

# Lab 2: Information Retrieval

CPE 466 Fall 2015

Jeff McGovern

Nicole Martin

## Abstract

An information retrieval system was implemented to process and query the dataset of utterances at hearings for the SB277 bill. Different term weighting, stemming, and stopword removal options were tested for their effect on the precision and accuracy of the results on a supplied set of queries. Information need sentences were converted into queries, and their results were analyzed. Using a stopword removal file consistently resulted in higher cosine similarity scores for query results as expected. While there were many options to filter data, we found that incorporating stemming and stopwords greatly improved relevance of results.

## Introduction

In this lab, we implemented an information retrieval system for a data set of legislative hearings and discussion regarding the California Senate Bill 277 (SB277). The purpose of SB277 was to amend and repeal parts of the Health and Safety Code, thereby removing several exemptions from immunization requirements for admittance to schools and other institutions.

Each entry in the dataset has the transcribed text of what each individual said at the different hearings (utterances) as well as metadata to describe the individual (a unique ID, first name, last name) and information on the hearing they spoke at (date, house, and committee) stored in a single JSON file. The information retrieval system processed this file, allowing for user supplied stopword file and optional stemming, to determine the term weights for each term in each utterance. Using this system, a user could query the system via text file, processed with the same stopword file and stemming options, and the retrieval system would output the top 10 utterances that matched the query along with their scores.

A set of ten queries and five information needs were run on the information retrieval system with different stopword and stemming options. Queries contained mainly keywords related to the desired information, while information needs were more expressive and complete sentences. The retrieval system processed these very similarly to the documents in the collection and results of these queries analyzed for precision and accuracy by us.

# System Design and Implementation

The information retrieval system runs on Python 2.6. Its structure closely resembles the vector space model (VSM), with Utterances representing documents and an Utterance Collection representing a document collection. Each Utterance had a corresponding Vocabulary, tracking the number of occurrences of each word in the document and JSON metadata in a Python dictionary. An Utterance Collection tracks and parses Utterances, calculating document frequency while it parses Utterances.

Each Utterance contains a sparse vector of weights for each term in its Vocabulary, calculated using the following scheme:

$$w = tf \cdot idf$$

where  $tf = \frac{f_t}{\max(f_t)}$  is the term frequency, normalized by the maximum frequency in the Utterance Collection and  $idf$  is inverse document frequency, calculated as follows:

$$idf = \log\left(\frac{N}{df}\right)$$

where  $N$  is the number of Utterances in the collection and  $df$  was the number of Utterances that term appeared in. Cosine similarity between a document's term weights and a query's term weights gives a similarity value used to rank documents by relevance. Inverse document frequency for a term in a query comes from the term's inverse document frequency in the collection.

Both Queries and Utterances share the same parsing engine, both with the same tokenization, stopwords, and stemming options. A compiled regex tokenizes the words, each of which feeds into the stopwords filter and stemming algorithm, returning a list of words to insert into the appropriate Vocabularies.

The execution of the implemented information retrieval system involved a 2-step process:

1. Parsing a supplied JSON file of hearing utterances with any of the following options enabled:
  - a. Stopword removal with supplied list of stopwords
  - b. Stemming using the Porter-Stemmer stemming algorithm
  - c. Metadata incorporation
  - d. Deduplication of entries with identical speakers, date, and text
2. Querying against the Utterance Collection subject to the above enabled options

The driver for the IR system is a Python file called `run.py` that handles the command line arguments and begins the parsing and querying process. Supplemental files `query.py`,

utterance.py, and Stemming was implemented via the Porter Stemming Algorithm using the Python implementation written by Vivake Gupta and available online (<http://tartarus.org/~martin/PorterStemmer>). The stopword removal implementation involved supplying a txt file containing one stopword per line.

After the creation of an Utterance Collection to store metadata about the number of utterances, the optionally supplied stopword file and if stemming was enabled, the parsing process for each utterance was as follows:

1. Create an Utterance object for each utterance
2. Store Utterance JSON metadata
3. Parse the Utterance text into a Vocabulary

The information retrieval system calculates the term weights for each Utterance once every Utterance has been parsed. Metadata incorporation and deduplication, if enabled, occur on the fly. Deduplication leverages a custom hash function. Once found and if metadata incorporation is enabled, the differing data merges into the old Utterance object, i.e., the metadata values are added to the old utterances vocabulary. Usually these situations occurred when the same person said the same text on the same date, but the `PersonType` attribute would differ. If deduplication is enabled and metadata incorporation is *not* enabled, the duplicate record is simply ignored. If neither are enabled, the duplicate document is treated like any other Utterance and incorporated in full into the Utterance Collection.

## Query Answering

Query 01: “Disneyland incident measles people sick”

We compared the results of the different parsing options (without stemming or stopword removal, with stemming but without stopword removal, without stemming and with stopword removal, and with both stemming and stopword removal) on the results of this query to determine the best parameters for querying our information retrieval system given this data set. The top 10 utterances and their cosine similarity scores without using stopword removal and without stemming are shown in Figure 1 below.

```

~~~~~
1 |0.351 pid:96  person:Legislator
   And I would also point out that this is not the first time measles have shown up in Disneyland.
   Disneyland is an international destination. People come from all over the world. There is still
   measles out in the world.
~~~~~
2 |0.275 pid:96  person:Legislator
   So when you have a high enough immunization rate, basically, when an exposer shows up, someone
   shows up with the disease, it doesn't go anywhere. In fact, let's just take Disneyland, for
   example. This is clearly not the first time someone has shown up to Disneyland with measles.
   Disneyland is an internationally renowned park. People come from all over the world.
~~~~~
3 |0.256 pid:4994 person:GeneralPublic
   So I wanted to let you know, for those of you who don't already know, being located in Orange
   County near Disneyland, we did have eight cases of Measles come into our emergency department,
   children with Measles come into our ED this winter. Five out of the eight were so sick that they
   needed to be hospitalized.
~~~~~
4 |0.239 pid:4992 person:GeneralPublic
   The recent measles outbreak illustrates this point. The measles vaccine works great. One dose
   results in about 93 or 95% effectiveness. Two doses gets that up to about 97%. But that's not
   100%. When measles was introduced in Disneyland in December, it spread.
~~~~~
5 |0.237 pid:96  person:Legislator
   And, this year the Disneyland Measles outbreak infected 134 Californians, as well as other
   people across the United States, and in Canada, and Mexico. And, it also included 15 infants who
   were too young to be immunized, here in California. It hospitalized at least 20 people.
~~~~~
6 |0.235 pid:96  person:Legislator
   That's what we saw with the Disneyland outbreak, the measles outbreak that started in
   Disneyland, that's what we saw with the pertussis outbreak, that's why we're doing this bill.
~~~~~
7 |0.232 pid:96  person:Legislator
   We've actually had deaths from measles over the past 12 years from these long term sequelae of
   measles, that every one of those people who got measles, the 130 people who got measles, they
   are at risk over the next 10 years of developing a deadly sequelae of SSPE, which is what the
   infant that you just heard the story from is dying of.
~~~~~
8 |0.229 pid:96  person:Legislator
   And you had hundreds of hospitalizations from that pertussis outbreak. So, pertussis clearly
   triggered legislative action. Measles, of course we had the Disneyland outbreak. But don't
   forget that the year before, in 2013, before the Disneyland outbreak, we had a record 644 cases
   of measles in the country.
~~~~~
9 |0.221 pid:96  person:Legislator
   And what we're seeing there is, you're seeing schools with immunization rates as low as 50% or
   below and in fact when people have analyzed the measles outbreak that happened to begin in
   Disneyland.
~~~~~
10|0.214 pid:72  person:Legislator
   I guess at the end of the day, what it comes down to, I think some of us just have fundamental
   distrust for science, sometimes distrust for government, and unfortunately it carries over to
   risks for other people. Sometimes decisions people make put other people at risk. I went on the
   WHO, World Health Organization website, got it right here in front of me on Sunday looking at
   measles. We talked a lot about measles here because of the measles outbreak.
~~~~~

```

Figure 1: Top 10 results, without stopword removal and without stemming.

Overall the results in Figure 1 show fairly successful matching with 8 out of 10 results directly talking about measles in context with Disneyland. Cosine similarity scores ranged from a high of 0.351 to a low of 0.214. All of the results included “measles”, 9 results included

“Disneyland”, 6 of the results included “people”, only one result included “sick”, and none of the results include “incident”.

The query “Disneyland incident measles people sick” was run on the system again, without stemming but this time using stopword removal with the file stopwords-short.txt to determine if removing stopwords improved the search results, as shown in Figure 2.

```
~~~~~
1 |0.392 pid:96  person:Legislator
   And I would also point out that this is not the first time measles have shown up in Disneyland.
   Disneyland is an international destination. People come from all over the world. There is still
   measles out in the world.
~~~~~
2 |0.302 pid:96  person:Legislator
   So when you have a high enough immunization rate, basically, when an exposor shows up, someone
   shows up with the disease, it doesn't go anywhere. In fact, let's just take Disneyland, for
   example. This is clearly not the first time someone has shown up to Disneyland with measles.
   Disneyland is an internationally renowned park. People come from all over the world.
~~~~~
3 |0.272 pid:4994 person:GeneralPublic
   So I wanted to let you know, for those of you who don't already know, being located in Orange
   County near Disneyland, we did have eight cases of Measles come into our emergency department,
   children with Measles come into our ED this winter. Five out of the eight were so sick that they
   needed to be hospitalized.
~~~~~
4 |0.267 pid:96  person:Legislator
   We've actually had deaths from measles over the past 12 years from these long term sequelae of
   measles, that every one of those people who got measles, the 130 people who got measles, they
   are at risk over the next 10 years of developing a deadly sequelae of SSPE, which is what the
   infant that you just heard the story from is dying of.
~~~~~
5 |0.265 pid:4992 person:GeneralPublic
   The recent measles outbreak illustrates this point. The measles vaccine works great. One dose
   results in about 93 or 95% effectiveness. Two doses gets that up to about 97%. But that's not
   100%. When measles was introduced in Disneyland in December, it spread.
~~~~~
6 |0.264 pid:96  person:Legislator
   That's what we saw with the Disneyland outbreak, the measles outbreak that started in
   Disneyland, that's what we saw with the pertussis outbreak, that's why we're doing this bill.
~~~~~
7 |0.262 pid:96  person:Legislator
   And, this year the Disneyland Measles outbreak infected 134 Californians, as well as other
   people across the United States, and in Canada, and Mexico. And, it also included 15 infants who
   were too young to be immunized, here in California. It hospitalized at least 20 people.
~~~~~
8 |0.249 pid:72  person:Legislator
   I guess at the end of the day, what it comes down to, I think some of us just have fundamental
   distrust for science, sometimes distrust for government, and unfortunately it carries over to
   risks for other people. Sometimes decisions people make put other people at risk. I went on the
   WHO, World Health Organization website, got it right here in front of me on Sunday looking at
   measles. We talked a lot about measles here because of the measles outbreak.
~~~~~
9 |0.248 pid:96  person:Legislator
   And you had hundreds of hospitalizations from that pertussis outbreak. So, pertussis clearly
   triggered legislative action. Measles, of course we had the Disneyland outbreak. But don't
   forget that the year before, in 2013, before the Disneyland outbreak, we had a record 644 cases
   of measles in the country.
~~~~~
10|0.248 pid:96  person:Legislator
   And what we're seeing there is, you're seeing schools with immunization rates as low as 50% or
   below and in fact when people have analyzed the measles outbreak that happened to begin in
   Disneyland.
~~~~~
```

Figure 2: Top 10 results, with stopword removal via stopwords-short.txt and without stemming.

