

## Finding or Building a Corpus Report

1. Describe how you created your knowledge base, include screen shots of the knowledge base, and indicate your top 10 terms

First, I chose a starter url to begin my web scraping. I decided to use the Wikipedia page of the Hamilton Musical as it is my favorite musical. Then I wrote a web crawler function that found 20 relevant URLs. In this function, I ensured that the URLs that were found had pages that were outside the Wikipedia domain, were not video links, started with 'http' and contained the word 'Hamilton' .

Then, I scraped all the text from these URLs and cleaned up the text. I cleaned up the text by removing newlines, tabs, carriage returns, as well as any sentences that were encapsulated in <>. I also removed all the sentences that started with "END " because those were unnecessary. After cleaning up the text, I combined all of the clean text from all the URLs and tokenized it. I preprocessed the tokens further to remove all stop words and non-alpha words and also converted all the tokens to lower case. I then found the count of each token in the text and found the 40 words with the highest count:

Order	Word	Count
1.	hamilton	386
2.	miranda	312
3.	new	132
4.	show	104
5.	musical	95
6.	first	94
7.	like	82
8.	time	76
9.	broadway	75
10.	one	72
11.	also	70

12.	burr	69
13.	york	64
14.	theatre	59
15.	close	58
16.	washington	55
17.	going	54
18.	get	53
19.	story	50
20.	says	50
21.	would	50
22.	see	50
23.	people	50
24.	heights	49
25.	culture	49
26.	years	49
27.	tickets	48
28.	us	47
29.	music	46
30.	end	45
31.	still	45
32.	way	45
33.	may	44
34.	know	44

35.	america	42
36.	history	38
37.	site	37
38.	song	36
39.	last	35
40.	early	34

From the 40 words shown above, I chose what I thought were the 10 most relevant and important words:

1. hamilton
2. miranda
3. musical
4. burr
5. history
6. broadway
7. culture
8. washington
9. theatre
10. america

Then, I made a dictionary with these ten words as the keys. And the value of each key contained all the sentences from the combined text that included that specific word. That is what became my knowledge base. I then pickled that dictionary and stored it in a file.

Below are the first 5 sentences in the knowledge base from each word to provide an example:

#### 1. Hamilton

```

hamilton:
1) Google Tag Manager (noscript) End Google Tag Manager (noscript) Accessibility
links Navigation
Menu Daily
Featured Lin-Manuel Miranda
Talks 'Hamilton': Once A 'Ridiculous' Pitch, Now A Revolution April 9, 2016 8:24
AM ET Heard on NPR Staff
Lin-Manuel Miranda Talks 'Hamilton': Once A 'Ridiculous' Pitch, Now A Revolution 7:27
7:27 iframe src="https://www.

2) org/player/embed/473503407/473623599" width="100%" height="290" frameborder="0" scrolling="no"
title="NPR embedded audio player"> Lin-Manuel Miranda appears in
Hamilton's opening night at the Richard Rodgers Theatre in New York City, in August 2015.

3) Right now, Hamilton is the hardest ticket to get on Broadway.

4) The Grammy-winning show portrays the life of Alexander Hamilton, a founder of the United States who was
once a poor, orphaned boy "dropped in the middle of a forgotten spot of the Caribbean" – and it does so in
the rhymes and music of hip-hop and pop.

5) And then once you start hearing about Hamilton's life story, it sort of makes sense.

