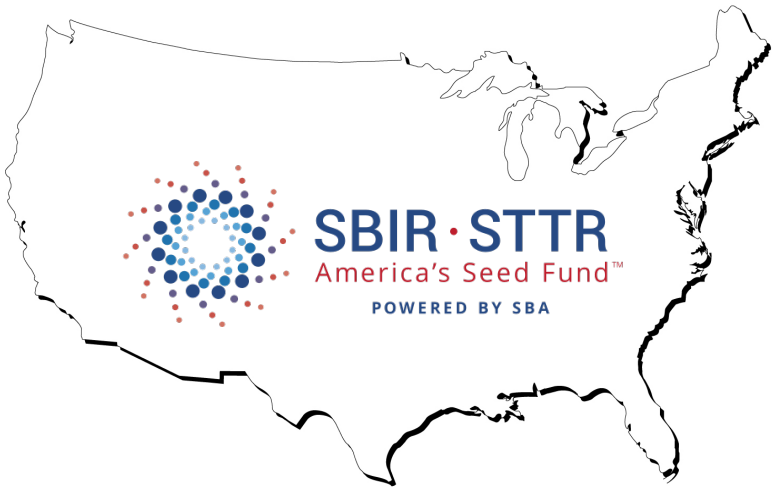
Instagrant

Helping small business predict federal grant awards

Cedric Herman

Small Businesses are Big!

- Vital role to the US economy
- ~90% of US businesses < 20 employees*



- Equity free!
- High risk - High Reward
- 35,000 applicants/year
- Must have potential for commercialization

SBIR = Small Business Innovative Research
STTR = Small Business Technology Transfer

