I	S
8	O	R	D	M
				T

Figure 7: Sample Output

CROSSWIND

OFFICIAL SOLUTION

1	2	3	4	
S	H	O	P	
5	E	U	R	O
6	A	N	G	U
	8	D	A	7
				S
	9	O	N	
				X

09/05/2021 | 18:47:52 | CROSSWIND

ACROSS

1 Hit the mail
5 Capital of 19 countries
6 Breed of black cattle
8 What the cloud stores
9 Black gemstone

DOWN

1 Mermaid's home
2 "___ P" (100%, in slang)
3 Heart or lung
4 In a sullen mood
7 Jazz instrument, for short

PROPOSED SOLUTION

1	2	3	4	
S	H	O	P	
5	P	A	R	T
6	L	I	G	7
		H		T
	8	M	A	S
				S
	9	N		

Figure 8: Sample Output

CROSSWIND

OFFICIAL SOLUTION

1	R		2	E		3	P		4	S	
5	O		B		A		M		6	A	
7	M		I		L		A		N		
8	A		L		E		R		T		
			9	L		O		T	S		

ACROSS

1 Count at the gym
5 Springsteen's podcast partner on "Renegades : Born in the U.S.A."
7 Italian fashion hub
8 Watchful for possible danger
9 A ton

DOWN

1 Tomato type
2 Online way to pay utilities
3 Kind of diet that mimics cavemen
4 Modern lead-in to phone, TV and even refrigerator
6 ___ on a log (celery snack)

PROPOSED SOLUTION

1	M		2	A		3	R		4	K	
5	I				A				6	M	
7	L				N				I		
8	L		A		K		E		S		
			9	U		S		E	S		

09/05/2021 | 21:46:06 | CROSSWIND

Figure 9: Sample Output

CROSSWIND

OFFICIAL SOLUTION

1	D		2	I		3	C		E	
	O			4	Y		A		5	M
6	Z		7	E		B		R	A	
8	E		N		E				Y	
			9	O		R		E	O	

ACROSS

1 Black-and-white equipment
4 Thanksgiving side dish
6 Black-and-white animal
8 Opposite of WSW
9 Black-and-white snack

DOWN

1 Catch some Z's
2 Internet prefix
3 Pointy part of Mr. Spock
5 BLT condiment
7 Rock musician Brian

PROPOSED SOLUTION

1	G		2	R		3	A		Y	
	L			4					5	F
6	O		7	V		E		R	O	
8	M		M		E				O	
			9	F		O		O	D	

09/05/2021 | 21:02:00 | CROSSWIND

Figure 10: Sample Output

CROSSWIND

OFFICIAL SOLUTION

	1	B	2	A	3	T	4	H
	5	O		D		I		E
6	P		O	O		C		H
7	A		L		P		O	
8	C		A		T		S	

09/05/2021 | 20:49:18 | CROSSWIND

ACROSS

1 Hard thing to give a pet
5 Dog in the comic strip "Garfield"
6 Doggie
7 Pet food brand
8 Scratching post users

DOWN

1 When repeated, Yale student's cheer
2 Take in from a pet shelter
3 Kiwis : New Zealanders :: ____ : Costa Ricans
4 Snickering sound
6 "Super" campaign funder

PROPOSED SOLUTION

	1		2	C	3		4	W
	5		A		L		S	O
6	P			A				W
7	A		L		S		O	
8	C		O		P		Y	

Figure 11: Sample Output

CROSSWIND

OFFICIAL SOLUTION

		1	D	2	O	3	T	
4	A	5	L		A	M	O	
6	S		E		I		N	E
7	P		A		R		I	S
8	S		K		Y			

09/05/2021 | 23:22:11 | CROSSWIND

ACROSS

1 When repeated three times, an ellipsis
4 Car rental agency
6 River through 7-Across
7 Climate agreement city
8 Beautiful view during a sunset