```

## 2. miranda

```
15 miranda:
16
17 1) Google Tag Manager (noscript) End Google Tag Manager (noscript) Accessibility
    Links Navigation
    Menu Daily
    Featured Lin-Manuel Miranda
    Talks 'Hamilton': Once A 'Ridiculous' Pitch, Now A Revolution April 9, 2016 8:24
    AM ET Heard on NPR Staff
    Lin-Manuel Miranda Talks 'Hamilton': Once A 'Ridiculous' Pitch, Now A Revolution 7:27
    7:27 iframe src="https://www.
18
19 2) org/player/embed/473503407/473623599" width="100%" height="290" frameborder="0" scrolling="no"
    title="NPR embedded audio player"> Lin-Manuel Miranda appears in
    Hamilton's opening night at the Richard Rodgers Theatre in New York City, in August 2015.
20
21 3) "I understand how ridiculous the elevator pitch for this show is," Lin-Manuel Miranda – the composer,
    lyricist and lead of the show – tells NPR's Scott Simon.
22
23 4) "That was what grabbed me about it," Miranda continues, "was this was a guy who used words to get
    everywhere and do what my favorite hip-hop artists do – if not write about their struggles, their lives,
    then transcend their circumstances by sheer virtuosity.
24
25 5) "Hamilton The Revolution by Lin-Manuel Miranda and Jeremy Mccarter Downloadable Audio, 287
    pages | Buy Featured Book Title Hamilton Subtitle The Revolution Author Lin-Manuel Miranda and
    Jeremy Mccarter Your purchase helps support NPR programming.
26
```

## 3. musical

```
27
28 musical:
29
30 1) Neilson Barnard/Getty Images The Broadway musical that's set during a revolution may
    have set off a revolution of its own, too.
31
32 2) In May 2009, at at the White House, Miranda was expected to perform a number from In the Heights , a
    musical running on Broadway at the time that he wrote and starred in.
33
34 3) "Interview Highlights McCarter, on his statement, when he was theater critic of New York Magazine ,
    that hip-hop could save musical theater Jeremy McCarter: Well, I'd grown up listening to hip-hop, and
    I'd grown up enjoying the theater, much as Lin had.
35
36 4) And we deal with it in a lot of ways, big and small, but it was an open thing and a thing to grapple
    with because it just was a way of life that no one knew how to deal with, and we grappled with it as much
    as we could – and still tell a story in 2 hours and 45 minutes in a musical.
37
38 5) On whether McCarter has any concern there will soon be a spate of copycat shows, like, hypothetically,
    Coolidge: The Musical McCarter: You know, I talked to Stephen Sondheim about this – not about the
    Coolidge musical specifically, although a Coolidge musical by Sondheim would be pretty amazing, I think.
39
```

## 4. burr

```
41 burr:
42
43 1) All that industry and drama unfolds in the shadow of a forgone conclusion: Hamilton will die in an 1804
    duel with Vice President Aaron Burr (Leslie Odom Jr.
44
45 2) Whereas Hamilton is about seizing the moment and taking a stand ("I am not throwing away my shot" is
    his mantra), Burr is a passive, taciturn opportunist whom life passes by.
46
47 3) Hamilton is the center, but Burr is his equally weighted Judas and Javert—and more complexly drawn than
    either.
48
49 4) Spoiler alert for the historically ignorant: It all crashes in 1804, when Hamilton agrees to a duel
    with then-Vice President Aaron Burr (Leslie Odom Jr.
50
51 5) All that industry and drama unfolds in the shadow of a forgone conclusion: Hamilton will die in an 1804
    duel with Vice President Aaron Burr (Leslie Odom Jr.
52
53
```

## 5. history

```
53 history:
54
55
56 1) A sublime conjunction of radio-ready hip-hop (as well as R&B, Britpop and trad showstoppers),
under-dramatized American history and Miranda's uniquely personal focus as a first-generation Puerto Rican
and inexhaustible wordsmith, Hamilton hits multilevel culture buttons, hard.
57
58 2) Follow David Cote on Twitter: +++++ Here is David
Cote's review of the Off Broadway premiere of Hamilton : Hamilton : Theater review by David Cote
History ticks to a syncopated beat in Lin-Manuel Miranda's jubilant, overflowingly rich Hamilton.
59
60 3) A sublime conjunction of radio-ready hip-hop (as well as R&B, Britpop and trad showstoppers),
under-dramatized American history and Miranda's uniquely personal focus as a son of immigrants and as an
inexhaustible wordsmith, Hamilton hits multilevel culture buttons, hard.
61
62 4) "Hamilton" is not a gimmicky transposition of early American history to a contemporary urban setting.
63
64 5) He sees a connection between Miranda's creation and the Henriad, Shakespeare's early cycle of history
plays: "What Lin is doing is taking the vernacular of the streets and elevating it to verse.
65
```