*2014 US Census Bureau Data

SBIR-STTR grant

Phase I

Concept
Development
6 months
< $150,000

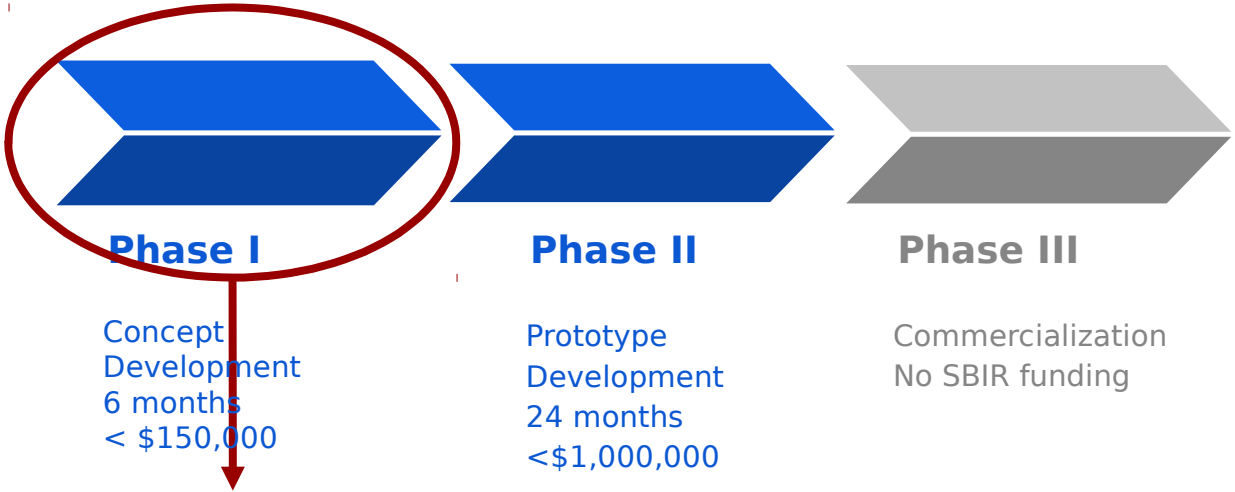
Phase II

Prototype
Development
24 months
<$1,000,000

Phase III

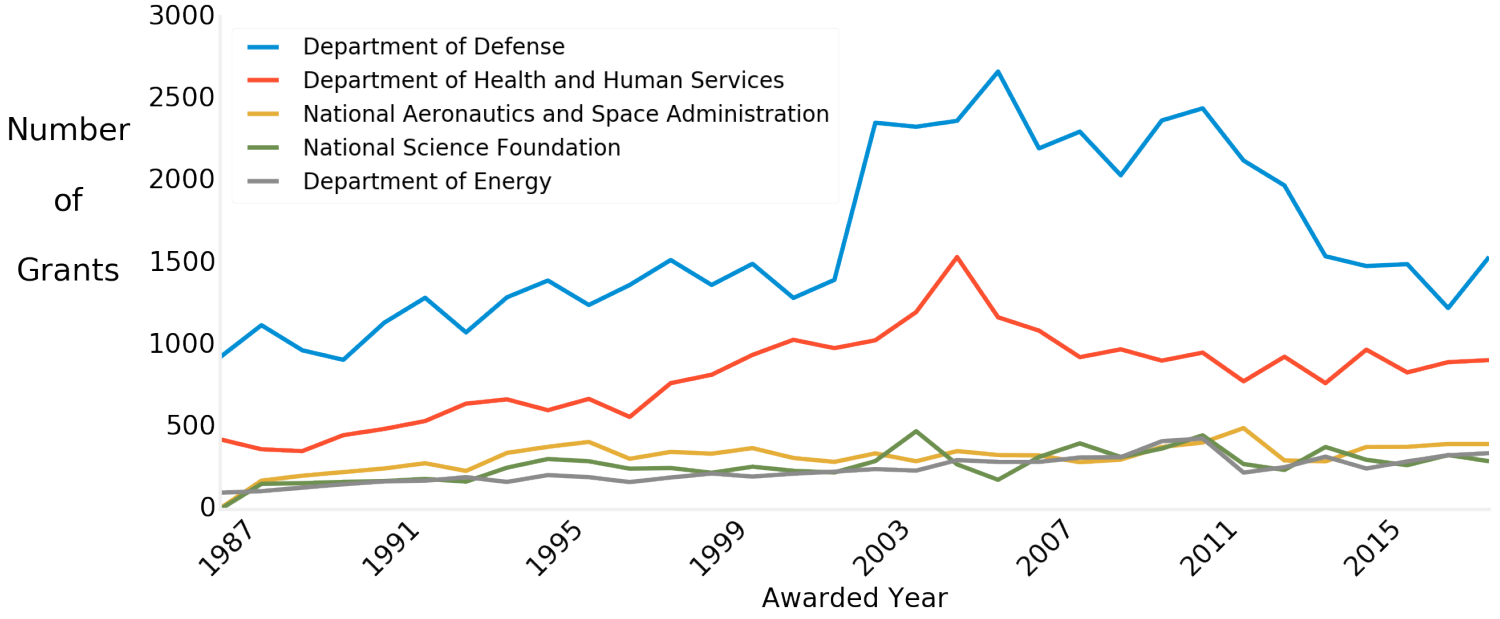
Commercialization
No SBIR funding

SBIR-STTR grant



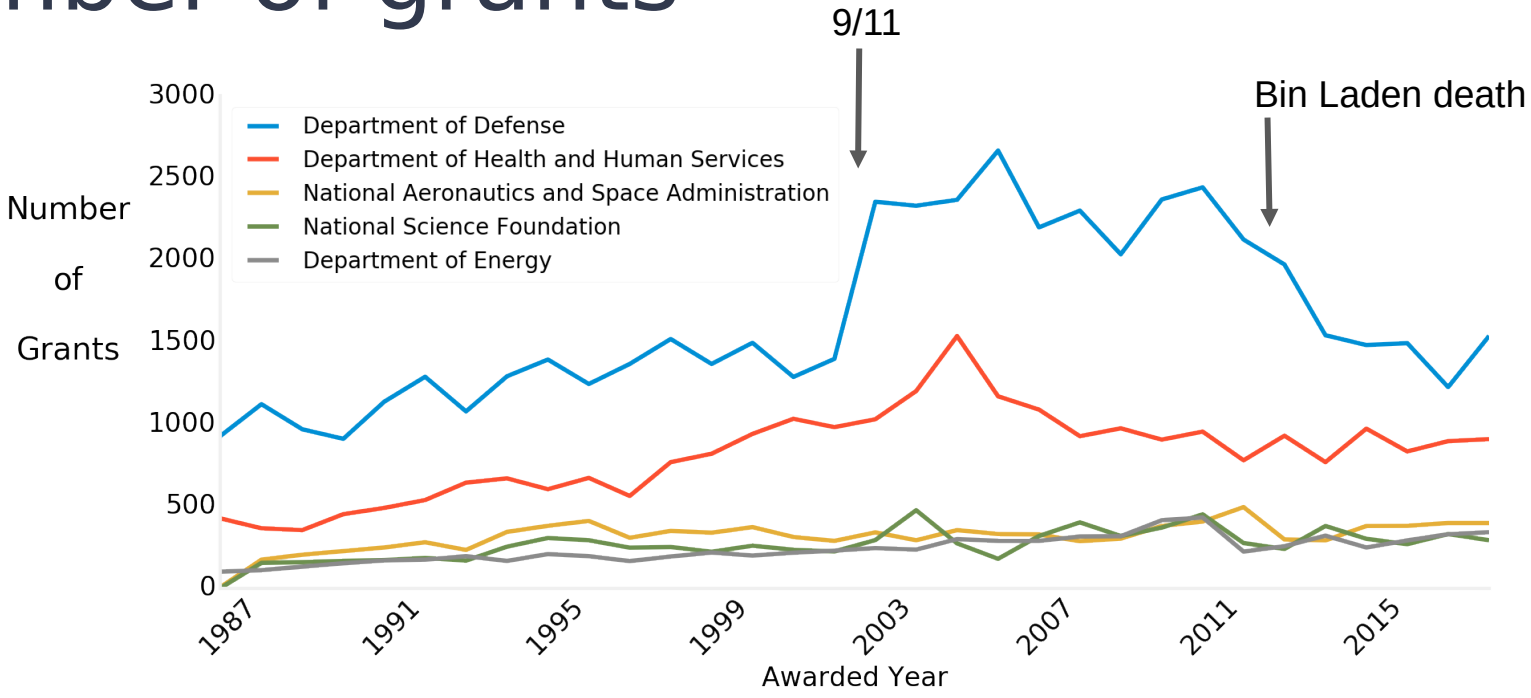
How much should a small business apply for?