DOWN

1 Grocery aisle with milk, cheese and butter
2 All: Prefix
3 Ten below?
4 Egyptian vipers
5 Unauthorized disclosure

PROPOSED SOLUTION

		1	D	2	A	3	Y	
4	S	5	L		A	N	G	
6	C		L		I		F	F
7	C		O		L		O	R
8	S		A		Y			

Figure 12: Sample Output

CROSSWIND

OFFICIAL SOLUTION

1	P	L	A	N	T
6	H	O	N	O	R
7	I	T	S	M	E
8	S	T	E	A	K
9	H	O	L	D	S

09/05/2021 | 23:30:13 | CROSSWIND

ACROSS

1 Test of responsibility before a pet or kid
6 Word before student or system
7 First line on the phone to someone you know well
8 Rare order at a restaurant
9 Waits on the phone

DOWN

1 Jam band fronted by guitarist Trey Anastasio
2 Scratch-off ticket game
3 "Moon And Half Dome" photographer Adams
4 Wanderer
5 Arduous journeys

PROPOSED SOLUTION

1	E	X	A	C	T
6	R	U	M	O	R
7	P	I	E	C	E
8	S	T	E	A	K
9	O	P	E	N	S

Figure 13: Sample Output

Bibliography

- [1] Littman, M.L., Keim, G.A., Shazeer, N., 2002. "A probabilistic approach to solving crossword puzzles." *Artificial Intelligence* 134, 23–55. [Online].
Available: <https://www.sciencedirect.com/science/article/pii/S000437020100114X>
doi:10.1016/S0004-3702(01)00114-X. [Accessed: 09-May-2021].

- [2] Thanasuan, K. and Mueller, S. T., "Crossword expertise as recognitional decision making: an artificial intelligence approach." *Frontiers*, 26-Aug-2014. [Online].
Available: <https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01018/full>.
[Accessed:09-May-2021].

- [3] Jobin, A., Menon, A., Sekhar, A., Damodaran, V., 2017. "Key to crossword solving : NLP." 929-933. 10.1109/ICACCI.2017.8125960. [Online].
Available: <https://ieeexplore.ieee.org/document/8125960>
[Accessed: 07-May-2021].

Appendix

```
from functools import cmp_to_key
from nltk.corpus import words
import copy
import re
import sys
from nltk.corpus import wordnet as wn

# global variables
all_words = wn.all_lemma_names()
word_scores = None

# This Node class is the main component of our search tree
# it contains the state of a solution which includes:
# current selections, remaining candidates, and depth of node
# constraints are also propagated through the tree for ease of coding,
# however constraints are passed by reference and each node is looking
# at the same constraints, these constraints are never modified
class Node():
    # A simple init function that stores the given values and does
    # reduction based on current selections and constraints
    def __init__(self, candidates, selections, depth, constraints):
        self.candidates = candidates
        self.depth = depth
        self.constraints = constraints
        self.selections = selections
        self.set_selections()
        self.set_constraint_counts()

    # this removes all words from candidates that do not obey the constraints
    # from the current selections
    def reduce(self):
        self.candidates = reduce_by_selections(self.candidates,
                                              self.selections,
                                              self.constraints)

    # this method is used to find any new implicit selections other than
    # the one given by parent
```

```

# an implicit selection would be a candidate list that has only one member
# left due to reduction by constraints
# each such implicit selection is set to the actual selections
# and reduction is done again
def set_selections(self):
    self.reduce()

    for i, selection in enumerate(self.selections[0]):
        if len(selection) == 0:
            if len(self.candidates[0][i]) == 1:
                self.selections[0][i] = list(self.candidates[0][i].keys())[0]
                self.reduce()

    for i, selection in enumerate(self.selections[1]):
        if len(selection) == 0:
            if len(self.candidates[1][i]) == 1:
                self.selections[1][i] = list(self.candidates[1][i].keys())[0]
                self.reduce()

# find the number of constraints each candidate will have to obey
# due to the current selections
# it is later used to decide which children will be generated
# see Node.get_children()
def set_constraint_counts(self):
    self.has_constraints = False
    self.constr_counts = ([0,0,0,0,0],[0,0,0,0,0])

    acc_selections = self.selections[0]
    down_selections = self.selections[1]

    for constraint in self.constraints:
        acc_idx = constraint[0]
        down_idx = constraint[2]

        acc_selection = acc_selections[acc_idx]
        down_selection = down_selections[down_idx]

        if len(acc_selection) > 0:
            self.has_constraints = True
            self.constr_counts[1][down_idx] += 1