The results shown in Figure 2 are interesting in that all the same utterances are present in the top 10 list as in Figure 1, though in a very different order and with very different similarity scores. Since it produced the same utterances as without removing any stopwords, it also had fairly successful matching with 8 out of 10 results directly talking about measles in context with Disneyland. As shown in Figure 2, cosine similarity scores ranged from 0.392 to 0.248, which are higher values overall than the 0.351 to 0.214 range seen in Figure 1. This stopword removal results in fewer terms being considered for term weights and inverse document frequency, thereby making the remaining terms more relevant in comparison.

This query was run on the system again without stemming but using the larger stopword file stopwords-medium.txt, shown in Figure 3 to see how removing more stopwords would alter the search results.

```

~~~~~
1 |0.448 pid:96  person:Legislator
   And I would also point out that this is not the first time measles have shown up in Disneyland.
   Disneyland is an international destination. People come from all over the world. There is still
   measles out in the world.
~~~~~
2 |0.338 pid:96  person:Legislator
   So when you have a high enough immunization rate, basically, when an exposé shows up, someone
   shows up with the disease, it doesn't go anywhere. In fact, let's just take Disneyland, for
   example. This is clearly not the first time someone has shown up to Disneyland with measles.
   Disneyland is an internationally renowned park. People come from all over the world.
~~~~~
3 |0.322 pid:4994 person:GeneralPublic
   So I wanted to let you know, for those of you who don't already know, being located in Orange
   County near Disneyland, we did have eight cases of Measles come into our emergency department,
   children with Measles come into our ED this winter. Five out of the eight were so sick that they
   needed to be hospitalized.
~~~~~
4 |0.320 pid:96  person:Legislator
   That's what we saw with the Disneyland outbreak, the measles outbreak that started in
   Disneyland, that's what we saw with the pertussis outbreak, that's why we're doing this bill.
~~~~~
5 |0.285 pid:96  person:Legislator
   And you had hundreds of hospitalizations from that pertussis outbreak. So, pertussis clearly
   triggered legislative action. Measles, of course we had the Disneyland outbreak. But don't
   forget that the year before, in 2013, before the Disneyland outbreak, we had a record 644 cases
   of measles in the country.
~~~~~
6 |0.285 pid:96  person:Legislator
   And, this year the Disneyland Measles outbreak infected 134 Californians, as well as other
   people across the United States, and in Canada, and Mexico. And, it also included 15 infants who
   were too young to be immunized, here in California. It hospitalized at least 20 people.
~~~~~
7 |0.279 pid:96  person:Legislator
   We've actually had deaths from measles over the past 12 years from these long term sequelae of
   measles, that every one of those people who got measles, the 130 people who got measles, they
   are at risk over the next 10 years of developing a deadly sequelae of SSPE, which is what the
   infant that you just heard the story from is dying of.
~~~~~
8 |0.274 pid:96  person:Legislator
   And what we're seeing there is, you're seeing schools with immunization rates as low as 50% or
   below and in fact when people have analyzed the measles outbreak that happened to begin in

```

Disneyland.

```

~~~~~
9 |0.270 pid:4992 person:GeneralPublic
  The recent measles outbreak illustrates this point. The measles vaccine works great. One dose
  results in about 93 or 95% effectiveness. Two doses gets that up to about 97%. But that's not
  100%. When measles was introduced in Disneyland in December, it spread.
~~~~~
10|0.267 pid:72  person:Legislator
   I guess at the end of the day, what it comes down to, I think some of us just have fundamental
   distrust for science, sometimes distrust for government, and unfortunately it carries over to
   risks for other people. Sometimes decisions people make put other people at risk. I went on the
   WHO, World Health Organization website, got it right here in front of me on Sunday looking at
   measles. We talked a lot about measles here because of the measles outbreak.
~~~~~

```

Figure 3: Top 10 results, with stopword removal via stopwords-medium.txt and without stemming.

Again we find that Figure 3 contains all the same utterances as in Figure 1 and Figure 2. Since it produced the same utterances as without removing any stopwords, it also had fairly successful matching at 8 out of 10 results. As shown in Figure 2, cosine similarity scores ranged from 0.448 to 0.267, which are higher values overall than removing stopwords from stopwords-short.txt, continuing the trend of stopword removal resulting in fewer terms being considered for term weights and inverse document frequency, thereby making the remaining terms more relevant in comparison.

```

~~~~~
1 |0.587 pid:96  person:Legislator
   And I would also point out that this is not the first time measles have shown up in Disneyland.
   Disneyland is an international destination. People come from all over the world. There is still
   measles out in the world.
~~~~~
2 |0.445 pid:4994 person:GeneralPublic
   So I wanted to let you know, for those of you who don't already know, being located in Orange
   County near Disneyland, we did have eight cases of Measles come into our emergency department,
   children with Measles come into our ED this winter. Five out of the eight were so sick that they
   needed to be hospitalized.
~~~~~
3 |0.429 pid:96  person:Legislator
   So when you have a high enough immunization rate, basically, when an exposé shows up, someone
   shows up with the disease, it doesn't go anywhere. In fact, let's just take Disneyland, for
   example. This is clearly not the first time someone has shown up to Disneyland with measles.
   Disneyland is an internationally renowned park. People come from all over the world.
~~~~~
4 |0.319 pid:96  person:Legislator
   And what we're seeing there is, you're seeing schools with immunization rates as low as 50% or
   below and in fact when people have analyzed the measles outbreak that happened to begin in
   Disneyland.
~~~~~
5 |0.316 pid:72  person:Legislator
   I guess at the end of the day, what it comes down to, I think some of us just have fundamental
   distrust for science, sometimes distrust for government, and unfortunately it carries over to
   risks for other people. Sometimes decisions people make put other people at risk. I went on the
   WHO, World Health Organization website, got it right here in front of me on Sunday looking at
   measles. We talked a lot about measles here because of the measles outbreak.
~~~~~
6 |0.314 pid:96  person:Legislator
   That's what we saw with the Disneyland outbreak, the measles outbreak that started in
   Disneyland, that's what we saw with the pertussis outbreak, that's why we're doing this bill.
~~~~~

```

```

~~~~~
7 |0.303 pid:96  person:Legislator
   And, this year the Disneyland Measles outbreak infected 134 Californians, as well as other
   people across the United States, and in Canada, and Mexico. And, it also included 15 infants who
   were too young to be immunized, here in California.  It hospitalized at least 20 people.
~~~~~
8 |0.302 pid:96  person:Legislator
   However, in past exposures in Disneyland, what happened is that the disease showed up and there
   were enough people who were vaccinated and it basically maybe affected one or two other people
   and it stopped and you didn't have to hear about it in the news because it wasn't spreading
   throughout our state.
~~~~~
9 |0.300 pid:96  person:Legislator
   We've actually had deaths from measles over the past 12 years from these long term sequelae of
   measles, that every one of those people who got measles, the 130 people who got measles, they
   are at risk over the next 10 years of developing a deadly sequelae of SSPE, which is what the
   infant that you just heard the story from is dying of.
~~~~~
10|0.286 pid:4992 person:GeneralPublic
   The recent measles outbreak illustrates this point. The measles vaccine works great. One dose
   results in about 93 or 95% effectiveness. Two doses gets that up to about 97%. But that's not
   100%. When measles was introduced in Disneyland in December, it spread.
~~~~~

```

Figure 4: Top 10 results, with stopword removal via stopwords-long.txt and without stemming.

Using stopwords-long.txt, a new utterance was finally seen in the top 10 results compared with Figures 1-3. It occurred at position 8 on the list, but since it replaced a different result directly references Disneyland and measles outbreaks the accuracy remained at 8 out of 10 results. As shown in Figure 4, cosine similarity scores ranged from 0.587 to 0.286, which matches the trend of higher cosine similarity scores with more stopword removal.

Overall the placement of the utterances and their similarity scores are reported in Table 1.

Table 1: Utterances, their ranking, and cosine similarity score with no stemming and different stopword files.

| Utterance  | no stopword |       | stopwords<br>-short |       | stopwords<br>-medium |       | stopwords<br>-long |       |
|--|-------------|-------|---------------------|-------|----------------------|-------|--------------------|-------|
|  | Rank        | Score | Rank                | Score | Rank                 | Score | Rank               | Score |
| And I would also point out that this is not the first time measles have shown up in Disneyland. Disneyland is an international destination. People come from all over the world. There is still measles out in the world.  | 1           | 0.351 | 1                   | 0.392 | 1                    | 0.448 | 1                  | 0.587 |
| So I wanted to let you know, for those of you who don't already know, being located in Orange County near Disneyland, we did have eight cases of Measles come into our emergency department, children with Measles come into our ED this winter. Five out of the eight were so sick that they needed to be hospitalized. | 3           | 0.256 | 3                   | 0.272 | 3                    | 0.322 | 2                  | 0.445 |
| So when you have a high enough immunization rate, basically, when an exposer shows up, someone shows up with the disease,  | 2           | 0.275 | 2                   | 0.302 | 2                    | 0.338 | 3                  | 0.429 |



|   |    |       |    |       |    |       |    |       |
|---|----|-------|----|-------|----|-------|----|-------|
| it doesn't go anywhere. In fact, let's just take Disneyland, for example. This is clearly not the first time someone has shown up to Disneyland with measles. Disneyland is an internationally renowned park. People come from all over the world.  |    |       |    |       |    |       |    |       |
| And what we're seeing there is, you're seeing schools with immunization rates as low as 50% or below and in fact when people have analyzed the measles outbreak that happened to begin in Disneyland.   | 9  | 0.221 | 10 | 0.248 | 8  | 0.274 | 4  | 0.319 |
| I guess at the end of the day, what it comes down to, I think some of us just have fundamental distrust for science, sometimes distrust for government, and unfortunately it carries over to risks for other people. Sometimes decisions people make put other people at risk. I went on the WHO, World Health Organization website, got it right here in front of me on Sunday looking at measles. We talked a lot about measles here because of the measles outbreak. | 10 | 0.214 | 8  | 0.249 | 10 | 0.267 | 5  | 0.316 |
| That's what we saw with the Disneyland outbreak, the measles outbreak that started in Disneyland, that's what we saw with the pertussis outbreak, that's why we're doing this bill.   | 6  | 0.235 | 6  | 0.264 | 4  | 0.320 | 6  | 0.314 |
| And, this year the Disneyland Measles outbreak infected 134 Californians, as well as other people across the United States, and in Canada, and Mexico. And, it also included 15 infants who were too young to be immunized, here in California. It hospitalized at least 20 people.   | 5  | 0.237 | 7  | 0.262 | 6  | 0.285 | 7  | 0.303 |
| However, in past exposures in Disneyland, what happened is that the disease showed up and there were enough people who were vaccinated and it basically maybe affected one or two other people and it stopped and you didn't have to hear about it in the news because it wasn't spreading throughout our state.  | -  |       | -  | -     | -  | -     | 8  | 0.302 |
| We've actually had deaths from measles over the past 12 years from these long term sequelae of measles, that every one of those people who got measles, the 130 people who got measles, they are at risk over the next 10 years of developing a deadly sequelae of SSPE, which is what the infant that you just heard the story from is dying of.   | 7  | 0.232 | 4  | 0.267 | 7  | 0.279 | 9  | 0.300 |
| The recent measles outbreak illustrates this point. The measles vaccine works great. One dose results in about 93 or 95% effectiveness. Two doses gets that up to about 97%. But that's not 100%. When measles was introduced in Disneyland in December, it spread.   | 4  | 0.239 | 5  | 0.265 | 9  | 0.270 | 10 | 0.286 |
| And you had hundreds of hospitalizations from that pertussis outbreak. So, pertussis clearly triggered legislative action. Measles, of course we had the Disneyland outbreak. But don't forget that the year before, in 2013, before the Disneyland outbreak, we had a record 644 cases of measles in the country.  | 8  | 0.229 | 9  | 0.248 | 5  | 0.285 | -  | -     |

Overall the more stopwords in the stopword removal file the higher the cosine similarity score, but the top 10 results for this query did not change significantly until a very large stopword file was provided.

The query “Disneyland incident measles people sick” was run on the system again, without stopword removal but using stemming to determine if stemming improved the search results, as shown in Figure 5.