Number of grants



11 participating Federal Agency in 2017

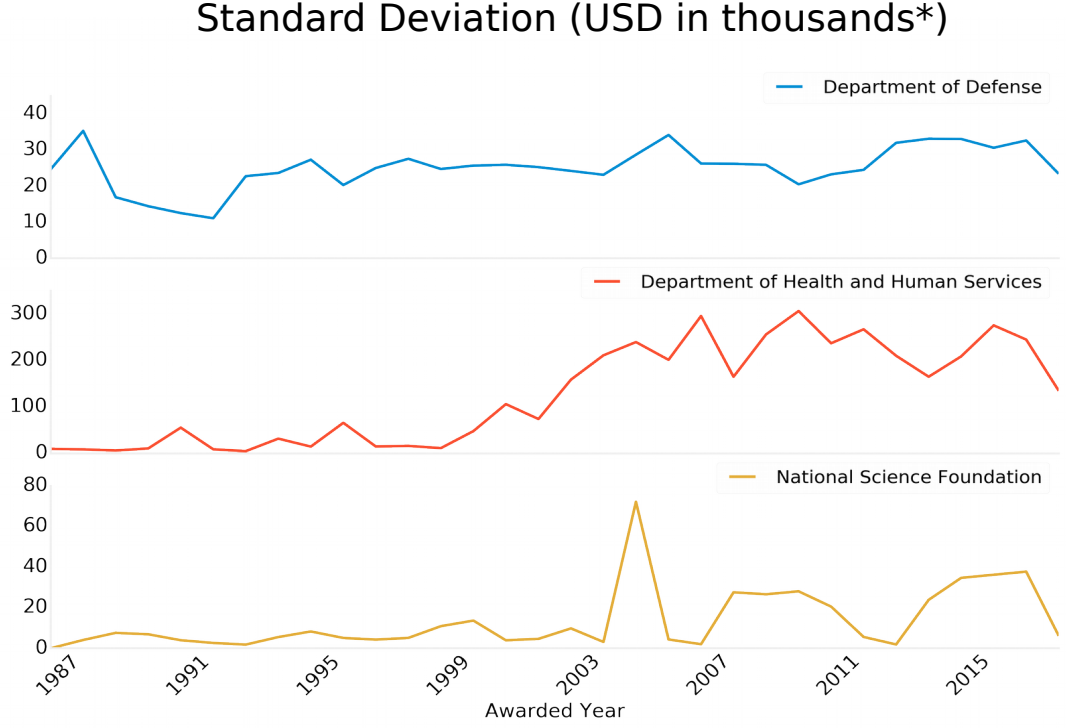
Number of grants



11 participating Federal Agency in 2017

Awarded USD

- DOD consistent broad distribution of dollar amount
- Prior to 2000, HHS awardees were granted the same dollar amount systematically
- NSF, very little variance until 2003



*Target dollar amount adjusted for inflation using Consumer Price Index (CPI)

Data workflow

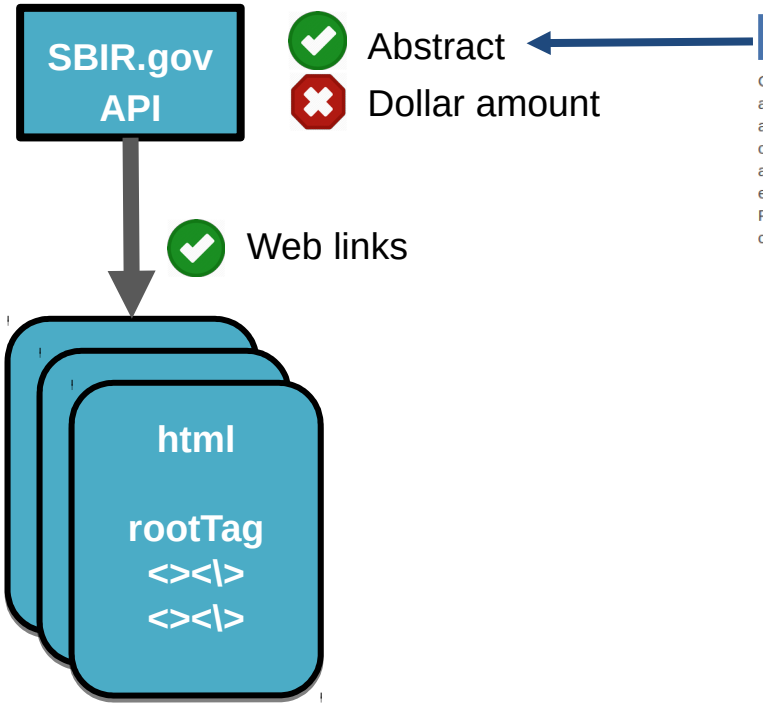
Abstract

Dollar amount

Abstract

Conduct feasibility study of design/development of terrestrial sensor system prototype for small tactical units for local area/perimeter security. Seismic and acoustic sensors will be analyzed for detection performance in various operational environment to include rural and urban areas to ensure awareness of personnel movements to cover a 500 sq ft area. An analysis of predicted range/sensitivity will be studied in order to determine operationally suitable employment. The effort will analyze new generation advanced data capture systems, ultra-low power microcontrollers, previous and new advanced detection algorithms, small low power comm modules (Bluetooth 4.0, Zigbee, et al) and smart phone or tablet sized controller for eventual incorporation with an android type device. Specifically, the effort will address known shortfall with previous systems in False Alarm Rate, low Probability of Detection, target discrimination, threshold/sensitivity adjustment, power consumption, form factor, SWaP, complexity of employment, concealment, cyber protection, physical tamper and ease of use.