```

```

        if len(down_selection) > 0:
            self.has_constraints = True
            self.constr_counts[0][acc_idx] += 1

    def is_leaf(self):
        for candidate_map_list in self.candidates:
            for candidate_map in candidate_map_list:
                if len(candidate_map) > 1:
                    return False

        return True

    def get_hashable(self):
        keys = []

        for candidate in self.candidates:
            for cand_map in candidate:
                keys += cand_map.keys()

        return hash(frozenset(sorted(keys)))

    def get_child(self, direction, index, word):
        # direction: 0->across, 1->down
        candidates = copy.deepcopy(self.candidates)
        candidates[direction][index] = {word: 1.0}
        selections = copy.deepcopy(self.selections)
        selections[direction][index] = word
        depth = self.depth + 1

        return Node(candidates, selections, depth, self.constraints)

    def get_children(self):
        acc_candidates = self.candidates[0]
        down_candidates = self.candidates[1]

        children = []

        # if the node is a leaf, return empty list to terminate
        # since it wont have any children

```

```

if self.is_leaf():
    return children

# generate children for all possible accross candidates
for i, cand_map in enumerate(acc_candidates):
    # if the new selection wont be obeying any constraints
    # we do not need to check that route since it will lead
    # to nodes with all accross selections
    # which is not desired
    # therefore we skip children that does not obey any new
    # constraints
    if self.has_constraints:
        if self.constr_counts[0][i] < 1:
            continue
        pass

    if len(cand_map) <= 1:
        continue

    for word in cand_map.keys():
        children.append(self.get_child(0, i, word))

# generate children for all possible down candidates
for i, cand_map in enumerate(down_candidates):
    # if the new selection wont be obeying any constraints
    # we do not need to check that route since it will lead
    # to nodes with all down selections
    # which is not desired
    # therefore we skip children that does not obey any new
    # constraints
    if self.has_constraints:
        if self.constr_counts[1][i] < 1:
            continue
        pass

    if len(cand_map) <= 1:
        continue

    for word in cand_map.keys():
        children.append(self.get_child(1, i, word))

```

```

    return children

def get_score(self):
    score = 0.0
    for selection in self.selections:
        for word in selection:
            if len(word) > 0:
                score += 0
            if word in word_scores:
                score += word_scores[word]

    return score

# This is a function used to sort leafs at the end in order to make
# the best guess
# leafs are sorted by their scores, for more info see Node.get_score()
def node_compare(node_1, node_2):
    if node_1.get_score() > node_2.get_score():
        return -1
    elif node_1.get_score() == node_2.get_score():
        return 0
    else:
        return 1

# given candidates, current selections and all constraints
# remove all candidates that do not obey the constraints of the current selections
def reduce_by_selections(candidates, selections, constraints):
    acc_selections = selections[0]
    down_selections = selections[1]

    acc_cands = candidates[0]
    down_cands = candidates[1]

    for constraint in constraints:
        acc_idx = constraint[0]
        down_idx = constraint[2]

        acc_at = constraint[1]

```

```

down_at = constraint[3]

acc_selection = acc_selections[acc_idx]
down_selection = down_selections[down_idx]

acc_len = len(acc_selection)
down_len = len(down_selection)

to_be_removed = []

if acc_len > 0 and down_len > 0:
    continue

if acc_len > 0:
    letter = acc_selection[acc_at]
    for candidate in down_cands[down_idx].keys():
        if candidate[down_at] != letter:
            to_be_removed.append(candidate)
    for word in to_be_removed:
        down_cands[down_idx].pop(word, None)

if down_len > 0:
    letter = down_selection[down_at]
    for candidate in acc_cands[acc_idx].keys():
        if candidate[acc_at] != letter:
            to_be_removed.append(candidate)
    for word in to_be_removed:
        acc_cands[acc_idx].pop(word, None)

return (acc_cands, down_cands)

# helper function of get_letter_counts
# if the word is n characters, returns a list of n dictionaries
# i'th dictionary hold the number of letters seen for the i'th character
# the dictionary has letters as keys, and number of letters seen as values
def get_letter_count_of_word_list(word_list):
    counts = []

    if len(word_list) == 0:
        length = -1