```
~~~~~
1 |0.288 pid:96  person:Legislator
   And I would also point out that this is not the first time measles have shown up in Disneyland.
   Disneyland is an international destination. People come from all over the world. There is still
   measles out in the world.
~~~~~
2 |0.230 pid:96  person:Legislator
   So when you have a high enough immunization rate, basically, when an exposer shows up, someone
   shows up with the disease, it doesn't go anywhere. In fact, let's just take Disneyland, for
   example. This is clearly not the first time someone has shown up to Disneyland with measles.
   Disneyland is an internationally renowned park. People come from all over the world.
~~~~~
3 |0.212 pid:4994 person:GeneralPublic
   So I wanted to let you know, for those of you who don't already know, being located in Orange
   County near Disneyland, we did have eight cases of Measles come into our emergency department,
   children with Measles come into our ED this winter. Five out of the eight were so sick that they
   needed to be hospitalized.
~~~~~
4 |0.201 pid:96  person:Legislator
   And, this year the Disneyland Measles outbreak infected 134 Californians, as well as other
   people across the United States, and in Canada, and Mexico. And, it also included 15 infants who
   were too young to be immunized, here in California. It hospitalized at least 20 people.
~~~~~
5 |0.198 pid:4992 person:GeneralPublic
   The recent measles outbreak illustrates this point. The measles vaccine works great. One dose
   results in about 93 or 95% effectiveness. Two doses gets that up to about 97%. But that's not
   100%. When measles was introduced in Disneyland in December, it spread.
~~~~~
6 |0.196 pid:96  person:Legislator
   That's what we saw with the Disneyland outbreak, the measles outbreak that started in
   Disneyland, that's what we saw with the pertussis outbreak, that's why we're doing this bill.
~~~~~
7 |0.192 pid:96  person:Legislator
   And what we're seeing there is, you're seeing schools with immunization rates as low as 50% or
   below and in fact when people have analyzed the measles outbreak that happened to begin in
   Disneyland.
~~~~~
8 |0.192 pid:96  person:Legislator
   We've actually had deaths from measles over the past 12 years from these long term sequelae of
   measles, that every one of those people who got measles, the 130 people who got measles, they
   are at risk over the next 10 years of developing a deadly sequelae of SSPE, which is what the
   infant that you just heard the story from is dying of.
~~~~~
9 |0.191 pid:96  person:Legislator
   And you had hundreds of hospitalizations from that pertussis outbreak. So, pertussis clearly
   triggered legislative action. Measles, of course we had the Disneyland outbreak. But don't
   forget that the year before, in 2013, before the Disneyland outbreak, we had a record 644 cases
   of measles in the country.
~~~~~
10|0.183 pid:96  person:Legislator
   over 9,000 people, hospitalized 809, killed 10 infants, and the Disneyland measles outbreak,
   which infected 134 Californians including 15 infants too young to be immunized, and hospitalized
   at least 20 people, have both been linked to communities with low vaccination rates and high use
   of the personal belief exemption.
~~~~~
```

Figure 5: Top 10 results, without stopword removal and with stemming.

In contrast with Figure 1, Figure 5 shows lower overall cosine similarity scores ranging from 0.288 to 0.183, which are lower values than without stemming, but more successful matching results having 9 out of 10 results references measles with Disneyland. The top 10 results include 9 of the same utterances as in Figure 1. The ordering of the top 2 results remain the same between the stemming and non-stemming cases without stopword removal, but the majority of the list in a different order. The #10 results are completely different between the stemming and non-stemming cases without stopword removal.

## Query 02: “Centers for Disease Control chance of a serious allergic reaction”

```

1 |0.568      pid:25  person:Legislator
    ~~~~~
    According to the CDC, the Centers for Disease Control, the chance of a serious allergic reaction
    due to DTap, MMR or varicella vaccines is less than four in one million doses.
    ~~~~~
2 |0.308      pid:9623  person:GeneralPublic
    ~~~~~
    Thank you, Senator. I would just like to respond by quoting the Centers for Disease Control
    website, and their question that is posed about unvaccinated children spreading the disease.
    ~~~~~
3 |0.274      pid:96  person:Legislator
    ~~~~~
    You would have conditions where, if there was uncertainty about so the child who may be perhaps
    allergic to something that may be a component of vaccine or they've developed allergic reaction
    ~~~~~
4 |0.246      pid:96  person:Legislator
    ~~~~~
    The Centers for Disease Control list immunizations as one of the ten great public health
    achievements, that resulted in adding 25 years, 25 years of life expectancy for Americans. The
    World Health Organization describes vaccination as second only to clean water in reducing the
    burden of infectious disease.
    ~~~~~
5 |0.232      pid:96  person:Legislator
    ~~~~~
    So if you want to talk about measles, they have...if they're exposed, they have a 90% chance of
    catching the disease. If they are vaccinated, they have about 1% chance of catching the disease.
    ~~~~~
6 |0.222      pid:25  person:Legislator
    ~~~~~
    Thank you. Well, these are helpful insights. I would ask, you referenced the CDC, Centers for
    Disease Control. Do you question their data as to the difference between adverse reactions
    versus fatalities for nonvaccinated who contract certain diseases?
    ~~~~~
7 |0.208      pid:9623  person:GeneralPublic
    ~~~~~
    So again, we can never be completely assured that we are providing no risk, but if you look at
    the AAP guidelines and the guidelines for the Centers for Disease Control and the Committee on
    Immunization Practices, it is recommended that it be given to those children.
    ~~~~~
8 |0.198      pid:57  person:Legislator
    ~~~~~
    Okay. So, that's one of those things that we can test for, but is it ever possible for a child
    to have a very strong reaction, a very negative reaction and they don't have that? Are there
    other symptoms that you can determine a cause and effect or is there some randomness to this
    where you have...that there's no reason to believe that this child is going to have a reaction
    and yet they do?
    ~~~~~
9 |0.190      pid:96  person:Legislator
    ~~~~~
    As I mentioned in my opening testimony they have determined that the risk of any sort of serious
  
```

incidents with vaccinations, at least 1000 times less frequent than any serious condition from the disease itself.

10|0.188      pid:96   person:Legislator

These findings are also supported by the World Health Organization, the European Union, the Centers for Disease Control, the American Medical Association and other leading medical scientific and public health organizations in the United States and across the world.

Figure 6: Top 10 results for Query 2 with no options enabled.

For Query 2, Figure 6 shows information retrieval *without* any options (stopwords, stemming, deduplication, or metadata) enabled. Of these 10 responses, the only ones that seem to be irrelevant are result 4 and 10 (if only because it is unclear what findings are at hand), giving us a possible precision of 0.8. Altering the option made minimal improvements to the results.

1 |0.630      pid:25   person:Legislator

According to the CDC, the Centers for Disease Control, the chance of a serious allergic reaction due to DTap, MMR or varicella vaccines is less than four in one million doses.

2 |0.372      pid:96   person:Legislator

You would have conditions where, if there was uncertainty about so the child who may be perhaps allergic to something that may be a component of vaccine or they've developed allergic reaction

3 |0.271      pid:25   person:Legislator

Thank you. Well, these are helpful insights. I would ask, you referenced the CDC, Centers for Disease Control. Do you question their data as to the difference between adverse reactions versus fatalities for nonvaccinated who contract certain diseases?

4 |0.262      pid:96   person:Legislator

So if you want to talk about measles, they have...if they're exposed, they have a 90% chance of catching the disease. If they are vaccinated, they have about 1% chance of catching the disease.

5 |0.261      pid:9623      person:GeneralPublic

Thank you, Senator. I would just like to respond by quoting the Centers for Disease Control website, and their question that is posed about unvaccinated children spreading the disease.