Data workflow

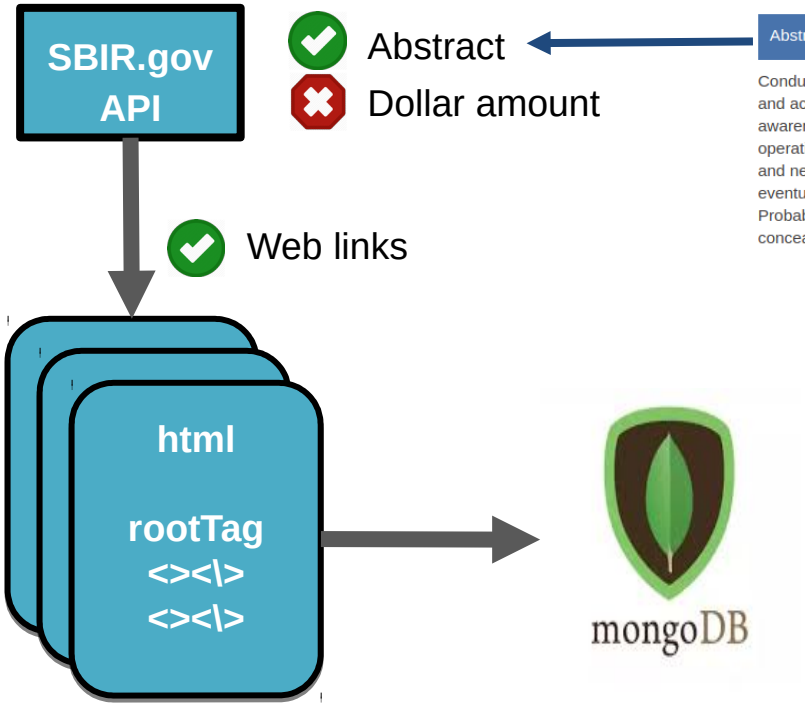


Abstract

Conduct feasibility study of design/development of terrestrial sensor system prototype for small tactical units for local area/perimeter security. Seismic and acoustic sensors will be analyzed for detection performance in various operational environment to include rural and urban areas to ensure awareness of personnel movements to cover a 500 sq ft area. An analysis of predicted range/sensitivity will be studied in order to determine operationally suitable employment. The effort will analyze new generation advanced data capture systems, ultra-low power microcontrollers, previous and new advanced detection algorithms, small low power comm modules (Bluetooth 4.0, Zigbee, et al) and smart phone or tablet sized controller for eventual incorporation with an android type device. Specifically, the effort will address known shortfall with previous systems in False Alarm Rate, low Probability of Detection, target discrimination, threshold/sensitivity adjustment, power consumption, form factor, SWaP, complexity of employment, concealment, cyber protection, physical tamper and ease of use.

Web scraper ~166k awards (multiprocessing)

Data workflow

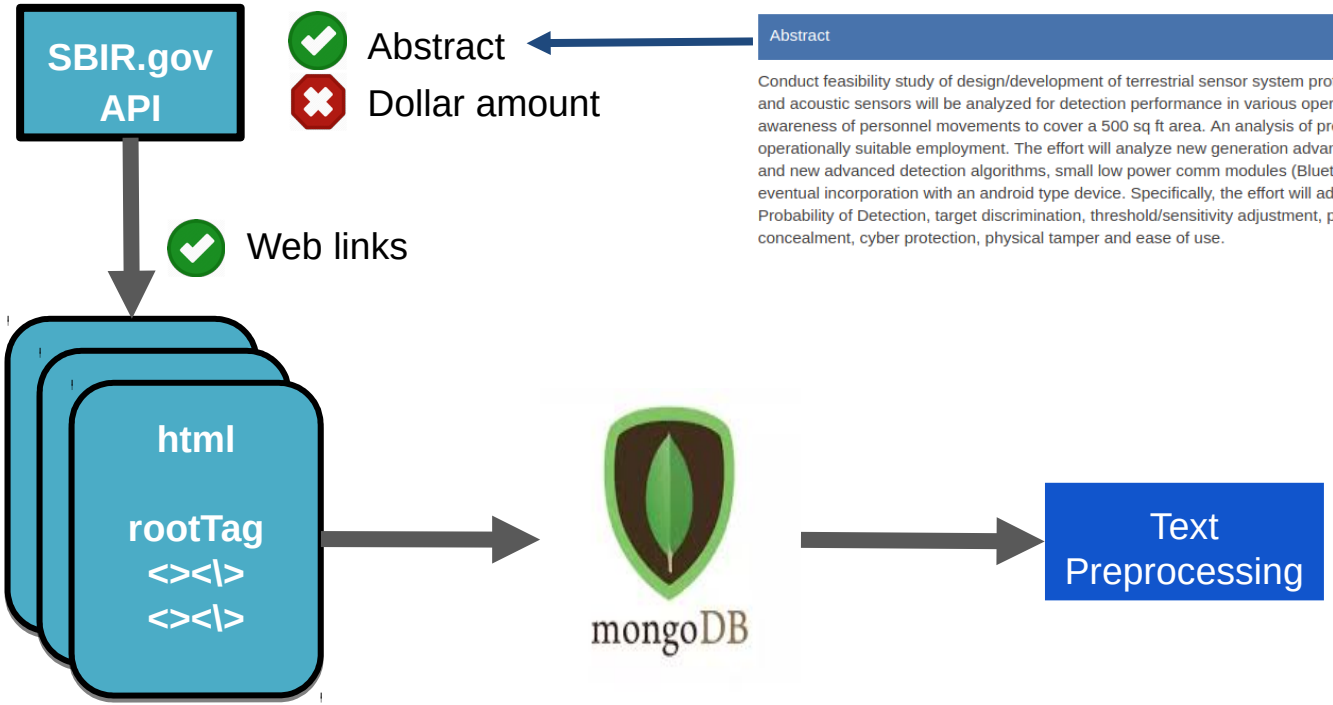


Abstract

Conduct feasibility study of design/development of terrestrial sensor system prototype for small tactical units for local area/perimeter security. Seismic and acoustic sensors will be analyzed for detection performance in various operational environment to include rural and urban areas to ensure awareness of personnel movements to cover a 500 sq ft area. An analysis of predicted range/sensitivity will be studied in order to determine operationally suitable employment. The effort will analyze new generation advanced data capture systems, ultra-low power microcontrollers, previous and new advanced detection algorithms, small low power comm modules (Bluetooth 4.0, Zigbee, et al) and smart phone or tablet sized controller for eventual incorporation with an android type device. Specifically, the effort will address known shortfall with previous systems in False Alarm Rate, low Probability of Detection, target discrimination, threshold/sensitivity adjustment, power consumption, form factor, SWaP, complexity of employment, concealment, cyber protection, physical tamper and ease of use.

