



Autocomplete and Lecture Keywords Presentation

Max Everest, Zhen Tan, and Nirjhar Deb
Jinwoo Park



Autocomplete Idea

Create an autocompleter that automatically fills in valid sentences as the user types:

- Begin with a stored set of many popular questions and queries
- Narrow down these initial questions for every letter the user types
- Weigh each entry; try to find the most helpful query
- Output the top couple of entries

We present the various methods we have come up with over the project. We have implemented and stored every method in Google Colab or our repository. The current method of interest is the SQL database method.



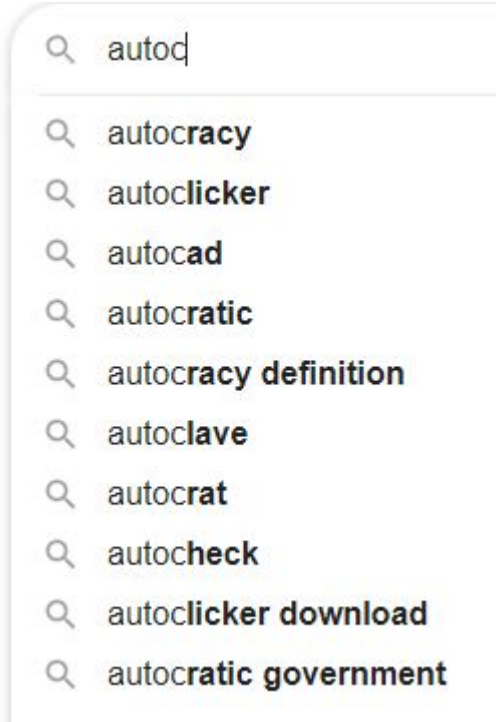
Initial Planning

Inspiration: Google's Autocompleter

- Fixes first few letters
- Outputs the best matches that start with these letters
- Uses pre-existing user queries to help predict new queries
- Works well with a large amount of stored queries

But we also need a dataset of pre-existing queries to work with.

We will use 6,000 questions generated from a ML T5 Model that read the entire textbook. Pretend these questions are pre-existing queries.

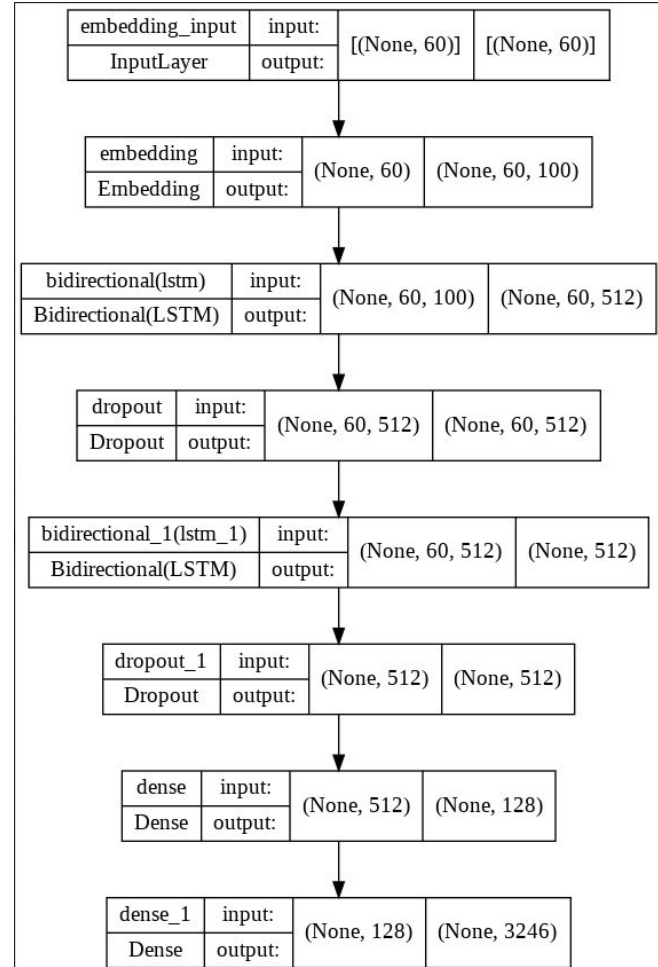
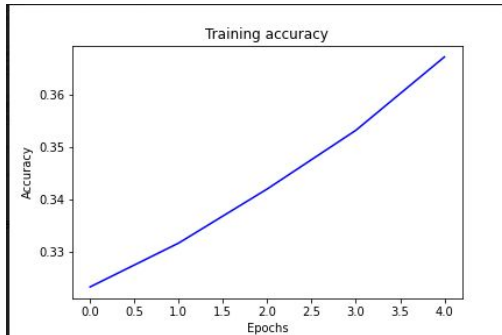