## 6. broadway

```
66
67 broadway:
68
69 1) Neilson Barnard/Getty Images The Broadway musical that's set during a revolution may
have set off a revolution of its own, too.
70
71 2) Right now, Hamilton is the hardest ticket to get on Broadway.
72
73 3) In May 2009, at at the White House, Miranda was expected to perform a number from In the Heights , a
musical running on Broadway at the time that he wrote and starred in.
74
75 4) You know what happens: Something like Hamilton comes along, it changes the sound of Broadway.
76
77 5) I love it like I love New York, or Broadway when it gets it right.
78
```

## 7. culture

```
80 culture:
81
82 1) A sublime conjunction of radio-ready hip-hop (as well as R&B, Britpop and trad showstoppers),
under-dramatized American history and Miranda's uniquely personal focus as a first-generation Puerto Rican
and inexhaustible wordsmith, Hamilton hits multilevel culture buttons, hard.
83
84 2) A sublime conjunction of radio-ready hip-hop (as well as R&B, Britpop and trad showstoppers),
under-dramatized American history and Miranda's uniquely personal focus as a son of immigrants and as an
inexhaustible wordsmith, Hamilton hits multilevel culture buttons, hard.
85
86 3) I keep telling everyone from our culture, 'You have no idea how lyrically amazing this show is from a
rap perspective.
87
88 4) I keep telling everyone from our culture, 'You have no idea how lyrically amazing this show is from a
rap perspective.
89
90 5) Deadpans Alex Balk, "On any of those rare occasions when you get a glimpse of someone wondering
whether Hamilton might not be the greatest thing that ever happened in 50,000 years of culture do you
find yourself stunned, surprised, or fearful for the heretic's safety?" Recommended Reading Spencer
Kornhaber Jo Boaler and Lang Chen Marina Koren But earlier this week, it appeared that backlash to the
musical may have finally arrived when a New York City lawyer a Hamilton ad encouraging "non-white"
performers to try out for lead parts.
91
```

## 8. ashington

```
93 washington:
94
95 1) "And it's right after the Washington Monument moment, and Washington hangs his head in shame and steps
96 back.
97 2) Assuming you don't know the basics, Hamilton is a (mostly) rapped-through biomusical about an orphan
98 immigrant from the Caribbean who came to New York, served as secretary to General Washington, fought
99 against the redcoats, authored most of the Federalist Papers defending the Constitution, founded the
100 Treasury and the New York Post and even made time for an extramarital affair that he damage-controlled
101 in a scandal-stanching pamphlet.
102
103 3) And yet the ambitious and brilliant young Hamilton emigrates north and becomes a successful lawyer,
104 General Washington's go-to aide and one of the Constitution's most eloquent interpreters, all the while
105 starting a family and weathering a sex scandal.
106
107 4) The cast in Thomas Kail's bravura, dance-filled staging is all-star: Christopher Jackson's Washington
108 is an affable demigod; Phillipa Soo's tremulous Eliza, the long-suffering wife to philandering Hamilton;
109 Daveed Diggs a deliciously conceited and fatuous Jefferson.
110
111 5) Assuming you don't know the basics, Hamilton is a (mostly) rapped-through biomusical about an orphan
112 immigrant from the Caribbean who came to New York, served as secretary to General Washington, fought
113 against the redcoats, authored most of the Federalist Papers defending the Constitution, founded the
114 Treasury and the New York Post and even made time for an extramarital affair that he damage-controlled
115 in a scandal-stanching pamphlet.
```