Web scraper ~166k awards (multiprocessing)

Data workflow

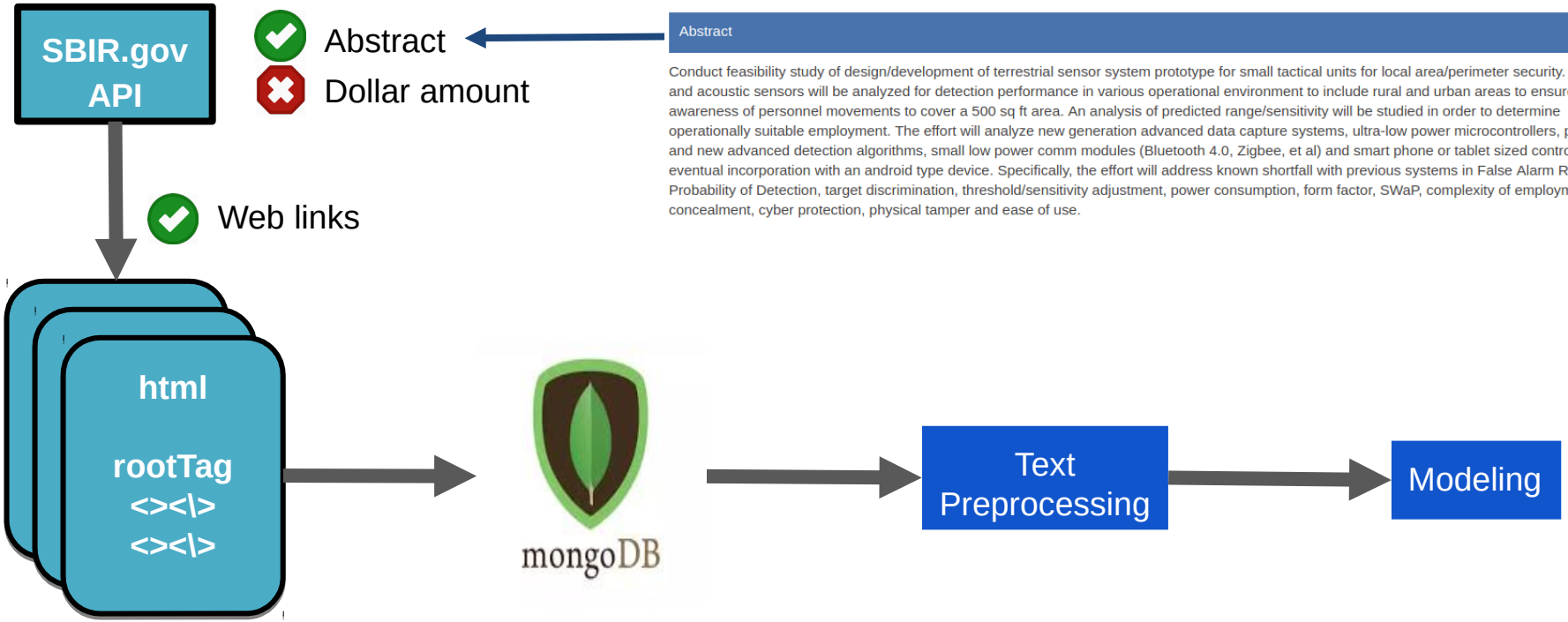


Abstract

Conduct feasibility study of design/development of terrestrial sensor system prototype for small tactical units for local area/perimeter security. Seismic and acoustic sensors will be analyzed for detection performance in various operational environment to include rural and urban areas to ensure awareness of personnel movements to cover a 500 sq ft area. An analysis of predicted range/sensitivity will be studied in order to determine operationally suitable employment. The effort will analyze new generation advanced data capture systems, ultra-low power microcontrollers, previous and new advanced detection algorithms, small low power comm modules (Bluetooth 4.0, Zigbee, et al) and smart phone or tablet sized controller for eventual incorporation with an android type device. Specifically, the effort will address known shortfall with previous systems in False Alarm Rate, low Probability of Detection, target discrimination, threshold/sensitivity adjustment, power consumption, form factor, SWaP, complexity of employment, concealment, cyber protection, physical tamper and ease of use.

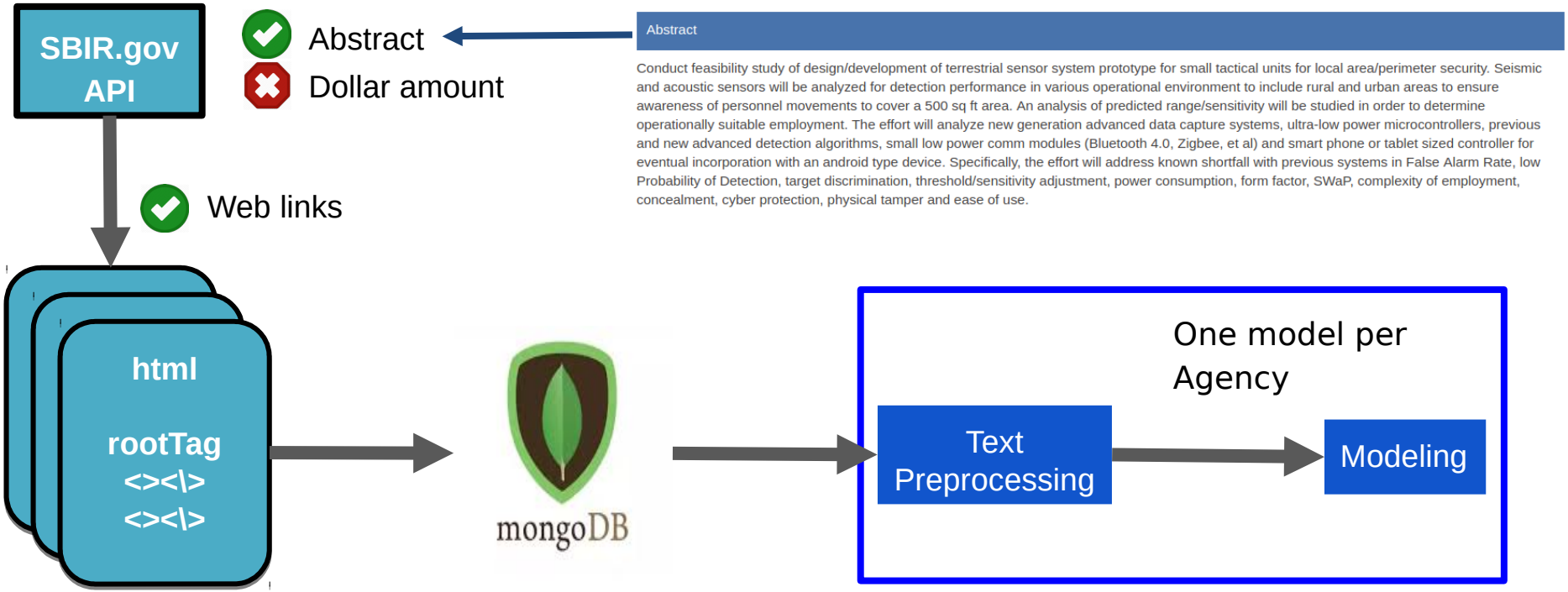
Web scraper ~166k awards (multiprocessing)

