

Content Based Recommender System as a Search Engine to Find Federal Contract Opportunities Based on Topic Similarities

Bibor Szabo

Inspiration

Search by:

- Keyword
- Solicitation ID
- Agency name

Filter results

The screenshot displays the SAM.gov website. At the top, the SAM.gov logo is visible with a 'beta' tag. To the right of the logo are navigation icons: a hamburger menu, a magnifying glass for search, a bell for notifications, and a 'Sign in' link. Below the header is a dark blue search bar. On the left of the search bar is a dropdown menu currently set to 'Contract Opportunities'. The search input field contains the text 'I'm looking for..'. To the right of the input field is a blue 'Search' button. Below the search bar, the 'Contract Opportunities' section is highlighted. It features a title 'Contract Opportunities' and a paragraph: 'View procurement notices from Federal contracting offices to find opportunities to bid for Federal business. These include pre-solicitation notices, solicitation notices, award notices and sole source notices. Manage your searches and track notices by creating an account.' To the right of this text are two circular icons: one with a building icon labeled 'Vendor Collaboration' and another with a calendar icon labeled 'Small Business Events'. At the bottom of the screenshot is a white box titled 'Contract Opportunities Search Tips' containing two bullet points: 'Search by keyword, solicitation ID, or agency name. Use filters to narrow your results.' and 'Accounts are not required to view contract opportunities. Create an account to save searches, add your entity to an opportunity's Interested Vendors List, and receive automatic updates when opportunities change.'


SAM^{beta} GOV


☰ 🔍 🔔 Sign in

Contract Opportunities ▾ I'm looking for.. 🔍 Search

Contract Opportunities

View procurement notices from Federal contracting offices to find opportunities to bid for Federal business. These include pre-solicitation notices, solicitation notices, award notices and sole source notices. Manage your searches and track notices by creating an account.

 Vendor Collaboration

 Small Business Events

Contract Opportunities Search Tips

- Search by keyword, solicitation ID, or agency name. Use filters to narrow your results.
- Accounts are not required to view contract opportunities. Create an account to save searches, add your entity to an opportunity's Interested Vendors List, and receive automatic updates when opportunities change.

Goals of the Project

1. Build an application that, based on word or topic similarities, would recommend contract notifications of interest.
2. Observe, how well a content based recommender system is capable of finding not only similar but also relevant notifications.

Data Source

Beta.SAM.gov

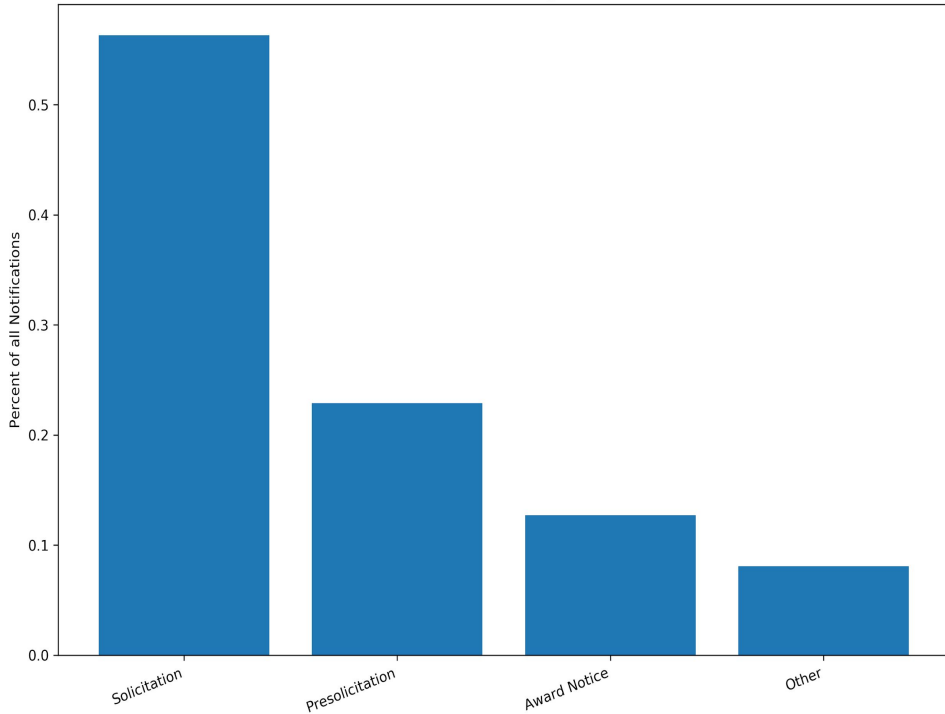
- ❑ Used the *Get Opportunities Public API*
- ❑ Gathered 55,000 documents