6 |0.212      pid:96   person:Legislator

The findings are also supported by the World Health Organization, the European Union, the Centers for Disease Control, the American Medical Association, and many other leading medical scientific and public health organizations in the United States and across the world.

7 |0.212      pid:96   person:Legislator

These findings are also supported by the World Health Organization, the European Union, the Centers for Disease Control, the American Medical Association and other leading medical scientific and public health organizations in the United States and across the world.

8 |0.208      pid:96   person:Legislator

And, these findings, they're also supported by the World Health Organization, the European Union, Centers for Disease Control, the American Medical Association, and other leading medical, scientific, and public health organizations in the United States and across the world.

9 |0.199      pid:96   person:Legislator

```

So, I point out that studies looking at significant adverse reactions, so you know, obviously
ones are more minor. Pain, swelling at the injection site, of course are fairly common. There
are other more significant ones. Particularly the most common would be actually allergic
reaction to the vaccine. Those are approximately one in a million, less than the risk of being
struck by lightning.
~~~~~
10|0.195      pid:10897      person:GeneralPublic
~~~~~
Rachel Center from Yucaipa, California, mother of 6, opposed SB 277.
~~~~~

```

Figure 7: Top 10 results for Query 2 with stemming, stopwords-long.txt, dedup and metadata incorporation.

For example, in Figure 7, we see that enabling the long list of stopwords as well as all of the other options gives us stranger results. Results 6, 7, and 8 appear to be duplicates, but there are subtle difference in the text of each utterance making it difficult to see that they're effectively saying the same thing. In fact, it looks like only results 1, 2, and 3 are relevant, giving us a precision of 0.3.

### Query 03: "requiring parents to jump through hoops"

```

~~~~~
1 |0.435      pid:105 person:Legislator
~~~~~
We don't say, "Have you got your 27 vaccines," or whatever it is, before they come and go to
Disneyland, or any other place. And so, you're requiring our own parents to jump through hoops
to do all these things in the interest of public health; and yet, we're still exposed.
~~~~~
2 |0.187      pid:4960      person:GeneralPublic
~~~~~
I urge you, Senators, Chairman, please, do not let this bill pass through this committee, and go
through all the other things we have to go through, because this is a bad bill, it needs to be
stopped now, and I appreciate your concern, because we know there's a risk.
~~~~~
3 |0.176      pid:9941      person:GeneralPublic
~~~~~
No. But what will do you with the parents? Many of the parents who are --
~~~~~
4 |0.169      pid:85  person:LegStaff
~~~~~
All right. Let me just jump in here. We talked about narrow tailoring earlier, and I want to ask
you to narrowly tailor your responses to the questions of the members. Answer their question.
And then stop and see if they have another question. And we'll move through the member questions
and get on to public comment. Dr. Wood, do you have any more questions?
~~~~~
5 |0.165      pid:10765     person:GeneralPublic
~~~~~
The thousands of Californians who have come to Sacramento this year to participate in our
democracy and testify against this bill are asking you to protect parental, civil and human
rights. The right of the state to tell parents what to do to their children's bodies ends with
the right of parents to protect their children's lives begins.
~~~~~
6 |0.160      pid:10765     person:GeneralPublic
~~~~~
Most parents do not have the money to exercise or the expertise to homeschool and those at
greatest risk for being denied an equal education guaranteed by the California constitution are
handicapped children requiring special education under civil rights laws.
~~~~~

```

```

7 |0.160      pid:105 person:Legislator
~~~~~
    If I was a parent and I had a child in that category, from what I'm hearing, I wouldn't trust
    the public school anyway, and yet it seems like you'd want to create a safe cocoon around them
    as much as you can; and yet, we're jumping through a lot of hoops to force a lot of other people
    out when we still don't have a safe situation in school.
~~~~~
8 |0.155      pid:9613      person:GeneralPublic
~~~~~
    I want to highlight that most religions don't actually oppose vaccines. There are certainly
    parents who have sincere religious opposition to vaccines, but the way our jurisprudence works,
    it's very hard to limit a religious exemption if we offer one just to those parents.
~~~~~
9 |0.150      pid:85  person:LegStaff
~~~~~
    Thank you. Okay, we're going to move on. We're going to go to public comment. I'm first going to
    set a couple of quick ground rules. Firstly, and thank you to all of our witnesses, to our
    authors I want to say that, I deeply appreciate that parents have brought children here today.
    It has exposed these children to our government and taught them an important lesson about civic
    engagement. And I hope they also leave knowing that nearly every issue has multiple sides to it,
    as we've clearly discovered through today's debate and that reasonable people can differ.
~~~~~
10|0.148     pid:85  person:LegStaff
~~~~~
    Wait a second, wait a second, let me jump in here
~~~~~

```

Figure 8: Top 10 results for Query 3 with no options enabled.

One cool example was Query 3. Just adding stemming seemed to bring more relevance to the results. In Figure 8, one can see that the only results that have to do with parents jumping through hoops are results 1 and 7. Figure 9 shows that adding stemming brings those two relevant results to the top and adds some more high level relevance to the query. Result 5 in Figure 9 seems to be somewhat relevant, but result 7 talks about the degree the a parent must “jump through a hoop” in a vague sense, while Result 9 mentions policies and enforcement with respect to parents. These may be happenstance, but they seem intuitively relevant.

```

1 |0.426      pid:105 person:Legislator
~~~~~
    We don't say, "Have you got your 27 vaccines," or whatever it is, before they come and go to
    Disneyland, or any other place. And so, you're requiring our own parents to jump through hoops
    to do all these things in the interest of public health; and yet, we're still exposed.
~~~~~
2 |0.266      pid:105 person:Legislator
~~~~~
    If I was a parent and I had a child in that category, from what I'm hearing, I wouldn't trust
    the public school anyway, and yet it seems like you'd want to create a safe cocoon around them
    as much as you can; and yet, we're jumping through a lot of hoops to force a lot of other people
    out when we still don't have a safe situation in school.
~~~~~
3 |0.220      pid:9613      person:GeneralPublic
~~~~~
    We don't change the curriculum to prevent parents from having to choose between a curriculum
    they don't like and homeschooling. It's problematic to say we need to do away with a reasonable
    health and safety requirement, vaccination to prevent parents from having to choose between
    following the requirement and choosing another educational option, again, home schooling or
    independent study.
~~~~~
4 |0.203      pid:4960      person:GeneralPublic
~~~~~
    I urge you, Senators, Chairman, please, do not let this bill pass through this committee, and go
~~~~~

```

through all the other things we have to go through, because this is a bad bill, it needs to be stopped now, and I appreciate your concern, because we know there's a risk.

5 | 0.194      pid:10765      person:GeneralPublic

The thousands of Californians who have come to Sacramento this year to participate in our democracy and testify against this bill are asking you to protect parental, civil and human rights. The right of the state to tell parents what to do to their children's bodies ends with the right of parents to protect their children's lives begins.

6 | 0.180      pid:9613      person:GeneralPublic

The second point is the courts don't require a religious exemption. Again emphasized that parental rights and religious rights don't allow the parents to expose the community to ill health and to disease and every court since have followed that precedent.

7 | 0.178      pid:29      person:Legislator

Okay, 'cause that kind of answered my question because it seems like with that amendment, it's a lot more looser interpretation than what the CDC is requiring. Okay, so my second question, just to be really fast, is under this bill, is the parent still required to sign a consent form for the child to be vaccinated? So, when you go to the doctor-

8 | 0.165      pid:85      person:LegStaff

All right. Let me just jump in here. We talked about narrow tailoring earlier, and I want to ask you to narrowly tailor your responses to the questions of the members. Answer their question. And then stop and see if they have another question. And we'll move through the member questions and get on to public comment. Dr. Wood, do you have any more questions?

9 | 0.161      pid:9282      person:GeneralPublic

The staff consistently worked with us to reinforce messaging about illness policies and educating parents about vaccinations. Again we had the same conversations with parents and teachers when Rhett started kindergarten. At a parent meeting in 2014, we continued to pressure the school district to take action.

10 | 0.161      pid:9941      person:GeneralPublic

No. But what will do you with the parents? Many of the parents who are --

Figure 9: Top 10 results for Query 3 with only stemming enabled.

Query 04: “Nearly everyone, from my father's generation and generations older than him, when I'd go and speak to groups, talk about the stories that they have both from their own childhood, from friends, personal stories of near-death experiences, friends they had who were maimed by communicable diseases.”

1 | 0.936      pid:70      person:Legislator

Nearly everyone, from my father's generation and generations older than him, when I'd go and speak to groups, talk about the stories that they have both from their own childhood, from friends, personal stories of near-death experiences, friends they had who were maimed by communicable diseases.

2 | 0.316      pid:70      person:Legislator

Some of you heard yesterday from a mother, Nina, whose friend can not be vaccinated. She's returned to my district but her story was for her search for a public school with a high immunization rate that her son could attend. Her story is not hers alone, nor are the stories

that many of you have told in this room. All these stories and these lives along with the ones that we don't hear are the reason why we're all here today.

3 | 0.291            pid:70   person:Legislator

He got polio when he was a little kid. He got it right before the vaccine became widely available. He doesn't even have the arm strength in his arms to be able to break his fall. When I go down to my district, and I talk to folks from his generation and older, I hear a story every single time about the horror stories of communicable diseases from measles to polio,

4 | 0.265            pid:70   person:Legislator

If you did, you know that what brought me to this story was the fact that my father had polio as a kid, and got it before the vaccine became widely available. Folks from his generation can't believe that we've allowed these communicable diseases back into American society.

5 | 0.212            pid:70   person:Legislator

So, I, it is a good one, it is a good one. So, today I also stand here before you to ask for you to vote SB 277, and part of my personal story for being here is the story of my father, who contracted Polio, as a little boy, before the vaccine became widely available.

6 | 0.208            pid:10641    person:GeneralPublic

I'm Michelle Parsons, we are three generations of unvaccinated family members, extremely healthy, I'm the first, my daughter is second generation, two children besides her at home, third generation.

7 | 0.168            pid:70   person:Legislator

I also come to this story as the son of a Polio survivor who got Polio as a child and was nursed to health by his very determined mother but who is from a generation that I think is amazed that we have allowed these communicable diseases back into American society when we have the tools to prevent that sort of human suffering and that sort of, of human suffering I suppose.

8 | 0.136            pid:9613    person:GeneralPublic

And that fits our general framework. Our courts have decided that a general law that's applicable to everyone doesn't require exemptions even if your religion opposes it. You don't get to not pay taxes because you have religious objection to them.