Data workflow



Web scraper ~166k awards (multiprocessing)

Data workflow

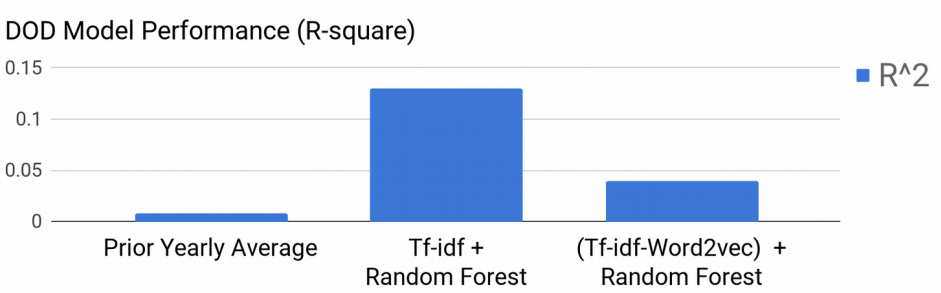


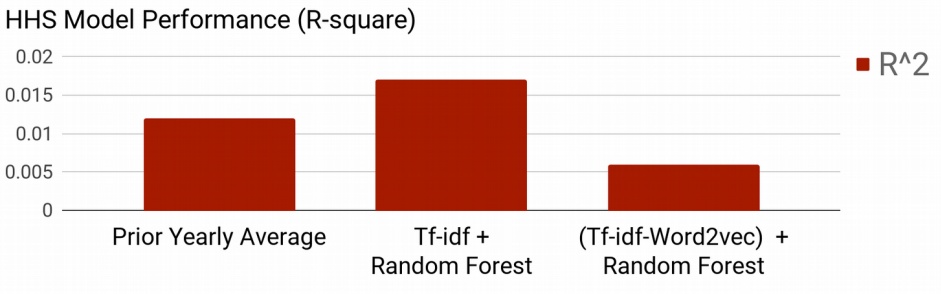
Web scraper ~166k awards (multiprocessing)

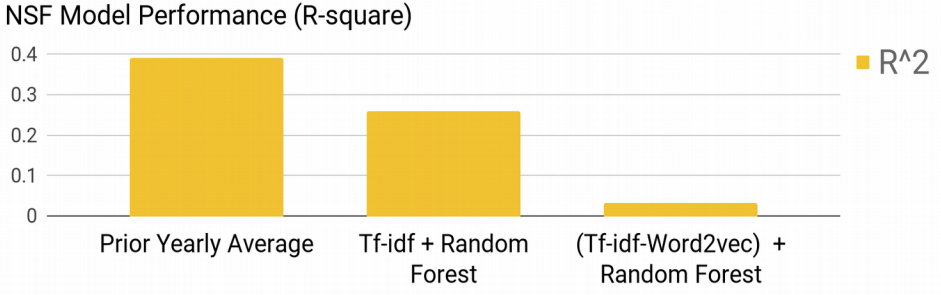
Modeling

Model evaluation and validation:

- Stratification on award year for train/test split
- 5-fold cross-validation





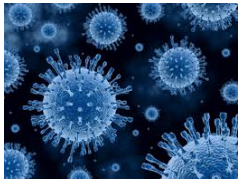


Lesson learned

- DoD and HHS tackles classical problems with new technique:

Recurring terms are “ship”, “vehicle”, “sensor”,...

Recurring terms are “vaccine”, “patient care”, “virus”,...



- NSF has less data and very diverse topics, e.g.:

“Mass-Production of 3D Printed Parts”
“Online Game to Assess Behavioral/Social Emotional Skills for Students in Kindergarten”
“Internally Microstructured Optical Films for Natural Lighting of Building Interiors”
“Machine Assisted Comparative Policy Analysis in Public Health”

Demo

About me - Cedric Herman

A couple of Masters…
in different countries

ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

Electrical Engineering/
Computer Science

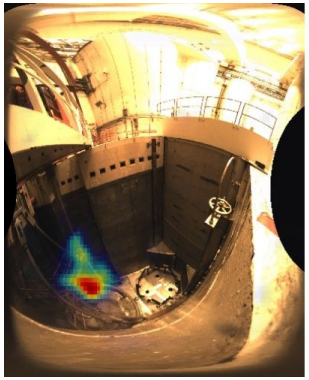
UNICAEN

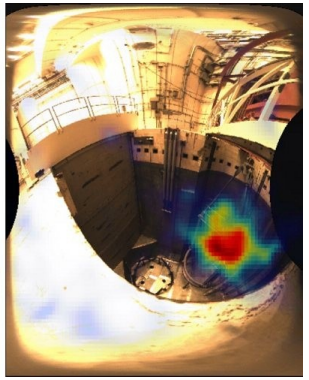
Physics

UNIVERSITY OF MICHIGAN

Nuclear Sciences

- 3 years, Research Assistant in Academia
- Gamma-ray Imager prototype delivered to DTRA





- 7 years, Research Scientist
- Global R&D in Radiation Detection Systems:
  - Modeling
  - Product development
  - Customer Support

App screen shots

Inbox (3) - herman... | Google Calendar - T... | Demo Group: Dia&... | Week 4 Demo Slide... | Cedric_Herman_We... | Home Page | chrome reload with... | cedric

127.0.0.1:5000

Instagrant Home About Contact

SBIR · STTR
America's Seed Fund™
POWERED BY SBA

Please choose Federal Agency below

Department of Education (ED)

Department of Commerce (DoC)

Department of Energy (DoE)

