Classifying Federal Contracting Notices

Arnob L. Alam

Overview

❖ The Federal Government posts more than 1,200 notices a day relating to contracting opportunities

❖ The notices are of several types

➢ Sources Sought: the government is seeking a list of suppliers who can provide particular goods or services

➢ Presolicitation: the government intends to solicit bids for goods or services in the future

➢ Solicitation: the government is actively accepting bids for goods or services

➢ Award: the government has awarded contract(s)

➢ Sole Source/Justification & Approval: the government has awarded a contract without competitive bidding

➢ Cancellation: cancellation of a previous presolicitation or solicitation

➢ Modification/Amendment: modification or amendment to an existing notice

➢ Special: Events, seminars, change of acquisition procedures or other special notices

➢ Other: (Intent to Bundle Requirements, Sale of Surplus Property, Foreign Government Standard, Fair Opportunity/Limited Sources Justification)

1560-015919298, PLATE,STRUCTURAL,AI

Solicitation Number: SPE4A516R0003
Agency: Defense Logistics Agency
Office: DLA Acquisition Locations
Location: DLA Aviation - BSM

Notice Type:
Presolicitation

Posted Date:
September 9, 2015

Response Date:
Oct 26, 2015 3:00 pm Eastern

Archiving Policy:
Automatic, 15 days after response date

Archive Date:
November 10, 2015

Original Set Aside:
N/A

Set Aside:
N/A

Classification Code:
15 -- Aircraft & airframe structural components

NAICS Code:
336 -- Transportation Equipment Manufacturing/336413 -- Other Aircraft Parts and Auxiliary Equipment Manufacturing

Synopsis:
Added: Sep 09, 2015 11:34 am
This solicitation will contain NSN 1560-015919298 for a PLATE,STRUCTURAL,AI. A Letter Request for Proposal (RFP) will be issued for award of a firm fixed price contract. This is a sole sourced item and must be acquired from an approved manufacturer. The OEM and sole approved source is Boeing, CAGE 82918, P/N: 5-83073-49. The quantity is as follows: 4 each. FOB: Origin; Inspection/Acceptance: Origin based on the Basic Order Agreement. Delivery: 182 DARO. The final contract award decision may be based upon a combination of price, past performance, and other evaluation factors as described in the solicitation. The solicitation issue date is on or about September 24, 2015; In accordance with 10 U. S. C. 2304 (c) (1), other than full and open competition applies. The proposed contract action is for supplies or services for which the Government intends to solicit and negotiate with only one source under the authority of FAR 6.302. Interested persons may identify their interest and capability to respond to the requirement or submit proposals. This notice of intent is not a request for competitive proposals. However, all proposals received within thirty days after date of publication of this synopsis will be considered by the Government. A determination by the Government not to compete with this proposed contract based upon responses to this notice is solely within the discretion of the Government. Information received will normally be considered solely for the purpose of determining whether to conduct a competitive procurement. Award will be made only if the offeror, the product/service or the manufacturer meets qualification requirement at time of award, in accordance with FAR clause 52.209-1 or 52.209-2. Please contact the Acquisition Specialist for additional information at Virginia.Bangit@dla.mil tel: (804) 279-2072, fax: (804)279-4949.

Contracting Office Address:
8000 Jefferson Davis Highway
Richmond, Virginia 23297-5000
United States

Place of Performance:
Defense Logistics Agency Richmond
8000 Jefferson Davis Highway
Richmond, Virginia 23297
United States

Primary Point of Contact:
Virginia Bangit
virginia.bangit@dla.mil
Phone: 8042792072

Secondary Point of Contact:
Brittany C. Gull,
Acquisition Specialist
Brittany.Gull@dla.mil
Phone: 8042794390
Fax: 8042796687

Opportunity History
- **Original Synopsis**
  Sep 09, 2015
  11:34 am

CANCELLATION OF RFQ

Solicitation Number: HTC711-15-Q-J-002
Agency: Other Defense Agencies
Office: United States Transportation Command
Location: USTRANSCOM Command Acquisition

Notice Type:
Special Notice

Posted Date:
August 27, 2015

Response Date:
-

Archiving Policy:
Automatic, on specified date

Archive Date:
September 3, 2015

Original Set Aside:
N/A

Set Aside:
N/A

Classification Code:
84 -- Clothing, individual equipment & insignia

NAICS Code:
315 -- Apparel Manufacturing/315990 -- Apparel Accessories and Other Apparel Manufacturing

Synopsis:
Added: Aug 27, 2015 1:18 pm
CANCELLATION NOTICE OF RFQ HTC711-15-Q-J002 BRAND NAME CRYE PRECISION PPE FOR JECC/JPASE.