```

```

else:
    length = len(word_list[0])

for i in range(length):
    count = {}
    for word in word_list:
        letter = word[i]
        if letter in count:
            count[letter] += 1
        else:
            count[letter] = 1
    counts.append(count)

return counts

# finds the letter counts for all candidate lists
# we will have 10 candidate lists, 5 for accross, 5 for down
# if a candidate list contains words of length n, its letter counts will
# consist of n dictionaries, see get_letter_count_of_word_list(word_list)
# these letter counts are then used for reduction based on constraints
def get_letter_counts(candidates):
    acc_candidates = candidates[0]
    down_candidates = candidates[1]

    acc_counts = []
    down_counts = []

    for acc_cand_map in acc_candidates:
        word_list = list(acc_cand_map.keys())
        acc_counts.append(get_letter_count_of_word_list(word_list))

    for down_cand_map in down_candidates:
        word_list = list(down_cand_map.keys())
        down_counts.append(get_letter_count_of_word_list(word_list))

    return acc_counts, down_counts

def reduce_by_constraints(candidates, letter_counts, constraints):
    # k is the number of constraints a candidate needs to violate in order to
    # be removed

```

```

# setting k higher than 1 increases number of remaining candidates
# which can help keep the correct answers in the initial list
# however, it also keeps more of wrong candidates which might lead to
# wrong answers
k = 2

acc_candidates = candidates[0]
down_candidates = candidates[1]

acc_counts = letter_counts[0]
down_counts = letter_counts[1]

# this flag is used to keep track of changes
# so that if any candidate was removed, constraint checking is done again
# if there are no candidates removed in an iteration, then we know
# there is no more reduction to be done and quit the loop
flag = True

while (flag):
    # lists of candidates to be removed
    # we need this list because we cannot remove candidates while
    # iterating over them
    to_be_removed = ([{}], [{}], [{}], [{}], [{}], [{}], [{}], [{}], [{}], [{}])

    flag = False

    for constraint in constraints:
        acc_idx = constraint[0]
        down_idx = constraint[2]

        acc_at = constraint[1]
        down_at = constraint[3]

        acc_candidate = acc_candidates[acc_idx]
        down_candidate = down_candidates[down_idx]

        acc_count = acc_counts[acc_idx]
        down_count = down_counts[down_idx]

        for candidate in acc_candidate.keys():

```



```

letter = candidate[acc_at]
count = 0
if len(down_count) > 0 and letter in down_count[down_at]:
    count = down_count[down_at][letter]

if len(down_candidate) > 0 and count == 0:
    if candidate in to_be_removed[0][acc_idx]:
        to_be_removed[0][acc_idx][candidate] += 1
    else:
        to_be_removed[0][acc_idx][candidate] = 1

for candidate in down_candidate.keys():
    letter = candidate[down_at]
    count = 0
    if len(acc_count) > 0 and letter in acc_count[acc_at]:
        count = acc_count[acc_at][letter]
    if len(acc_candidate) > 0 and count == 0:
        if candidate in to_be_removed[1][down_idx]:
            to_be_removed[1][down_idx][candidate] += 1
        else:
            to_be_removed[1][down_idx][candidate] = 1

# remove the candidates that were set to be removed
for direction in range(2):
    for i in range(5):
        removing = to_be_removed[direction][i]
        for word in removing:
            if to_be_removed[direction][i][word] >= k:
                flag = True
                candidates[direction][i].pop(word, None)
                for j, letter in enumerate(word):
                    letter_counts[direction][i][j][letter] -= 1

return candidates, letter_counts

# uses either DFS or BFS method to find all leaves from the given root node
# note that it does not stop at some leaf, instead it generates all leaves
# and returns them all
def get_leafs(root):