Department of Health and Human Services (HHS)

Department of Transportation (DOT)

Department of Homeland Security (DHS)

Department of Defense (DoD)

Enviromental Protection Agency (EPA)

National and Aeronautics and Space Administration (NASA)

National Science Foundation (NSF)

Department of Agriculture (USDA)

Inbox (4) - herman... | Google Calendar - T... | Demo Group: Dia&... | Week 4 Demo Slide... | Cedric_Herman_We... | Home Page | Wound healing pro...

cedric

127.0.0.1:5000/input?type=hhs

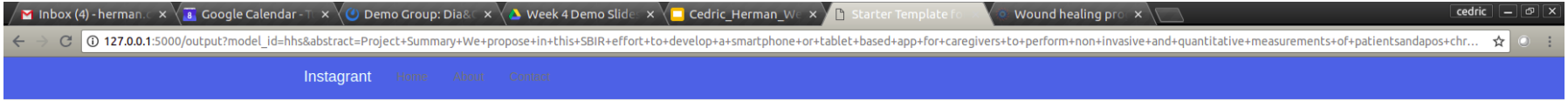
Instagrant Home About Contact

Department of Health and Human Services

Enter summary here:

Project Summary We propose in this SBIR effort to develop a smartphone or tablet based app for caregivers to perform non invasive and quantitative measurements of patientsandapos chronic wounds especially on diabtetic foot ulcers either at care facilities or at patientsandapos home Since chronic wounds take months or even longer to heal it is highly likely that more and more patients will be discharged early from the hospitals To follow up with the wound healing progress outpatients must routinely visit clinicians for a long period of time in order to have their wound healing progress assessed In addition current wound assessment methods mostly provide information with poor accuracy and are invasive The lack of precise wound data makes it difficult for clinicians to track subtle wound changes thus hindering the correct assessment of the treatment effectiveness The proposed software technology will be the first reported to offer precise wound measurement as well as analysis capabilities on mobile platforms It will provide great benefit to chronic wound patients by providing a low cost effective and personalized care solution With its unique framework and many advantages over existing methods the proprietary technology can be easily applied to the benefit of diagnosis and treatment on other

Estimate $$



Result Page

Estimated dollar amount:

Estimated dollar amount: $ 100,697
You are in the 17th percentile

Backup slides

Cost proposal

COST ELEMENT		Year 1
DIRECT LABOR:	Rate	Hours Amt.
Labor Category		
(Title and Name-- use additional pages as necessary)		
DIRECT LABOR COST		$110,000
MATERIAL COST		$50,000
TRAVEL COST		$ 3,000
OTHER (Specify)		$ 0
GRAND TOTAL ESTIMATED COST (PLUS FIXED FEE)		$ 153,000

Cost proposal

COST ELEMENT		Year 1
DIRECT LABOR:	Rate	Hours Amt.
Labor Category		
(Title and Name-- use additional pages as necessary)		$ 120,000
DIRECT LABOR COST		$ ~~110,000~~
MATERIAL COST		$ ~~50,000~~ $ 74,000
TRAVEL COST		$ 3,000
OTHER (Specify)		$ 0
GRAND TOTAL ESTIMATED COST (PLUS FIXED FEE)		$ 187,000 $ ~~153,000~~

Term importance

Department of Defense

abstract	0.0567
ii	0.0075
navy	0.0073
available	0.0073
polymeric	0.0069
high	0.0058
phase	0.0056
technology	0.0055
develop	0.0052
internet	0.0048
based	0.0046
treated	0.0044
application	0.0044
organization	0.0042
development	0.0041
design	0.0041
proposed	0.0041
benefit	0.0037
using	0.0036
cost	0.0035

Department of Human Health and Services

vaccine	0.0313
health relevance	0.0236
humanized	0.0195
patient specific	0.0122
patient care	0.0089
category	0.0076
future	0.0071
develop commercialize	0.0065
overall objective	0.0055
approval	0.0054
ind	0.0052
office	0.0048
anthrax	0.0047
panel	0.0045
objective	0.0045
fast track	0.0045
collaborator	0.0045
based	0.0045
virus	0.0044
threat	0.0044

Word2vec

Google News Pre-trained model:

- 3 million words vocabulary
- 300 dimensional vectors
- Skip-gram (as opposed to cbow)
- Made in 2013

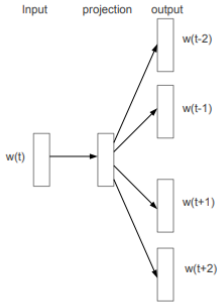
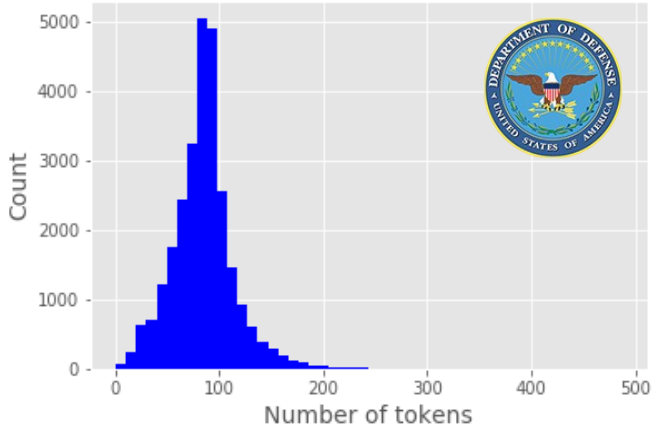
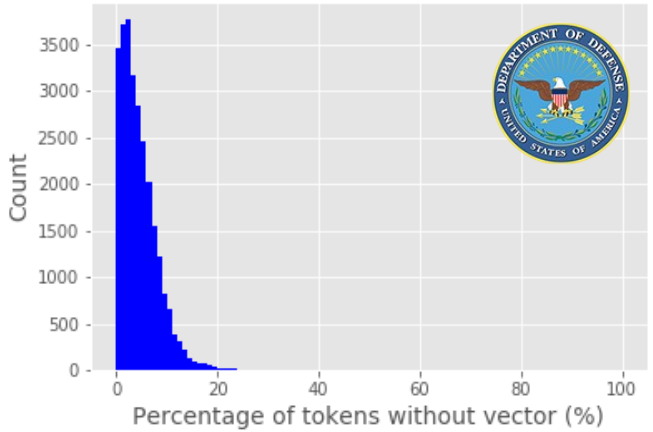