9 | 0.133            pid:9282    person:GeneralPublic

It's not about him anymore, it's about all those hundreds of kids in chemotherapy, all those babies. We're so pleased you're taking on this issue. Our culture, through news stories and other kind of story telling emphasizes an individual, individualism to such a degree, it's hard for people to understand how it impacts

10 | 0.132           pid:96   person:Legislator

I think that the, again, the purpose is to try to protect their general community.

Figure 10: Top 10 results for Query 4 with stopwords-long.txt and every option enabled.

Query 4, as shown in Figure 10 with every option enabled, seems to test if our algorithm can pick up exactly identical texts to our query. The top result never changes position when mixing together options and the only way we could get the cosine similarity to be less than 1 was to enable the long stopwords list, deduplicate, and incorporate metadata into the vocabulary.



Calculating precision in this case is unclear. For a query so specific, it seems to be accurate, since it returned that exact document as the first result.

### Query 05: “parents' right to make healthcare decisions for their child”

```
~~~~~
1 |0.560      pid:85  person:LegStaff
~~~~~
    Every child deserves to be able to go to school in a safe and healthy environment. The parents'
    right to make healthcare decisions for their child is not entirely boundless, and without
    limitation. It must take into consideration the effect those decisions have on our neighbors,
    and others in the community.
~~~~~
2 |0.464      pid:25  person:Legislator
~~~~~
    But, where we have a balance here with this bill, that is a tough one for us, is on the one
    hand, many people voiced their strong belief that they have a right to make the decision for
    their child. If it were just a decision about their child, I think you would find no quarrel
    with having a right to make that decision.
~~~~~
3 |0.437      pid:4995  person:GeneralPublic
~~~~~
    There are parents who disagree with me. I know they want to prevent their children from getting
    sick. I know some of them personally and I know them to be loving and kind. I respect they are
    doing what they believe is the best for their child, and if those decisions only affected their
    own child that might be okay, but their decision not only jeopardizes the health and safety of
    their children, it puts my child at risk and that is unfair.
~~~~~
4 |0.425      pid:57  person:Legislator
~~~~~
    So, what types of questions as a parent would you be asking me about my child if that was my
    child?
~~~~~
5 |0.411      pid:36  person:Legislator
~~~~~
    So. Okay. So then, if the parent, let's say the decision of the physician is to not allow.
~~~~~
6 |0.369      pid:9941  person:GeneralPublic
~~~~~
    No. But what will do you with the parents? Many of the parents who are --
~~~~~
7 |0.352      pid:96  person:Legislator
~~~~~
    So remember, this bill is about enrolling children in school. The decision whether to immunize
    or not is something the parent makes with their doctors or other healthcare providers providing
    the immunization. That is separate from school entry. When they then bring their child to
    school, the school will say, please present your immunization record and your school form, and
    they would say, okay your child may not.
~~~~~
8 |0.346      pid:29  person:Legislator
~~~~~
    That's separate from this bill. Obviously, the act of deciding whether to immunize your child or
    not is something you need to discuss with your healthcare professional-
~~~~~
9 |0.329      pid:6440  person:Lobbyist
~~~~~
    That child is my child, and he's more important-
~~~~~
10|0.326      pid:9105  person:GeneralPublic
~~~~~
    Carrie Baylor, mother, former healthcare professional, home school and vaccinated-injured child.
    Strongly opposed.
~~~~~
```

Figure 11: Top 10 results for Query 5 with stopwords-long.txt, stemming, and metadata enabled.

~~~~~  
1 |0.441           pid:85   person:LegStaff,Legislator  
~~~~~

Every child deserves to be able to go to school in a safe and healthy environment. The parents' right to make healthcare decisions for their child is not entirely boundless, and without limitation. It must take into consideration the effect those decisions have on our neighbors, and others in the community.

~~~~~  
2 |0.386           pid:25   person:Legislator  
~~~~~

But, where we have a balance here with this bill, that is a tough one for us, is on the one hand, many people voiced their strong belief that they have a right to make the decision for their child. If it were just a decision about their child, I think you would find no quarrel with having a right to make that decision.

~~~~~  
3 |0.341           pid:4995           person:GeneralPublic  
~~~~~

There are parents who disagree with me. I know they want to prevent their children from getting sick. I know some of them personally and I know them to be loving and kind. I respect they are doing what they believe is the best for their child, and if those decisions only affected their own child that might be okay, but their decision not only jeopardizes the health and safety of their children, it puts my child at risk and that is unfair.

~~~~~  
4 |0.337           pid:96   person:Legislator  
~~~~~

So remember, this bill is about enrolling children in school. The decision whether to immunize or not is something the parent makes with their doctors or other healthcare providers providing the immunization. That is separate from school entry. When they then bring their child to school, the school will say, please present your immunization record and your school form, and they would say, okay your child may not.

~~~~~  
5 |0.275           pid:96   person:Legislator  
~~~~~

There's nothing in this bill that changes that part of the relationship. This bill is about entry into school and entry into, basically it's about enrollment into school. So there is continuous to be informed consent for immunizations and if the parents say, I don't want to get my child immunized. Then the child is not immunized. That is what will happen. And then when they decide they want to enroll their child in school, then the school will say, well, is your child immunized? Then if they decide not to have their child immunized, under this bill, they would have to seek out other options like home schooling or independent study. That's what this bill is about. But, the decision whether to immunize or not, is a decision that's made between the child's parents or guardian, and the health profession that is providing the immunization.

~~~~~  
6 |0.267           pid:10763           person:GeneralPublic  
~~~~~

So that they can intelligently participate in decisions about their baby's healthcare. I believe that SB 277 might have the opposite effect on these discussions. Homeschooling is a viable option for very, very few families.

~~~~~  
7 |0.267           pid:96   person:Legislator  
~~~~~

Of course, the parent may also receive a note from the physician saying that there's a medical exemption. So they will review that conditional and then they will decide whether to allow the child to enroll or not, but that is not about consent for immunization. Consent for immunization happens between the parent of the child, or the guardian. And the healthcare professional who's providing the immunization. That has not changed at all in this bill.

~~~~~  
8 |0.264           pid:29   person:Legislator  
~~~~~

That's separate from this bill. Obviously, the act of deciding whether to immunize your child or not is something you need to discuss with your healthcare professional-

~~~~~  
9 |0.227           pid:36   person:Legislator  
~~~~~

So. Okay. So then, if the parent, let's say the decision of the physician is to not allow.

```

10|0.220      pid:4999      person:GeneralPublic
~~~~~
There are numerous personal decisions parents make for their children. Disposable vs. cloth
diapers, formula vs. breast milk, etc. Vaccination is not a comparable decision. The choice not
to vaccinate your children is a threat to public health. Choosing not to vaccinate your child is
like, making a choice to risk the health of everyone around you.
~~~~~

```

Figure 11: Top 10 results for Query 5 with stopwords-long.txt, stemming, and metadata enabled.

## Query 06: “disabled kids children exceptional needs right to education”

```

1 |0.666      pid:10764      person:GeneralPublic
~~~~~
In California we say, kids with exceptional needs. The kids with disabilities must be educated
in the least restrictive environment and what that means is to the maximum extent possible, they
have a right to receive an education with non-disabled peers and the only exception for that is
if the child's disability prevents it. That's the only exception.
~~~~~
2 |0.439      pid:10764      person:GeneralPublic
~~~~~
So there's serious Constitutional problems and under the Federal law, under the Individual with
Disabilities Educational Act. One of the pillars of that law, which protects the rights of
students with disabilities, is that kids with disabilities have a right to learn with their non-
disabled peers.
~~~~~
3 |0.367      pid:96      person:Legislator
~~~~~
Both protection of the public and children from death and disability due to preventable
communicable diseases, and ensuring children with special needs their right to an education, are
compelling interests to require all students be vaccinated unless granted a medical exemption.
~~~~~
4 |0.362      pid:96      person:Legislator
~~~~~
Both protection of the public and children from death and disability due to preventable
communicable diseases and assuring that children with special needs, their right to an
education, are compelling interests to require all students be vaccinated unless they are
granted a medical exemption. Thank you.
~~~~~
5 |0.308      pid:85      person:LegStaff
~~~~~
Thank you. The next is with respect to individualized education plans and special education.
That is, jumping down now to section H. And as discussed in the analysis under Federal and state
law, disabled children are guaranteed the right to a free appropriate public education,
including necessary services for a child to benefit from his or her education.
~~~~~
6 |0.291      pid:10764      person:GeneralPublic
~~~~~
The second part of the test is the law has to be narrowly tailored to limit rights, the right to
education as little as possible, which I think this bill is the opposite. It's very broad.
That's just the constitutional analysis under California's constitution. In addition, we've got
the Individuals with Disabilities Educational Improvement Act, the IDEA, which is a federal law.
California mirrors federal law in this respect, that requires that kids with special needs.
~~~~~
7 |0.277      pid:59      person:Legislator
~~~~~
So let's focus, like a laser, as my colleagues have done, on just the education issue, we're the
education committee. And we have here clearly a rights in conflict, we've got the rights of
parents who wanna be able to send their children to schools they perceive as safe, and to them,
safe means having nearly 100% vaccinated children, those parents have a right to send their kids
to schools they perceive as safe. On the other hand, the parents, many of them here, who don't
want to vaccinate their children, have a right not to vaccinate, so what do you do?
~~~~~
8 |0.275      pid:9941      person:GeneralPublic
~~~~~
I also want to touch briefly on the legal problems that would arise in taking away education
~~~~~

```

```

from special needs children under the Individuals with Disability Education Act. Under federal
law, special needs children are entitled to a free and public education that includes annual
individualized educational plan reviews and a host of special services, including occupational
therapy, physical therapy, and speech therapy.
~~~~~
9 |0.272      pid:4995      person:GeneralPublic
~~~~~
When SB 277 passes, more children will be vaccinated and so our kids, our families, and our
communities will be more protected against these terrible diseases. We have rights. We have the
right not to have our children endangered at school by preventable disease.
~~~~~
10|0.265     pid:4995     person:GeneralPublic
~~~~~
His right to an education, to interact with his community, should not be restrained because
other people aren't vaccinating their children. In working on this bill over the last few months
I've been privileged to hear the stories of many of the supporters. They tell me we need to
protect our children.
~~~~~

```

Figure 12: Top 10 results for Query 6 with stopwords-medium.txt, stemming, metadata, and dedup enabled.