## 9. Theatre

```
106 theatre:
107
108 1) org/player/embed/473503407/473623599" width="100%" height="290" frameborder="0" scrolling="no"
109 title="NPR embedded audio player"> Lin-Manuel Miranda appears in
110 Hamilton's opening night at the Richard Rodgers Theatre in New York City, in August 2015.
111
112 2) 🙌 Awesome, you're subscribed! Thanks for subscribing! Look out for your first newsletter in your inbox
113 soon! Search More Hamilton Theater, Musicals Richard Rodgers Theatre, Midtown West Until Dec 31, 2022 5 out
114 of 5 stars Recommended Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10
115 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton
116 /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus
117 Hamilton /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10 Advertising Time
118 Out says 5 out of 5 stars Hamilton : Theater review by David Cote What is left to say? After Founding
119 Father Alexander Hamilton's prodigious quill scratched out 12 volumes of nation-building fiscal and
120 military policy; after Lin-Manuel Miranda turned that titanic achievement (via Ron Chernow's 2004
121 biography) into the greatest American in decades; after every critic in town (including me) the Public
122 Theater world premiere to high heaven; and after seeing this language-drunk, rhyme-crazy dynamo a second
123 time, I can only marvel: We've used up all the damn words.
124
125 3) Richard Rodgers Theatre ( ).
126
127 4) Sign in Hi,
128
129 Hamilton Theater Drama Critics' pick 2 0 Love It Share
130
131 Tweet 0 Shares 1/10 Photograph: Joan Marcus Hamilton 2/10 Photograph:
132 Joan Marcus Hamilton 3/10 Photograph: Joan Marcus Hamilton 4/10 Photograph:
133 Joan Marcus Hamilton 5/10 Photograph: Joan Marcus Hamilton 6/10 Photograph:
134 Joan Marcus Hamilton 7/10 Photograph: Joan Marcus Hamilton 8/10 Photograph:
135 Joan Marcus Hamilton 9/10 Photograph: Joan Marcus Hamilton 10/10 Photograph:
136 Joan Marcus Hamilton Richard Rodgers Theatre, Midtown West Until Thursday December 31
137 2015 Time Out says 5 out of 5 stars Details Dates And Times Users
138 say ( 2 ) 5 out of 5 stars Time Out says 5 out of 5
139 stars Hamilton : Theater review by David Cote What is left to say? After Founding Father
140 Alexander Hamilton's prodigious quill scratched out 12 volumes of nation-building fiscal and military
141 policy; after Lin-Manuel Miranda turned that titanic achievement (via Ron Chernow's 2004 biography) into
142 the greatest American in decades; after every critic in town (including me) the Public Theater world
143 premiere to high heaven; and after seeing this language-drunk, rhyme-crazy dynamo a second time, I can
144 only marvel: We've used up all the damn words.
145
146 5) -David Cote Richard Rodgers Theatre ( ).
147
```