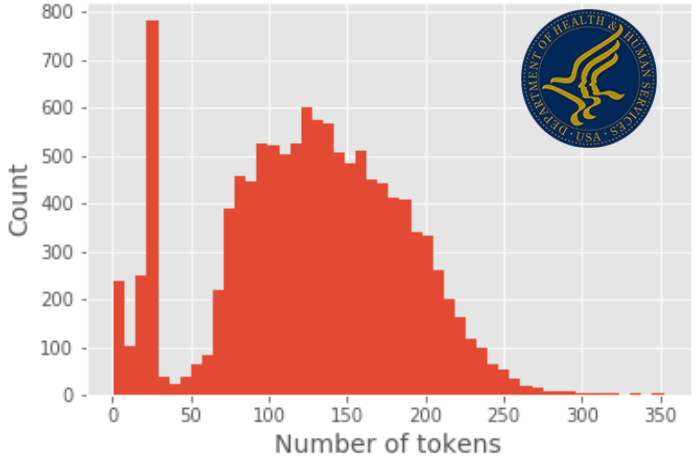
Figure 1: The Skip-gram model architecture. The training objective is to learn word vector representations that are good at predicting the nearby words.

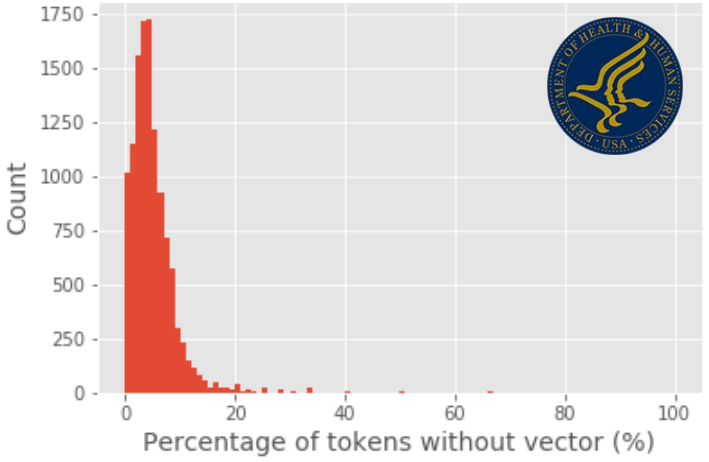
Missing word vector	Department of Defense	Department of Human Health and Services
Joined words	efficientbattery, volatilesolvents	Resultin, fdacompliance
Misspelling	capbilities, integrtion, trhoughput	Reproducibil, correcty, projectd
Acronyms	hgu, vtv, mlnn, rbf, sja	ekg, rna, pcr
Technical Jargon	themoluminescence, sublattice	Hypoadrenalism, cytolysis
Average missing tokens	5-6 tokens	4-5 tokens

Word2vec


Count
Number of tokens


Count
Percentage of tokens without vector (%)


Count
Number of tokens


Count
Percentage of tokens without vector (%)

Dispersion Index

- Aka Variance-to-mean ratio, relative variance, Fano factor:
  -> Measure the dispersion of observed occurrences

$$D = \frac{\sigma^2}{\mu}.$$

Tf-idf

$$tfidf(w,D) = tf(w,D) \times idf(w,D) = tf(w,D) \times log\left(\frac{C}{df(w)}\right)$$

Term frequency:

Normalize term count for each document by number of terms in that document

-> Adjust for documents length (number of terms varies)

Inverse Document Frequency:

Total number of document/number of documents with term w

-> Adjust for term importance across documents

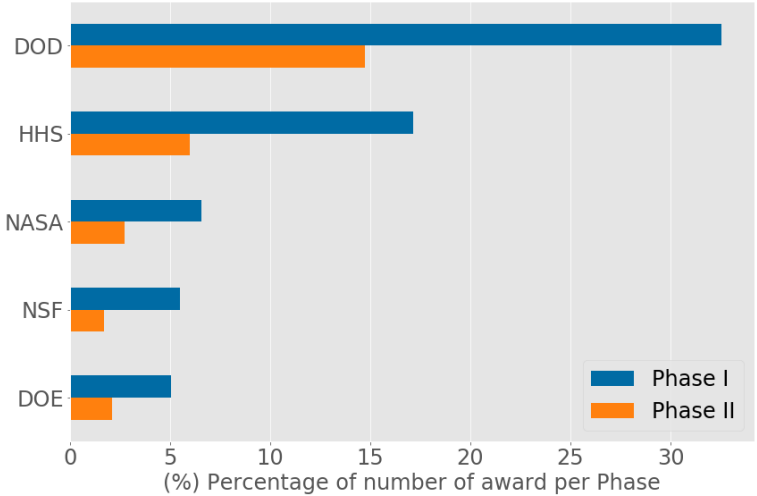
Tf-idf

tfidf(w, D) is the TF-IDF score for word ***w*** in document ***D***. The term ***tf(w, D)*** represents the term frequency of the word ***w*** in document ***D***, which can be obtained from the Bag of Words model. The term ***idf(w, D)*** is the inverse document frequency for the term ***w***, which can be computed as the log transform of the total number of documents in the corpus ***C*** divided by the document frequency of the word ***w***, which is basically the frequency of documents in the corpus where the word ***w*** occurs. There are multiple variants of this model but they all end up giving quite similar results.

SBIR-STTR grant

11 participating Federal Agency in 2017

Top 5 Federal Agency provider
1983 - 2018



SBIR-STTR grant

11 participating Federal Agency in 2017