Search results for "autod":

- autod
- autocracy
- autoclicker
- autocad
- autocratic
- autocracy definition
- autoclave
- autocrat
- autocheck
- autoclicker download
- autocratic government

Long Short-Term Memory

- Trained on the generated questions based on the DSPFirst textbook
- Bidirectional layer
 - provide additional context to the network and result in faster and even fuller learning on the problem
- Word-level generation



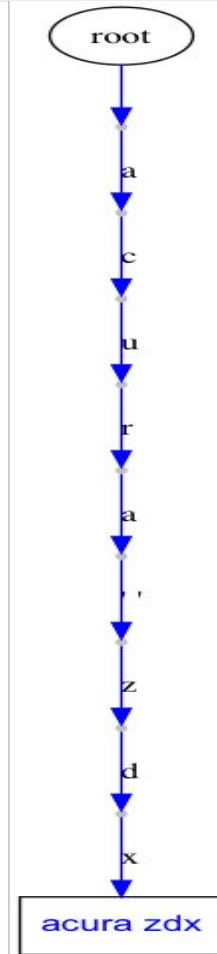
Directed Word Graph Method

- Thousands of keywords inserted into a graph
 - Extracted from generated questions and DSPFirst textbook using YAKE
- Each letter typed travels down the graph, narrowing the output
- Similar to a trie, but suggests keywords instead of full queries
- Context not considered: Previous words do not affect output

```
✓ [56] auto.set_string('what are sig')
```

```
✓ auto.get_suggestions()
```

```
[['sign'],  
 ['signs'],  
 ['signal'],  
 ["signal's"],  
 ['signify'],  
 ['signifies'],  
 ['significant'],  
 ['significance']]
```



0.015 sec / 0.016 sec

SQL Database Method

- Stored queries contain fields attached such as frequency
- Can search for both prefix matches and deep string matches
- Sort Order: Prefix matches sorted by length, then deep matches sorted by length.
- More dynamic than the trie; has better functionality
- Effectiveness limited by amount of stored queries

```
VIP-ITS: SELECT question, title, answer, frequency FROM questions WHERE question LIKE "what are signal%" X
```

question	title	answer	frequency
Filter...	Filter...	Filter...	Filter...
what are signals of different amplitudes, phases, and frequencies?	Chapter_2_Section_9	sinusoidal	1
what are signals of this class called?	Chapter_3_Section_9	FM signals	1
what are signals stored as in sp first lab exercises?	Chapter_-1_Section_1	vectors	1
what are signals that are audible to humans?	Chapter_2_Section_0	unknown	1

0.000 sec / 0.000 sec (times measured on an ACER Nitro 5 with a 4 core i5-8300H 2.3GHz CPU)

```
VIP-ITS: SELECT question FROM Questions WHERE INSTR(question, 'signal') > 0
```

question

signal

a discrete-time signal is a sequence of what?
a short segment of an orchestra music signal.
a signal may also be represented in terms of what?
as frequency increases, the signal has higher frequency oscillations.
as the frequency oscillations increase, the signal gets closer to what?
at what frequency does a signal begin?
audio cds use what sampling rate for storing music signals?
breaking the time-domain signal into two terms is called what?
can we hear that the variation causes the signal to fade?
determine the -transform of the input signal.
digital computers cannot deal with what kind of signals directly?
each delayed signal is multiplied by what?
filling in the signal values for in this example we obtain what?
functions serve as what for signals?
good examples of indefinite long signals are what?
how are certain sinusoidal input signals removed or nulled by a filter?
how can complicated signals be formed from relatively simple spectra?
how can fir filters be used to remove rapid fluctuations in signals?
how can one reconstruct the original bandlimited periodic signal for all?
how can periodic signals be synthesized?
how can real-world signals be represented?
how can sampling be used to obtain a discrete-dimensional signal?
how can the dft be used to analyze continuous-time periodic signals?
how can we analyze a signal mathematically?
how can we compute the transform of an infinite-duration signal?
how can we compute the values of a discrete-time signal?
how can we create a signal with quadratic angle function?
how can we form a third signal by subtraction?

