CS532 Homework 8

Archana Machireddy

Question 1

def print\_tree(self,indent=""):

print("%s%s"%(indent,self.value))

next = self.child

while next is not None:

ret = next.print\_tree(indent+" "\*(len(self.value)))

if ret is not None:

return ret

next = next.sibling

return None

Question 2

def split\_node(self,prefix):

b = self.value[len(prefix):]

self.value = prefix

c = self.child

self.child = chars(b)

self.child.parent = self

self.child.child = c

if c != None:

self.child.child.parent = self.child

next = self.child.child.sibling

while next is not None:

next.parent = self.child

next = next.sibling

Question 3

def find\_partial(self,string,indent=""):

if string.find(self.value) != 0:

return (None,string)

string = string[len(self.value):]

last = self

if string == "":

return (self,string)

next = self.child

while next is not None:

ret,sub = next.find\_partial(string,indent+" ")

if ret is not None:

return (ret,sub)

next = next.sibling

return (last,string)

Question 4

def add\_word(self,word):

node, sub = self.find\_partial(word)

next = node.child

while next is not None:

com = self.common\_sub(next.value,sub)

if com != "":

next.split\_node(com)

if sub[len(com):] != "":

next.add\_child(chars(sub[len(com):]))

self.set\_word(next,sub[len(com):])

else:

next.word = True

return

else:

next = next.sibling

if sub != "":

node.add\_child(chars(sub))

self.set\_word(node,sub)

return

def set\_word(self,next,sub\_added):

node1 = next.child

while node1 is not None:

if node1.value == sub\_added:

node1.word = True

break

else:

node1 = node1.sibling

def print\_tree(self,indent=""):

if self.word == False:

print("%s%s"%(indent,self.value))

else:

print("%s%s W"%(indent,self.value))

next = self.child

while next is not None:

ret = next.print\_tree(indent+" "\*(len(self.value)))

if ret is not None:

return ret

next = next.sibling

return None

def split\_node(self,prefix):

b = self.value[len(prefix):]

self.value = prefix

c = self.child

self.child = chars(b)

self.child.parent = self

if self.word == True:

self.child.word = True

self.child.child = c

if b != "":

self.word = False

if c != None:

self.child.child.parent = self.child

next = self.child.child.sibling

while next is not None:

next.parent = self.child

next = next.sibling

def \_\_init\_\_(self,value):

self.parent = None

self.child = None

self.sibling = None

self.value = value

self.word = False

Question 5

def create\_trie():

l = lexicon()

start = ""

top = chars("")

while True:

words = l.next5(start)

if len(words) < 1:

break

for w in words:

top.add\_word(w)

start = words[-1]

return top

def tokenize(trie,string):

queue = [(0,[])]

str\_len = len(string)

answers = []

while queue:

start, tokens = queue.pop(0)

print(start,tokens)

if start == str\_len:

answers.append(tokens)

else:

node, rem = trie.find\_partial(string[start:])

done = len(string)-len(rem)

found = string[start:done]

if tokens == []:

new\_tokens=[found]

else:

new\_tokens = tokens+[found]

if node.word == True:

queue.append((done,new\_tokens))

print('Found word %s'%(found))

print('Adding to frontier %s %i'%(found,done))

next = node.parent

l = len(node.value)

while next is not None:

new\_l = len(found)-l

new\_ll = done - l

if next.word == True:

if tokens == []:

new\_tokens=[found[0:new\_l]]

else:

new\_tokens = tokens + [found[0:new\_l]]

queue.append((new\_ll,new\_tokens))

print('Adding to frontier %s %i'%(found[0:new\_l],new\_ll))

else:

print(' Node %s is not a word'%(found[0:new\_l]))

l = l+len(next.value)

next = next.parent

return answers