This resulted in 7 applicable results explicitly discussing accessibility for disabled children, however the results did provide much context for how the SB277 would affect them. The irrelevant results (7, 9, 10) did focus on children’s rights, but did not talk about disabled children specifically.

Query 07: “can a licensed physician in the state of California provide an exemption from immunization?”

```

~~~~~
1 |0.592      pid:96      person:Legislator
~~~~~
So the exemption that is in statute basically says that any licensed physician, any licensed
physician in the state of California can grant a medical exemption.
~~~~~
2 |0.541      pid:96      person:Legislator
~~~~~
So, actually, existing law for the medical exemption allows a licensed physician in the State of
California to exempt a child with a medical exemption for vaccination. And so, if the licensed
physician believes that that child is at risk for injury from the vaccine, they can simply give
a medical exemption. In many ways the law has no restrictions on a licensed physician in being
able to do that.
~~~~~
3 |0.526      pid:96      person:Legislator
~~~~~
that their licensed physician would not write a medical exemption, again any licensed physician
can write that medical exemption.
~~~~~
4 |0.522      pid:96      person:Legislator
~~~~~
They certainly are, any licensed physician in the state of California.
~~~~~
5 |0.488      pid:96      person:Legislator
~~~~~
The law clearly states that any licensed physician in the state of California can provide a
medical exemption. Mainly what they have to do is document the reason. They document the
duration, which there's no limitation. That can be indefinite. And they have to sign it, of
course, saying that they are a licensed physician. And you get a medical exemption.
~~~~~
6 |0.407      pid:96      person:Legislator
~~~~~

```

```

~~~~~
    That is, it's the professional challenge of the physician in terms of what they believe, that if
    the risk of the immunization is gonna be such that it's going to put that child at certain or,
    near, basically increased harm. Then they can provide that exemption. And so, that's to the
    judgment of actually any licensed physician in the state of California. There's no requirement
    that you even have to go to a physician that you've seen multiple times in the past.
~~~~~
7 |0.347      pid:96  person:Legislator
~~~~~
    Again I would urge them to talk to their physician I am astonished when a person says their
    physician, a licensed physician,
~~~~~
8 |0.339      pid:96  person:Legislator
~~~~~
    There is no other restriction on the judgement of a licensed physician, so anyone who said that
    they have a licensed physician who told them they couldn't get immunized, should be able to get
    a medical exemption, and then those children, because they're not immunized, require protection.
    When you talk about do you have children who may be partially immunized, they may have certain
    vaccines, they may not take other vaccines.
~~~~~
9 |0.297      pid:96  person:Legislator
~~~~~
    And so, what this bill has been crafted, working with the chair and the staff, looking at what
    the bill actually says, not what people would like to assert it says. But it actually says, how
    it will be interpreted, we have left that discretion to licensed physicians in the State of
    California, including whether it's your own physician, a specialist you're seeing, another
    physician, we have left that discretion open. Physicians, there's no limitation in the law. We
    just heard from the medical board. We are not aware of any physician who has been disciplined or
    investigated because they provided medical exemption. So there's no cloud hanging over them to
    be able to do this.
~~~~~
10|0.269      pid:9613  person:GeneralPublic
~~~~~
    In the past, California was extremely generous in providing exemptions from immunization
    requirements. Before AB 2109 it was even easier than it is now. It still is reasonably easy to
    get an exemption, but again this balance is something states can and do consider as things
    change.
~~~~~

```

Figure 13: Top 10 results for Query 7 with short.txt, stemming, metadata, and dedup enabled.

Query 7 is interesting in that it is formed as a question. The question appears to be answered by the first query and relevant results are apparent as far as the end of the list.

#### Query 08: “association oppose bill”

```

~~~~~
1 |0.267      pid:4992  person:GeneralPublic
~~~~~
    Sometimes it's difficult to sort those out so for some of those associations are very strong
    other associations in my opinion are not so strong and because of that the federal government
    does have this no-fault vaccine compensation system that's been in place.
~~~~~
2 |0.261      pid:6061  person:GeneralPublic,Lobbyist
~~~~~
    Farrah McDavid Ting representing the California State Association of Counties as well as the
    County Health Executives Association of California in support.
~~~~~
3 |0.249      pid:5062  person:Lobbyist,GeneralPublic
~~~~~
    Phillip Bautista, elected director to the American Nurses Association California, and as a board
    of directors on behalf of the association we strongly support this bill.
~~~~~
4 |0.248      pid:4992  person:GeneralPublic
~~~~~

```

```

~~~~~
such as thrombocytopenia, that's important for clotting factors it could be an allergic
reaction, it may be caused by the vaccine, it may be associated, there's that temporal
association or possibly it's coincidental.
~~~~~
5 |0.241      pid:1946      person:Lobbyist
~~~~~
Lee Angela Reid on behalf of the San Fransisco Unified School District, the Small School
Districts Association and the School Employers Association of California in support.
~~~~~
6 |0.239      pid:10804     person:Lobbyist
~~~~~
Mr. Chair and Members, David Ford behalf of the Association of Northern California Oncologists,
The Medical Oncology Association of Southern California, and The Silicon Valley Leadership Group
all in support.
~~~~~
7 |0.233      pid:4960      person:GeneralPublic
~~~~~
Dr. Brian Stenzler, President of the California Chiropractic Association. Nice to see some of
you again. Chairmen Hernandez, members of the committee, we appreciate you giving us this
opportunity. I want to stress that the California Chiropractic Association does not have a
position on vaccinations.
~~~~~
8 |0.197      pid:5440      person:GeneralPublic
~~~~~
Julie Schiffman, representative of the Homeschool Association of California, and we strongly
oppose this bill.
~~~~~
9 |0.195      pid:4960      person:GeneralPublic
~~~~~
Dr. Brian Stenzler, President of the California Chiropractic Association. We oppose this bill.
~~~~~
10|0.189      pid:4960      person:GeneralPublic
~~~~~
Dr. Brian Stenzler, from San Diego, president of the California Chiropractic Association, we
oppose this bill.
~~~~~

```

Figure 14: Top 10 results for Query 8 with stemming, metadata, and dedup enabled.

Out of 10 results, 8 mention associations that oppose the bill, as shown in Figure 14.

#### Query 09: “mercury in vaccines”

```

~~~~~
1 |0.550      pid:10763     person:GeneralPublic
~~~~~
Mercury has been eliminated from virtually every single vaccine. However, multiple dose vials of
the influenza shot given to children can still contain I believe 25 micrograms of mercury.
Unusual vaccines, I believe Japanese encephalitis vaccines contain mercury. And also I believe
that the tetanus shot can contain mercury.
~~~~~
2 |0.514      pid:4        person:Legislator,LegStaff
~~~~~
One, on the issue of Thimerosal and mercury. We've been presented with all kinds of conflicting
reports about mercury being in vaccinations that could cause various illness, but we've also
been provided with very definitive science that says there is no mercury. I want the presenter
to address it, and I want the witnesses for the opposition to address it.
~~~~~
3 |0.308      pid:10763     person:GeneralPublic
~~~~~
But the vast majority of childhood vaccines have no Thimerosal at all and trace amounts of
mercury at best.
~~~~~
4 |0.286      pid:10712     person:GeneralPublic
~~~~~

```

```

~~~~~
    Hi, my name's Lisa Mercury Ray. I'm from San Leandro, California. Mother of two, a chiropractor,
    and strongly oppose.
~~~~~
5 |0.281      pid:4    person:Legislator,LegStaff
~~~~~
    That there is no longer any mercury in the immunization is the information that we've been
    provided. If you believe that's different, please tell us and what science you are basing that
    on.
~~~~~
6 |0.242      pid:96   person:Legislator
~~~~~
    And that childhood vaccines contain mercury that causes autism. Opponents have made statements
    comparing vaccination to forced labor and rape. And falsely insinuated that children will be
    forcibly vaccinated and non-vaccinating parents would be thrown into prison.
~~~~~
7 |0.209      pid:10764 person:GeneralPublic
~~~~~
    Yeah, so my understanding is that there are trace amounts of not only mercury, but also other
    toxins and metals, I believe I have a list here, that are still permitted in vaccines that are
    administered within the state. Okay, so trace amounts, and then, yes sir?
~~~~~
8 |0.100      pid:10763 person:GeneralPublic
~~~~~
    So that I could be appropriately vaccinated. I, in fact, received a hepatitis A vaccination and
    a typhoid vaccine. SB 277 is not about whether vaccines are right or wrong, I wanna make that
    very clear. As a pediatrician, I give vaccines in my office every day. I receive vaccines.
~~~~~
9 |0.079      pid:10763 person:GeneralPublic
~~~~~
    My family has received vaccines. The children of all my friends have received vaccines. I give
    vaccines every day, and yet I oppose this bill. Because it's not about whether vaccines are good
    or bad. Instead, this bill asks a completely different question. Should all children in
    California be forced to be fully vaccinated?
~~~~~
10|0.074      pid:96   person:Legislator
~~~~~
    And so, when you talk about people who were vaccinated, who got pertussis, yes, we know
    pertussis wears off. All right. Measles vaccination, chicken pox vaccination, many of the other
    vaccines don't. A flu vaccine only lasts for a year. That doesn't mean that we don't get people
    vaccinated and that doesn't mean that they don't provide protection.
~~~~~

```

Figure 15: Top 10 results for Query 9 with stemming, metadata, and dedup enabled.

These query results were interesting in that only 6 of the 10 results mentioned mercury. Results 4, 8, 9, 10 don't mention mercury at all.