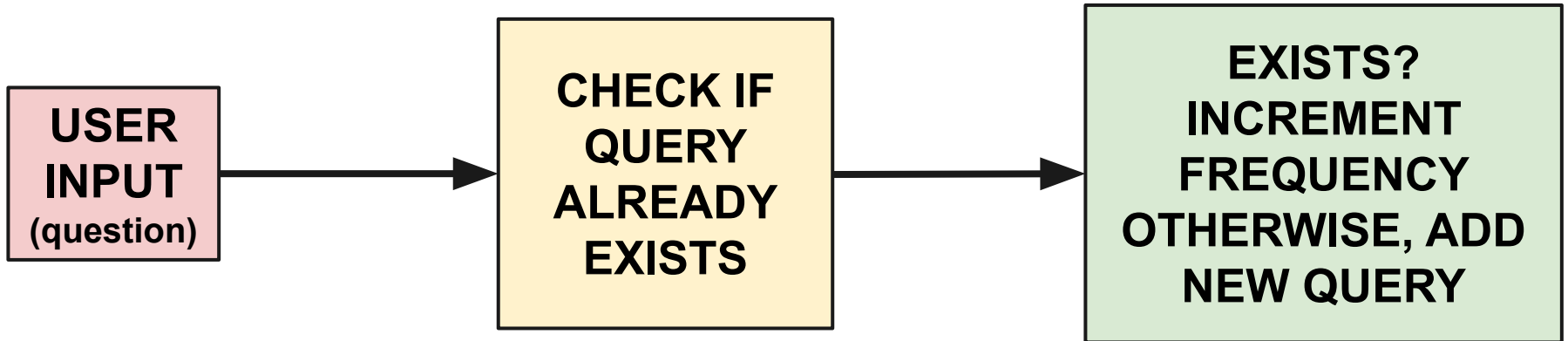
CONSOLE RE-RUN QUERY EXPORT 1-50 of 424

OPEN

count_id	question	title	answer	frequency
1	what is an extension of the real number system?	Chapter_-2_Section_0	A complex number system	1
2	what are complex numbers necessary to solve?	Chapter_-2_Section_0	equations	1
3	how many solutions does the previous equation have?	Chapter_-2_Section_0	two	1
4	what are numbers needed to solve for the two roots of a quadratic equation?	Chapter_-2_Section_0	complex numbers	1
5	whenever the discriminant is negative, the solution must be expressed as what?	Chapter_-2_Section_0	a complex number	1
6	how many different mathematical notations can be used to represent complex numbers?	Chapter_-2_Section_1	Several	1
7	what are the two basic types of polar form and rectangular form?	Chapter_-2_Section_1	unknown	1
8	converting between the two types quickly and easily is an important skill.	Chapter_-2_Section_1	unknown	1
9	rectangular form in rectangular form, all of the following notations define the same complex num...	Chapter_-2_Section_1	unknown	1
10	the ordered pair can be interpreted as what?	Chapter_-2_Section_1	a point in the two-dimensi...	1
11	a complex number can also be drawn as a vector whose tail is at the origin where?	Chapter_-2_Section_1	unknown	1
12	what is represented as a vector from the origin to?	Chapter_-2_Section_1	Complex number	1
13	what shows numerical examples of complex numbers?	Chapter_-2_Section_1	Figure~※	1
14	what are complex numbers plotted as vectors in the two-dimensional "complex plane"?	Chapter_-2_Section_1	Each is represented by a ...	1
15	each is represented by a vector from the origin to the point with what in the complex plane?	Chapter_-2_Section_1	coordinates	1
16	the point lies in the first quadrant of what?	Chapter_-2_Section_1	two-dimensional plane	1
17	where does the complex number notation represent the point in the two-dimensional plane?	Chapter_-2_Section_1	unknown	1
18	what is drawn as a vertical vector from the origin up to?	Chapter_-2_Section_1	unknown	1
19	what is another name for rectangular form?	Chapter_-2_Section_1	Cartesian form	1
20	what is the horizontal coordinate called?	Chapter_-2_Section_1	the real part	1
21	the vertical coordinate is called what?	Chapter_-2_Section_1	imaginary part	1
22	what is called the real part?	Chapter_-2_Section_1	The horizontal coordinate	1
23	what is the vertical coordinate called?	Chapter_-2_Section_1	imaginary part	1
24	the operators and are provided to extract what?	Chapter_-2_Section_1	real and imaginary parts of	1
25	in what form is the vector defined by its length and direction?	Chapter_-2_Section_1	polar	1