```

```

# the method for searching can be selected here
# BFS lets us see how the stack is growing
# DFS is going to take less memory and hence be faster
# however, their time complexities would be the same in this situation
method = "bfs" # "bfs"

# initialize the stack and list of found leaves
stack = [root]
leafs = []

# these dicts are used to store previously seen nodes
# they are used to prevent expansion of the same node twice
# if it is acquired through some other path before
# hence they reduce the total runtime
seen_dict = {}
leaf_dict = {}

# while nodes are remaining in the stack
while len(stack) > 0:
    #print(len(stack))

    # pop and expand a node
    node = stack.pop()
    node_hashable = node.get_hashable()
    children = node.get_children()

    # add the leaves to leaves list
    if len(children) == 0 and node_hashable not in leaf_dict:
        leaf_dict[node_hashable] = None
        leafs.append(node)

    # add the non-leaves to the stack to later be expanded
    for child in children:
        child_hashable = child.get_hashable()
        if child_hashable not in seen_dict:
            seen_dict[child_hashable] = None
            if method == "dfs":
                stack.append(child)
            elif method == "bfs":
                stack.insert(0, child)

```

```

        else:
            print("\nSearch method is invalid, exiting:")
            sys.exit()

    return leafs

def get_top_k_helper(cands, k):
    for key in cands.keys():
        cands[key] = word_scores[key]

    return dict(sorted(cands.items(), key=lambda item: item[1])[-k:])

def get_top_k(candidates, k):
    for i, cand_map_list in enumerate(candidates):
        for j, cand_map in enumerate(cand_map_list):
            candidates[i][j] = get_top_k_helper(cand_map, k)

    return candidates

# this is the main function to be called from somewhere else
# it takes in a puzzle which holds clues and constraints
# constraints are a list of tuples that store which entries have crossings
# at which indices
# candidates contains two lists of dictionarites
# each list contains 5 dictionaries and each of these dictionaries holds
# the candidates for the corresponding entry
# the function then calls all the other helper functions to find a solution
# print it on the console and return it so it can be shown on the gui
def solve_for(puzzle, candidates):
    constraints = puzzle.get_constraints()

    # store the word scores globally so they can later be used for ranking
    # possible solutions
    global word_scores
    word_scores = get_word_scores(candidates)

    #print_candidates(candidates, puzzle.clues)

    # reduce the candidates by constraints

```

```

letter_counts = get_letter_counts(candidates)
# candidates, letter_counts = reduce_by_constraints(candidates, letter_counts,
constraints)

candidates = get_top_k(candidates, 30)
print_candidates(candidates, puzzle.clues)
# sys.exit()

print("\nStarting Solving, The candidates we currently have:")
print()
# set the initial selections as empty
# these selections will be held by each node in the search tree
# and will be making up the state of a node
selections = [["", "", "", "", ""], ["", "", "", "", "]]

# construct root node and get the leafs of the tree
# each leaf will be a possible solution, the best scored leaf will
# be returned as the proposed solution
root = Node(candidates, selections, 0, constraints)
leafs = get_leafs(root)
leafs = sorted(leafs, key=cmp_to_key(node_compare))

# printing the best 10 leafs and their scores for tracing purposes
print("\n%d leaves were found: " % (len(leafs)))
print("Following were the 10 leaves with highest scores: ")
for i, leaf in enumerate(leafs):
    print(leaf.selections)
    print(leaf.get_score())
    if i > 10:
        break

# print the best guess and return it
print("\nBest Guess:")
print(leafs[0].selections)
# sys.exit()
return leafs[0].selections

## Helper methods

```

```

def get_word_scores(candidates):
    word_scores = {}
    for cand_map_list in candidates:
        for cand_map in cand_map_list:
            for word in cand_map:
                word_scores[word] = cand_map[word]

    return word_scores

def print_candidates(candidates, clues):
    acc_cands = candidates[0]
    down_cands = candidates[1]
    acc_clues = clues[0]
    down_clues = clues[1]

    for i, cands in enumerate(acc_cands):
        print()
        print("Clue: %s" % (acc_clues[i]))
        print(cands)

    for i, cands in enumerate(down_cands):
        print()
        print("Clue: %s" % (down_clues[i]))
        print(cands)

```

This project report includes work done in partial fulfillment of the requirements for CS 461 -- Artificial Intelligence. The software in the appendix is, to a large extent, original (with borrowed code clearly identified) and was written solely by members of CROSSWIND.

Word Count: 2006