Query 10: “disease exposure sustained transmission opt-out personal belief exemption preventable”

```

~~~~~
1 |0.383      pid:4992   person:GeneralPublic
~~~~~
    Opting out tends to cluster. Although the overall opting-out rate is less than 3% in California,
    some schools have opt-out rates of 10 or 20%, some 50% or greater. This leaves pockets of
    children susceptible to preventable diseases. When disease exposure occurs, sustained
    transmission is the result.
~~~~~
2 |0.305      pid:4992   person:GeneralPublic
~~~~~
    And resulting disease exposure leads to sustained transmission. The consequences are disease
    among those who choose not to be vaccinated, but also in those who are vaccinated, because,

```

```

remember vaccines don't work 100% of the time. And in addition, there are those who are too
young to be vaccinated, and they may be infected.
~~~~~
3 |0.197      pid:9624      person:GeneralPublic
~~~~~
and the vaccination rates are increasing. Since the restrictions have been put on personal
belief exemptions in California, personal belief exemptions have been dropping. I'm here to
advocate for the fact that religious exemptions are a necessary religious exemptions, legitimate
religious exemptions
~~~~~
4 |0.175      pid:5439      person:GeneralPublic
~~~~~
by the Department of Health and supporters of this bill as unvaccinated. My children, like
thousands of children across the state of California, use a personal belief exemption because it
is the only one we can use to maintain access to education. Eliminating the personal belief
exemption will not increase school or public safety.
~~~~~
5 |0.166      pid:4992      person:GeneralPublic
~~~~~
Depending on the vaccine, they're about 80 to 99% effective, and immunity may wane over time.
That's why it's important to have as many in the population immunized as possible, to have the
highest level of immunity, so that if an infectious exposure occurs, transmission is limited.
~~~~~
6 |0.162      pid:4992      person:GeneralPublic
~~~~~
Depending on the vaccine, however, they are about 80 to 99% effective, and immunity may wane
over time. That's why it's important to have as many in the population immunized as possible, to
achieve the highest level of immunity, so that if an infectious exposure occurs, transmission is
limited.
~~~~~
7 |0.162      pid:10765     person:GeneralPublic
~~~~~
Only 2.5% of children entering kindergarten in 2014 had a personal belief exemption and many of
these children are vaccinated, but use a modified vaccine schedule. SB 277 eliminates the
personal belief vaccine exemption, while federal health officials tell doctors to deny medical
exemptions to 99.99% of children.
~~~~~
8 |0.162      pid:5003      person:GeneralPublic
~~~~~
Neither IPV, the Polio vaccine, nor the acellular pertussis, the whooping cough vaccine, nor the
diphtheria toxoid vaccines are capable of preventing transmission of infection. They are
intended to prevent disease symptoms only.
~~~~~
9 |0.151      pid:9624      person:GeneralPublic
~~~~~
of the certain orthodoxy, not to give them. So, personal belief exemptions are necessary.
Religious belief exemptions are necessary, simply as a matter of American freedom. Thank you.
~~~~~
10|0.145      pid:5003      person:GeneralPublic
~~~~~
My personal experience prompted me to examine peer-reviewed research on Measles and other
vaccines in much greater detail than what is generally being taught in medical schools and PhD
programs. Based on my in-depth analysis of vaccines and the protection they afford, I'm here to
testify that families who utilize personal belief exemptions do not pose a significant threat to
the public, and that personal belief exemptions must be preserved.
~~~~~

```

Figure 16: Top 10 results for Query 10 with stemming, metadata, and dedup enabled.



# Information Need Matching

Information Need 01: “I would like to know who among the legislators have been strongly opposed to the vaccination bill, and specifically, I would like to see what their arguments were”

1 |0.394 pid:72 person:Legislator

I appreciate the answer but I would like Dr. Pan, to give feedback on that. I have the right, hey, you know.

2 |0.375 pid:119 person:Legislator

I appreciate that, but I would like to ask, if it's possible for me to ask the opposition to respond to the, I'm not going to re-ask the questions and all of that. You've heard the discussions back and forth, the questions and concerns, can we give them a minute or so just to respond to those questions in general? You can pick whichever ones you'd like.

3 |0.369 pid:96 person:Legislator

I would like Mrs. Fisher to answer the question too.

4 |0.364 pid:52 person:Legislator

I do think that this is-- this measure is very necessary to make our profession just a little bit easier on teachers who go to work day in and day out to do the best for their students. So as a former school teacher, I appreciate what you've done and what you will continue to do and with that Madam Chair I would like to move the bill.

5 |0.363 pid:100 person:Legislator

We don't want to have any problems with the fire marshal, and want to make sure that everybody has their chance to be heard. And I want to thank you all for coming today. I know that everyone here has a busy schedule. I like to see little ones.

6 |0.350 pid:96 person:Legislator

to the vaccination, we would then do a medical exemption as well because it would be dangerous for them to receive the vaccine.

7 |0.343 pid:100 person:Legislator

This is a special order of business, SB 277 by Senators Pan and Allen. And before we turn to the authors and invite them up, I would like to note this bill has been previously heard both by the senate health and senate education committees for issues that are specific to the jurisdiction of those committees. We are not here to rehear the questions and the issues that were discussed there, that were appropriately discussed.

8 |0.342 pid:103 person:Legislator

The problem that we're gonna have because of codes, fire marshal codes, we can't have this many people in, so as soon as you've either testified, and if you don't have a seat, you can leave to make room for those who would like, or to the side door, to allow those who would like to come in and speak, 'cause we need to have the flow, but we also wanna make it fair for everybody. Anyway, I'm sorry, go ahead.

9 |0.338 pid:100 person:Legislator

Thank you very much, is there anyone else who would like to speak, if someone is here and want's to speak in support who didn't get here in time please come forward, if there are any people still who would like to speak in opposition who haven't we'll give you one moment to step forward, I want to make sure everybody has been heard.

```
~~~~~
10|0.336 pid:96   person:Legislator
~~~~~
```

Actually, I would like to have Dr. Blumberg address this particular issue.

Figure 17: Top 10 results for Query 10 with stemming, metadata, and dedup enabled.

InfoNeed01.txt, shown in Figure 17 with no stopwords. Starting from no options enabled and moving up to this didn't alter the results very much, or really make them more relevant. No one really states their opposition to the bill, until we add stopwords in Figure 18. These stopwords probably help filter the information need as well, focusing on only the relevant terms. Of the 10 results, results 2 and 3 are directly relevant, giving us a precision of 0.2. While this better than previously, it's still not a confident number.

```
~~~~~
1 |0.434 pid:10763   person:GeneralPublic
~~~~~
```

I don't mean to be argumentative, but how do they document that they received vaccines before they came here? I don't-

```
~~~~~
2 |0.408 pid:29   person:Legislator
~~~~~
```

So arguments can be made on both sides and I'm not gonna get into the science of it because we could debate it all day, and it's really not about whether vaccines are right or wrong.

```
~~~~~
3 |0.364 pid:62   person:Legislator
~~~~~
```

I do believe in vaccines, but I just don't think this is the right way to get at it.

```
~~~~~
4 |0.342 pid:96   person:Legislator
~~~~~
```

Well I think certainly that could be something we can take a look at, 'cause as I said, it has never been done by the Department without legislative action in the past.

```
~~~~~
5 |0.332 pid:25   person:Legislator
~~~~~
```

Not just to those who haven't been vaccinated, but to even those who have been vaccinated. Or who can't be vaccinated because they suffer from an autoimmune disease.

```
~~~~~
6 |0.316 pid:80   person:Legislator
~~~~~
```

Looking at the arguments, listening to the testimony here and also listening to all the constituents that went through the room, and also I want to thank my constituents who came up to testify for and against this bill.

```
~~~~~
7 |0.302 pid:100  person:Legislator
~~~~~
```

Thank you. And you're opposed to the bill, thank you.

```
~~~~~
8 |0.297 pid:83   person:Legislator
~~~~~
```

I didn't mean this bill, I'm talking about the last bill.

```
~~~~~
9 |0.297 pid:100  person:Legislator
~~~~~
```

And I think, given the argument that education is fundamental, right, in California that we need to show a compelling state interest and I think that that has been shown. That's...when we talk about what some of the languages that the US Supreme Court has already identified is that there is the right of the public to be protected.

```
~~~~~
10|0.282 pid:100  person:Legislator
~~~~~
```

```

~~~~~
Thank you very much you are opposed to the bill, sir?
~~~~~

```

Figure 18: Top 10 results for InfoNeed01 with stopwords-long.txt, stemming, metadata, and dedup enabled.

Information Need 02: “I would like to find out the chief arguments of the proponents of the legislation. Why did they think it was important to require the vaccination of every child attending school in the state?”

```

~~~~~
1 |0.370 pid:96   person:Legislator
~~~~~
    They may, I believe they may attend school, yes.
~~~~~
2 |0.321 pid:57   person:Legislator
~~~~~
    Yeah. If my child has hepatitis B, may they attend school?
~~~~~
3 |0.307 pid:57   person:Legislator
~~~~~
    But if they're not vaccinated for hepatitis B, can they attend school?
~~~~~
4 |0.302 pid:96   person:Legislator
~~~~~
    Let me be clear, SB 277 does not mandate vaccinations. But, that choice requires responsibility.
    And, we have the responsibility to protect all children attending school by requiring
    vaccination when children go attend school with other children. Thank you, and I'd like to now
    pass this on to my co-author, joint-author, Senator Ben Allen. Thank you.
~~~~~
5 |0.269 pid:96   person:Legislator
~~~~~
    Well, if they have a medical exemption, then they will be able to attend school under this bill.
~~~~~
6 |0.259 pid:103  person:Legislator
~~~~~
    So how many opponents do we have up here? We should have equal amount. So we need two more
    proponents. Come up. Okay, I think we have four and four now, I think.
~~~~~
7 |0.251 pid:10763 person:GeneralPublic
~~~~~
    I don't mean to be argumentative, but how do they document that they received vaccines before
    they came here? I don't-
~~~~~
8 |0.248 pid:96   person:Legislator
~~~~~
    And that you are, even though they have not actually achieved the requirement, that you are
    working with the family with the intention to actually achieve that requirement in the near
    future. And that therefore, that's why they should be permitted to start school now and that
    they can continue to attend.
~~~~~
9 |0.241 pid:57   person:Legislator
~~~~~
    When you're looking at hepatitis B. So in your current bill, if my child has hepatitis B, can
    they attend school?
~~~~~
10|0.241 pid:96   person:Legislator
~~~~~
    vaccinations when enrolling them in the school attended by other children, but will instead have
    to home school them. SB 277 retains a medical exemption for children for whom a licensed

```

physician determines cannot receive a vaccine. And these children who cannot be vaccinated depend on other children at school to be vaccinated

Figure 19: Top 10 results for InfoNeed02 with stopwords-medium.txt, stemming, and dedup enabled.

For this information need, my first query iteration was "support approve require vaccination" which was run on the system without stemming but with stopwords removal via stopwords-medium.txt. Results are shown below in Figure 20.