## 10. america

```

119 america:
120
121 1) "When the White House calls and says, we'd love for you to perform, or if you have anything else on the
    American experience, and you have a hot 16 bars about Alexander Hamilton in your back pocket, my choice
    was clear," Miranda says.
122
123 2) 🙌 Awesome, you're subscribed! Thanks for subscribing! Look out for your first newsletter in your inbox
    soon! Search More Hamilton Theater, Musicals Richard Rodgers Theatre, Midtown West Until Dec 31, 2022 5 out
    of 5 stars Recommended Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10
    Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton
    /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus
    Hamilton /10 Photograph: Joan Marcus Hamilton /10 Photograph: Joan Marcus Hamilton /10 Advertising Time
    Out says 5 out of 5 stars Hamilton : Theater review by David Cote What is left to say? After Founding
    Father Alexander Hamilton's prodigious quill scratched out 12 volumes of nation-building fiscal and
    military policy; after Lin-Manuel Miranda turned that titanic achievement (via Ron Chernow's 2004
    biography) into the greatest American in decades; after every critic in town (including me) the Public
    Theater world premiere to high heaven; and after seeing this language-drunk, rhyme-crazy dynamo a second
    time, I can only marvel: We've used up all the damn words.
124
125 3) A sublime conjunction of radio-ready hip-hop (as well as R&B, Britpop and trad showstoppers),
    under-dramatized American history and Miranda's uniquely personal focus as a first-generation Puerto Rican
    and inexhaustible wordsmith, Hamilton hits multilevel culture buttons, hard.
126
127 4) And just as syncopation achieves its energizing effect by disturbing the expected flow, so Miranda's
    biomusical on founding father Alexander Hamilton is a rhythm-and-rhyme intervention for American
    iconography and ideology.
128
128
129 5) Sign in Hi,
    Hamilton Theater Drama Critics' pick 2 0 Love It Share
    Tweet 0 Shares 1/10 Photograph: Joan Marcus Hamilton 2/10 Photograph:
    Joan Marcus Hamilton 3/10 Photograph: Joan Marcus Hamilton 4/10 Photograph:
    Joan Marcus Hamilton 5/10 Photograph: Joan Marcus Hamilton 6/10 Photograph:
    Joan Marcus Hamilton 7/10 Photograph: Joan Marcus Hamilton 8/10 Photograph:
    Joan Marcus Hamilton 9/10 Photograph: Joan Marcus Hamilton 10/10 Photograph:
    Joan Marcus Hamilton Richard Rodgers Theatre, Midtown West Until Thursday December 31
    2015 Time Out says 5 out of 5 stars Details Dates And Times Users
    say ( 2 ) 5 out of 5 stars Time Out says 5 out of 5
    stars Hamilton : Theater review by David Cote What is left to say? After Founding Father
    Alexander Hamilton's prodigious quill scratched out 12 volumes of nation-building fiscal and military
    policy; after Lin-Manuel Miranda turned that titanic achievement (via Ron Chernow's 2004 biography) into
    the greatest American in decades; after every critic in town (including me) the Public Theater world
    premiere to high heaven; and after seeing this language-drunk, rhyme-crazy dynamo a second time, I can
    only marvel: We've used up all the damn words.
130

```

I outputted the first 5 sentences of each word in the knowledge base into a file using the following code:

```

import pickle

with open('knowledgeBase.pickle', 'rb') as handle:
    knowledgeDict = pickle.load(handle)

with open("KnowledgeBaseExample.txt", 'w') as f:
    for key in knowledgeDict:
        f.write('\n' + key + ": " + '\n\n')
        for i in range(5):
            f.write(str(i+1) + ") " + knowledgeDict[key][i] + ".\n")
        f.write('\n')

```

2. Write up a sample dialog you would like to create with a chatbot based on your knowledge base

### **Initial Dialog 1**

Hi, I'm a fan-bot of the Hamilton Musical. What is your name?

*I'm Hamna.*

Hi there Hamna! It's nice to meet you. Do you like Hamilton?

*I do. It's a good musical.*

Did you know that Hamilton was a poor, orphaned boy?

*Yes, I did.*

That's great! If you had to choose between Aaron Burr and George Washington, who would be your favorite character?

*Washington. I didn't really like Burr's character. He was mean.*

Washington is a good choice!

### **Initial Dialog 2**

Hi, I'm a fan-bot of the Hamilton Musical. What is your name?

*I'm Hamna*

Hi there Hamna! It's nice to meet you. Do you like Hamilton?

*Not really.*

I'm sorry to hear that! Did you know that Hamilton is a hip-hop musical?

*Yes, I did. I don't like hip-hop.*

I see. So you didn't like Hamilton because it has hip-hop?

*Yes.*

Did you know that Lin Manuel Miranda also wrote another musical, 'In the Heights'?

*No, I didn't know that he did anything besides Hamilton.*