SQL Database Frequency



We will call this process when the user finishes typing and presses enter.



User Error

- All of these methods could be heavily impacted by any spelling errors or unusual formatting by the user.
- Thus, we have also implemented an auto correction algorithm that will pre-process all user input and correct any common misspellings and standardize formatting.

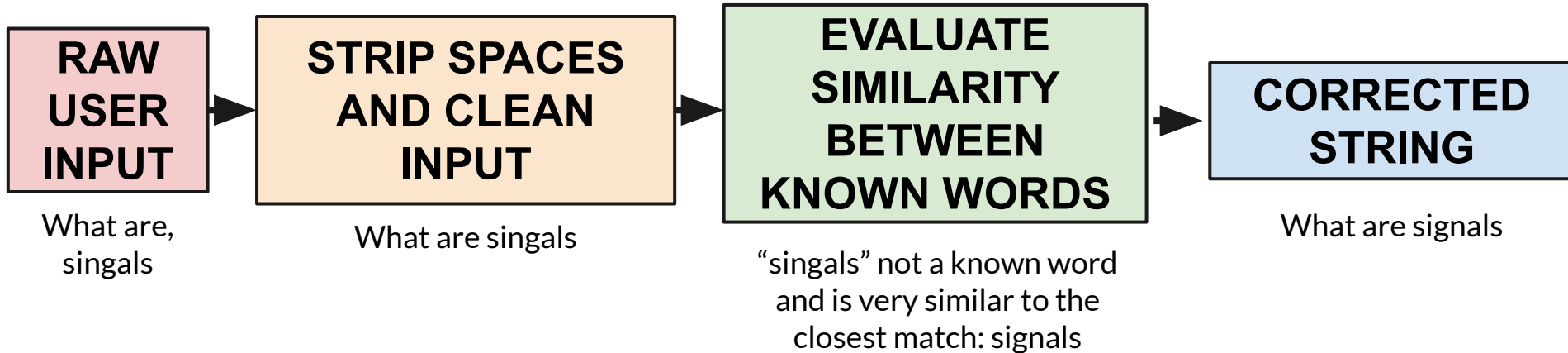
INPUT: what are, singals

AUTOCORRECT: 'what are, singals' CORRECTED TO 'what are signals'

```
['what are signals of different amplitudes, phases, and frequencies?',  
'what are signals of this class called?',  
'what are signals stored as in sp first lab exercises?',  
'what are signals that are audible to humans?']
```



Input/Output Flow of Autocorrection





Next Steps

- Design a function that will incorporate frequency and length among other factors simultaneously when finding the best matches
- Integrate the backend with a front end search interface (see next slide)
- Set up a free cloud SQL server to unify all additions and queries
- Package the project into the main chatbot repository
- Perhaps try more advanced metrics like Hamming distance or F1 score



React Search Bar

Search test

You can use `firpm` only once by inputting the `f` correctly. You can find an example of the bandpass filter design in the Matlab documentation for `firpm`.

what is a filter?