THIS RFQ IS CANCELLED

IF YOU HAVE ANY QUESTIONS PLEASE FEEL FREE TO CONTACT THE CONTRACTING OFFICER (arlene.y.davis.civ@mail.mil)

Contracting Office Address:
508 Scott Drive
Bldg 1900
Scott AFB, Illinois 62225
United States

Primary Point of Contact.:
Arlene Y. Davis
arlene.y.davis.civ@mail.mil
Phone: 7578369889

Opportunity History
- **Original Synopsis**
  Aug 27, 2015
  1:18 pm

Annual Maintenance for Boeing Twister, Spatial Query Server, and Datamaster Software

Solicitation Number: 15T4943
Agency: Other Defense Agencies
Office: National Geospatial-Intelligence Agency
Location: OCSW- GEOINT/Enterprise Operations

Notice Type:
Combined Synopsis/Solicitation

Posted Date:
September 9, 2015

Response Date:
Sep 14, 2015 11:00 am Central

Archiving Policy:
Manual Archive

Archive Date:
-

Original Set Aside:
N/A

Set Aside:
N/A

Classification Code:
70 -- General purpose information technology equipment

NAICS Code:
511 -- Publishing Industries (except Internet)/511210 -- Software Publishers

Synopsis:
Added: Sep 09, 2015 11:16 am
National Geospatial Intelligence Agency
Request for Quote (RFQ)
The National Geospatial Intelligence Agency (NGA) has a requirement as detailed below:

This is a combined synopsis/solicitation for Maintenance on Boeing Software as detailed in the table below prepared in accordance with the format in FAR subpart 12.6 and FAR part 13, as supplemented with additional information included in this notice. This announcement constitutes the only solicitation; quotes are being requested and a written solicitation will not be issued. The North American Industry Classification System (NAICS) code for this acquisition is 511210 and the size standard is $38.5M.

Part Number Software Title Qty PoP
TDF-10N-LIC TWISTER DATA FRAMEWORK 1 7-1-2015 - 6-30-2016
TS-SQS-WP-S WORKPLACE NT 7 7-1-2015 - 6-30-2016
TS-SQS-WP-1-S SQS NAMED USER WORKPLACE 775 7-1-2015 - 6-30-2016
TS-SQS-E-1-S SQS NAMED USER ENTERPRISE 16 7-1-2015 - 6-30-2016
TS-SQS-E-FL-S SQS COCURRENT USER ENTERPRISE 50 7-1-2015 - 6-30-2016
TS-SQS-E-S SQS SERVER ENTERPRISE 123 7-1-2015 - 6-30-2016
TS-SQS-E-1-S SQS NAMED USER ENTERPRISE 369 7-1-2015 - 6-30-2016
TS-SQS-E-FL-S SQS CONCURRENT USER ENTERPRISE 534 7-1-2015 - 6-30-2016
TS-DATA-24 15+ DATAMASTER WORKSTATION MAINTENANCE - WINDOWS 81 4-1-2015 - 3-31-2016
TS-DATA-DEVCL DATAMASTER COMMAND LINE MAINTENANCE - UNIX 4 4-1-2015 - 3-31-2016
TS-DATA-S01 DATAMASTER WORKSTATION MAINTENANCE - UNIX 13 4-1-2015 - 3-31-2016
TS-DATA-14 5-14 DATAMASTER WORKSTATION MAINTENANCE - WINDOWS 9 4-1-2015 - 3-31-2016
TS-DATA-24 REINSTATEMENT OF 15+ DATAMASTER WORKSTATION MAINTENANCE - WINDOWS 127 1-1-2015 - 6-30-2015
TS-DATA-24 15+ DATAMASTER WORKSTATION MAINTENANCE - WINDOWS 127 7-1-2015 - 6-30-2016

Period of performance: 1 Jan 2015 - 30 Jun 2016.