```
~~~~~
1 |0.224   pid:100   person:Legislator
~~~~~
    Others in support let's have you line up over here. We're going to make it very, everybody name,
    affiliation, and obviously this is to indicate your sup- that you are in support. Strong
    support, very strong support, modest support, but just very briefly.
~~~~~
2 |0.181   pid:10763  person:GeneralPublic
~~~~~
    I would say we want no more outbreaks. Again, 130 cases amongst 39 million people. And it
    stopped because of our high vaccination rate. And we have a high vaccination rate because we've
    informed parents. And we have a higher vaccination rate because of 2109 which, no, I did not
    support initially. But the impact was at least partially positive, and it's working.
~~~~~
3 |0.168   pid:25    person:Legislator
~~~~~
    Several of the people who testified in opposition said they were opposed because they have an
    autoimmune disease or disorder. Would your bill permit, would the medical exception allow
    somebody with an autoimmune disease not require vaccination?
~~~~~
4 |0.159   pid:9619   person:GeneralPublic
~~~~~
    My name is Suzanne Russin, in the 1950's I participated in the polio vaccination trials, and I
    strongly support this bill.
~~~~~
5 |0.146   pid:119    person:Legislator
~~~~~
    With respect to the level of vaccination in the state of California. Is it or is it not high?
~~~~~
6 |0.145   pid:96     person:Legislator
~~~~~
    Certainly, these are all vaccinations for diseases that can certainly put children at risk so
    for example, polio is a required vaccination, hopefully we don't want polio to come back, we
    also require DPT, which is pertussis, tetanus, and diphtheria, and while some people have
    pointed out that tetanus
~~~~~
7 |0.143   pid:96     person:Legislator
~~~~~
    So, I would argue is that's one of the reasons why we require hepatitis B vaccination because
    there's other children who were vaccinated against hepatitis B and their chance of getting
    hepatitis B from another child is extremely low.
~~~~~
8 |0.138   pid:10820  person:GeneralPublic
~~~~~
    Teresa Stark on behalf of Kaiser Permanente, in strong support. And also here as a mother in
    strong support.
~~~~~
9 |0.134   pid:105    person:Legislator
~~~~~
    We don't require tourists coming into California, and by the way, we're the number one tourist
```

```

destination. Our economy depends on it.
~~~~~
10|0.134    pid:96    person:Legislator
~~~~~
to the vaccination, we would then do a medical exemption as well because it would be dangerous
for them to receive the vaccine.
~~~~~

```

Figure 20: Top 10 results for InfoNeed02 with stopwords-medium.txt, stemming, and dedup enabled.

This query had mixed results with only 6 results clearly in support of the bill. However, the supporting results did show the arguments of preventing outbreaks, that high vaccination rates prevent outbreaks, and that the diseases themselves put children at risk.

With that in mind the revised query "support approve require vaccination prevent disease" was run on the system with stopwords-medium.txt and produced the following results in Figure 21:

```

1 |0.255    pid:25    person:Legislator
~~~~~
Several of the people who testified in opposition said they were opposed because they have an
autoimmune disease or disorder. Would your bill permit, would the medical exception allow
somebody with an autoimmune disease not require vaccination?
~~~~~
2 |0.201    pid:96    person:Legislator
~~~~~
This is really about protecting the children in school and that's why we wanted to get children
vaccinated to prevent these kinds of outbreaks and the spread of disease.
~~~~~
3 |0.200    pid:96    person:Legislator
~~~~~
95 percent from Measles, prevent the disease from spreading throughout the community. And, that
is why every state in the United States, including California, has passed laws requiring
vaccination for school entry; to protect our children and the community.
~~~~~
4 |0.190    pid:100   person:Legislator
~~~~~
Others in support let's have you line up over here. We're going to make it very, everybody name,
affiliation, and obviously this is to indicate your sup- that you are in support. Strong
support, very strong support, modest support, but just very briefly.
~~~~~
5 |0.183    pid:96    person:Legislator
~~~~~
the recommendation it's to immunize them as children because immunizations don't work after you
catch the disease, you have to do it before you catch the disease, but every one of these
immunizations is one that I think we can talk about that actually causes harm and the benefits
of vaccination certainly far outweigh having the disease continue.
~~~~~
6 |0.171    pid:96    person:Legislator
~~~~~
And that's 95% for measles, to prevent the disease from spreading throughout the community,
known as community immunity or herd immunity. And that is why every state in the United States
including California, has passed laws requiring vaccination for school entry to protect our
children and our community.
~~~~~
7 |0.163    pid:10763  person:GeneralPublic
~~~~~
Even if every school child is fully vaccinated, whooping cough and flu will spread through
schools until we get better vaccines for our kids and our teens. Our current measles vaccination
rate is easily high enough to prevent school spread. We have some of the highest vaccination
rates in history.
~~~~~

```

```

~~~~~
8 |0.163    pid:9613    person:GeneralPublic
~~~~~
    We don't change the curriculum to prevent parents from having to choose between a curriculum
    they don't like and homeschooling. It's problematic to say we need to do away with a reasonable
    health and safety requirement, vaccination to prevent parents from having to choose between
    following the requirement and choosing another educational option, again, home schooling or
    independent study.
~~~~~
9 |0.154    pid:96    person:Legislator
~~~~~
    when they actually had a tremendous drop in vaccination rates during the chaos after the fall of
    the Soviet Union, there actually was diphtheria that reemerged, in Russia, so hopefully we will
    never see diphtheria come back to the United States, it's known as the "silent killer," but
    that's another disease we require.
~~~~~
10|0.153    pid:96    person:Legislator
~~~~~
    That is correct, as long as we can achieve what is called community immunity, herd immunity that
    will prevent the disease from actually reaching that child and that's what we want to try and do
    to protect the children that are unvaccinated.
~~~~~

```

Figure 21: Top 10 results for InfoNeed02 with stopwords-medium.txt, stemming, and dedup enabled.

This resulted in 7 supportive arguments and one confused person who claimed that whooping cough would spread through schools despite high vaccination rates until we got "better vaccines". A few of the supportive arguments included the term "immunity", so with that in mind the final query "support approve require vaccination prevent disease immunity" was supplied to the system, with stopword removal via stopwords-medium.txt and with stemming turned on, as shown in Figure 22.

```

~~~~~
1 |0.321    pid:96    person:Legislator
~~~~~
    That is proof of why SB 277 is necessary to protect the public from vaccine preventable
    contagions. In addition, vaccine preventable infections are particularly dangerous to children
    who have cancer, may have transplants, or immune diseases. These children are particularly
    vulnerable to infection and they depend on community immunity to prevent the contagion from
    actually reaching them.
~~~~~
2 |0.304    pid:96    person:Legislator
~~~~~
    And that's 95% for measles, to prevent the disease from spreading throughout the community,
    known as community immunity or herd immunity. And that is why every state in the United States
    including California, has passed laws requiring vaccination for school entry to protect our
    children and our community.
~~~~~
3 |0.302    pid:96    person:Legislator
~~~~~
    the recommendation it's to immunize them as children because immunizations don't work after you
    catch the disease, you have to do it before you catch the disease, but every one of these
    immunizations is one that I think we can talk about that actually causes harm and the benefits
    of vaccination certainly far outweigh having the disease continue.
~~~~~
4 |0.291    pid:9613    person:GeneralPublic
~~~~~
    First to remind you the right to education, as was already mentioned here, also covered those
    that can't be immunized for medical reasons. We have a medical exemption that requires an MD

```

```

approval, nothing more. And those with medical reasons not to vaccinate have to face the choice
between going to a school that has low immunization rates and risking an outbreak, or not going.
~~~~~
5 |0.290    pid:9621    person:GeneralPublic
~~~~~
Good morning Catherine Flores-Martin, Executive Director for the California Immunization
Coalition working to protect all Californians from preventable diseases, in support.
~~~~~
6 |0.278    pid:96     person:Legislator
~~~~~
And maybe many other conditions, who cannot be immunized. And so, if a child cannot be
immunized, they depend of the rest of us to be immunized so that those diseases don't reach
them.
~~~~~
7 |0.270    pid:96     person:Legislator
~~~~~
That is correct, as long as we can achieve what is called community immunity, herd immunity that
will prevent the disease from actually reaching that child and that's what we want to try and do
to protect the children that are unvaccinated.
~~~~~
8 |0.268    pid:96     person:Legislator
~~~~~
No. The vaccine requirements have not been changed in terms of which vaccines are required at
what time in this particular bill, it's based on previous actions of this legislature. Every
single vaccine that's required in the state of California has been passed by this legislature
and signed by a governor.
~~~~~
9 |0.267    pid:5003    person:GeneralPublic
~~~~~
Neither IPV, the Polio vaccine, nor the acellular pertussis, the whooping cough vaccine, nor the
diphtheria toxoid vaccines are capable of preventing transmission of infection. They are
intended to prevent disease symptoms only.
~~~~~
10|0.262    pid:25     person:Legislator
~~~~~
Not just to those who haven't been vaccinated, but to even those who have been vaccinated. Or
who can't be vaccinated because they suffer from an autoimmune disease.
~~~~~

```

Figure 22: Top 10 results for InfoNeed02 with stopwords-medium.txt, stemming, and dedup enabled.

This by far returned the best results with 9 out of 10 utterances presenting clear arguments in relation to the information need “I would like to find out the chief arguments of the proponents of the legislation. Why did they think it was important to require the vaccination of every child attending school in the state?”

## Analysis and Conclusions

Overall the inclusion of a file of stopwords to remove resulted in increased cosine similarity values, which makes sense given that it reduces the overall number of terms per document, thereby increasing the relative weight of each remaining term. Stemming greatly improved results, particularly in Query 4, which involved parents jumping through hoops. In queries that had very distinct keywords that did not have a large variety of possible stems, such

as Query 1 with “Disneyland” and “Measles” stemming did not have as great an effect on the query results since the search terms were not greatly affected by the stemming algorithm. Stemming produced results that felt more intuitively correct at a high level.

Queries with many words in them -- particularly words that one would not necessarily consider stop words -- garnered lengthy results that also included a lot of words. Increasing the words in a query to the point that the query is a document in the collection itself resulted in the retrieval system grabbing the exact document in question. The cosine similarity was, of course, 1.00, until heavier filtering methods -- stopwords, stemming, and metadata in particular -- were applied, which ever so slightly altered the score of the top document.

For the majority of queries, we were able to find parameters that resulted in 6-10 of the 10 results appearing relevant to the search. In particular for the information need queries we were able to tune the precision since they allowed us to iteratively change out query until we received mostly relevant results. In general the most effective system for this dataset incorporated stemming and stopword removal, and generating a custom stopword list specific to this dataset could result in higher precision.

Deduplication in this dataset was a concern of ours. We worried that the presence of the same utterance in the collection more than once would skew the results. We didn’t notice any duplicates, but some *almost* duplicates. We had to inspect very closely to make sure, but there were slight changes in wording that Python could not catch. Furthermore, we may not have seen any in most of the top 10 because the duplication would have increased the document frequency of the terms in those utterances, lowering its weight.