Notifications

- ❑ Between January 1, 2020 - June 8, 2020
- ❑ Daily updates

Data

Notification Categories



Notification Types

- ❑ 55 % Solicitations
 - ❑ Solicitations
 - ❑ Combined Synopsis/ Solicitations
- ❑ 23% Pre-solicitations
 - ❑ Pre-solicitations
 - ❑ Sources Sought
- ❑ 12% Award Notice
- ❑ Other
 - ❑ Special Notices
 - ❑ Justification
 - ❑ Sale of Surplus Property

The Challenge

Raw Data

10769
George Washington Memorial Parkway
6550
48--VALVE, SOLENOID
8194
Solicitation for 2 each NSN: 2915-00-632-7932
30533
Air Combat Evolution (ACE)
38180
Software Licenses for Coinbase Supplies
7189 Indefinite Delivery Indefinite Quantity (IDIQ), Multiple Award Task Order Contract
tion Services within the Louisville Districts Area of Responsibility
25424
FUSE, CONTROL CRKT
39205
59--LIGHT EMITTING DIOD

Cleaned Data

'Solicitation For Each Nsn'

'Indefinite Delivery Indefinite Quantity Idiq Multiple Award Task Order
Contract Matoc For Design Bid Build Construction Services Within
The Louisville Districts Area Of Responsibility'

'Fuse Control Crkt'

Model 1 -- Frequency Based Embeddings

Tf-Idf Vectorizer

- ❑ Stop words removed
- ❑ N-grams: Uni-grams, Bi-grams, Tri-grams

Cosine Similarity Matrix

- ❑ Cosine similarity of 0 \Rightarrow no similarity
- ❑ Cosine similarity of 1 \Rightarrow identical vectors

Model 2 -- Prediction Based Embeddings

Doc2Vec Model

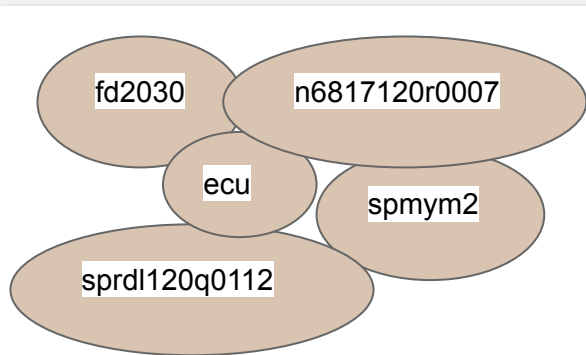
- ❑ Creates feature vectors for each document
- ❑ Used to find similarities between sentences, paragraphs or longer documents
- ❑ Stopwords left in - for added context

Cosine Similarity

- ❑ For measuring distance between document vectors

Limitations

One word titles, letter-number sequences do not provide context.



Evaluation and comparison of models is difficult.

Good
(in statistical terms)



Relevant
(in human terms)

Accuracy
of classification



Accurately and
consistently
finding similar
notifications

Let's Investigate!

What next?

- ❑ Despite the limited amount of context data – the recommender system finds similar and relevant documents in the database

Next:

- ❑ Automate daily data retrieval and model-rebuild – to keep the content of the search engine up-to-date
- ❑ Hypertuning the Doc2Vec model – to find the most effective combination of tuning parameters
- ❑ Add further features to the application – to increase its functionality

References

Natural Language Processing

<https://www.oreilly.com/library/view/applied-text-analysis/9781491963036/ch04.html>

<https://towardsdatascience.com/document-embedding-techniques-fed3e7a6a25d#da9f>

<https://medium.com/wisio/a-gentle-introduction-to-doc2vec-db3e8c0cce5e>

Content Based Filtering

<http://recommender-systems.org/content-based-filtering/>

<https://heartbeat.fritz.ai/recommender-systems-with-python-part-i-content-based-filtering-5df4940bd831>

Information Retrieval Systems

<https://www.youtube.com/watch?v=F5mRW0jo-U4>

https://developer.mozilla.org/en-US/docs/Learn/Server-side/Django/Home_page

<https://docs.djangoproject.com/en/3.0/>

<https://docs.djangoproject.com/en/dev/topics/db/queries/>

https://www.tutorialspoint.com/django/django_url_mapping.htm