ENTER

LSTM Model (Per Word) ▾

what is a filter designed

what is a filter that

what is a filter to

what is a filter of

what is a filter design

Demo



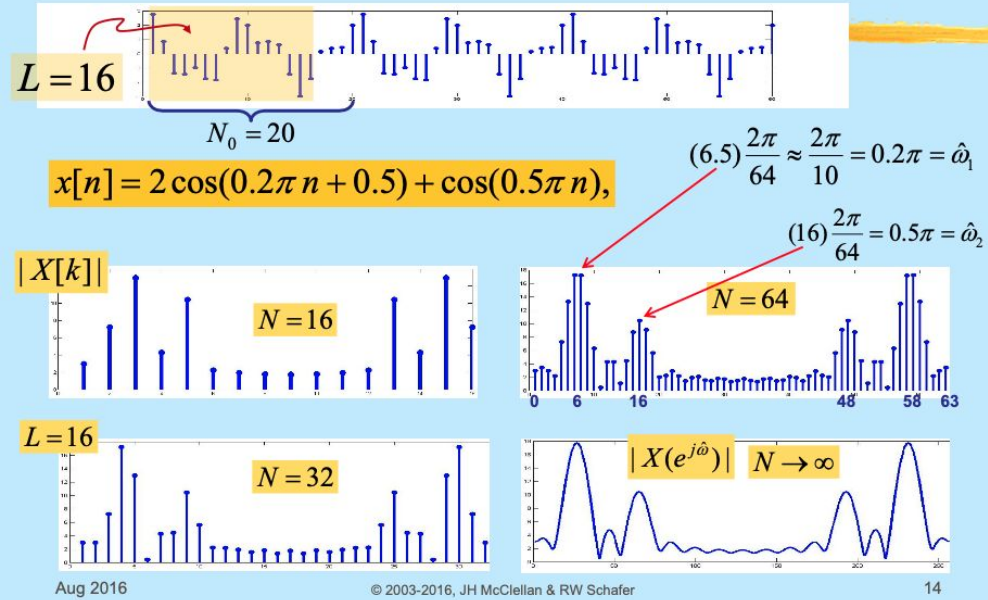
Semester Goals: Lecture Keywords Team

- Extract keywords from lecture slides on dspfirst.gatech.edu.
- Map each keyword to the lecture number and page it appears in.
- Parse chatbot questions to detect these keywords.
- Return image/link to the appropriate slide.

Keyword Generation

- Textbook and lectures contain complicated equations that make it difficult to analyze
 - semantic analysis will not be very useful with LaTeX expressions!
- Use titles of the lecture slides
- pdfplumber library to extract pdf data
- map in convenient json format
- Assumes student will be able to understand lecture material on their own

Imprecise Period Example (2)



NEGATIVE FREQUENCY

- Is negative frequency real?
- Doppler Radar provides intuition
 - Police radar measures speed by using the Doppler shift principle
 - Let's assume 400Hz \leftrightarrow 60 mph
 - +400Hz means towards the radar
 - -400Hz means away (opposite **direction**)
 - Think of a train whistle

pdf parser
reads line
by line

```
{'keyword':  
'NEGATIVE  
FREQUENCY',  
'lecture': 5,  
'page': 13}
```

TutorBot Response

- Simple keyword contains model
 - Could possibly combine with more complicated semantic analysis/auto-recommendation feature
- Generates image and clickable link to lecture on dspfirst.gatech.edu

The screenshot shows the TutorBot interface. At the top, there is a blue header with the text "TUTORBOT" and a close button (X). Below the header, there is a user query in a light gray bubble: "FS for a SQUARE WAVE {a}". To the right of the query is a user profile icon. Below the query, there is a dark blue response bubble containing the text: "This is the textbook answer". Below the response bubble, there is a dark blue box with the text "collapse content". Below the "collapse content" box, there is a white box containing the following text and mathematical formulas:

FS for a SQUARE WAVE {a_k}

$$a_k = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-j(2\pi/T_0)kt} dt \quad (k \neq 0)$$

$a_k = \frac{1}{0.04} \int_0^{0.04} 1 e^{-j(2\pi/0.04)kt} dt = \frac{1}{0.04} \int_0^{0.04} e^{-j(2\pi/0.04)kt} dt$

$$= \frac{1}{(-j2\pi k)} (e^{-j(2\pi k)0.04} - 1) = \frac{1 - (-1)^k}{j2\pi k} \quad (k \neq 0)$$

At the bottom of the interface, there is a white input field with the text "Write your message here" and a blue arrow icon on the right.



Next Steps

- Better keyword generation
 - multiline titles get cut off
 - use new library
 - filter out useless slides
 - add textbook index support
- Differences between offline and online slides
 - indices are off
 - lectures are different?
- Implement auto-completion of keywords

Demo



Thank you