This is a request for quote for the items listed above. NGA intends to select awardee on a brand name basis for the required items. You must be the manufacturer or authorized reseller of the Boeing Software listed above in order to respond to this RFQ. This notice is a request for competitive proposals. The Government reserves the right to cancel this solicitation, with no obligation to the vendor by the Government.

Maintenance shall include bug fixes, patches, and technical support.

If responding as a large business for NAICS code 511210, please submit a copy of your subcontracting plan or a letter stating no small business subcontracting possibilities exist for this requirement.
If any interested party is capable of fulfilling all of NGA's requirements, please explain in a two-page capability statement submitted to Doug Verseman at Douglas.H.Verseman@nga.mil by 11 SEP 2015 11:00 am Central Time.

The challenge

❖ We would like to classify a project (a collection of notices relating to a single acquisition effort) into one of several states

➢ Presolicitation/Sources Sought
➢ Open for bidding
➢ Awarded
➢ Cancelled

❖ First attempt: Just look at the notice type, and classify accordingly

➢ Issue: notice types are not being used correctly

1560-015919298, PLATE,STRUCTURAL,AI
Solicitation Number: SPE4A516R0003
Agency: Defense Logistics Agency
Office: DLA Acquisition Locations
Location: DLA Aviation - BSM

Notice Type:
<mark>Presolicitation</mark>

Posted Date:
September 9, 2015

Response Date:
Oct 26, 2015 3:00 pm Eastern

Archiving Policy:
Automatic, 15 days after response date

Archive Date:
November 10, 2015

Original Set Aside:
N/A

Set Aside:
N/A

Classification Code:
15 -- Aircraft & airframe structural components

NAICS Code:
336 -- Transportation Equipment Manufacturing/336413 -- Other Aircraft Parts and Auxiliary Equipment Manufacturing

Synopsis:
Added: Sep 09, 2015 11:34 am
This solicitation will contain NSN 1560-015919298 for a PLATE,STRUCTURAL,AI. A Letter Request for Proposal (RFP) will be issued for award of a firm fixed price <mark>contract. This is a sole sourced item</mark> and must be acquired from an approved manufacturer. The OEM and <mark>sole approved source is Boeing,</mark> CAGE 82918, P/N: 5-83073-49. The quantity is as follows: 4 each. FOB: Origin; Inspection/Acceptance: Origin based on the Basic Order Agreement. Delivery: 182 DARO. The final contract award decision may be based upon a combination of price, past performance, and other evaluation factors as described in the solicitation. The solicitation issue date is on or about September 24, 2015; In accordance with 10 U. S. C. 2304 (c) (1), other than full and open competition applies. The proposed contract action is for supplies or services for which the Government intends to solicit and negotiate with only one source under the authority of FAR 6.302. Interested persons may identify their interest and capability to respond to the requirement or submit proposals. This notice of intent is not a request for competitive proposals. However, all proposals received within thirty days after date of publication of this synopsis will be considered by the Government. A determination by the Government not to compete with this proposed contract based upon responses to this notice is solely within the discretion of the Government. Information received will normally be considered solely for the purpose of determining whether to conduct a competitive procurement.
Award will be made only if the offeror, the product/service or the manufacturer meets qualification requirement at time of award, in accordance with FAR clause 52.209-1 or 52.209-2. Please contact the Acquisition Specialist for additional information at Virginia.Bangit@dla.mil tel: (804) 279-2072, fax: (804)279-4949.

Contracting Office Address:
8000 Jefferson Davis Highway
Richmond, Virginia 23297-5000
United States

Place of Performance:
Defense Logistics Agency Richmond
8000 Jefferson Davis Highway
Richmond, Virginia 23297
United States

Primary Point of Contact.:
Virginia Bangit
virginia.bangit@dla.mil
Phone: 8042792072

Secondary Point of Contact:
Brittany C. Gull,
Acquisition Specialist
Brittany.Gull@dla.mil
Phone: 8042794390
Fax: 8042796687

CANCELLATION OF RFQ
Solicitation Number: HTC711-15-Q-J-002
Agency: Other Defense Agencies
Office: United States Transportation Command
Location: USTRANSCOM Command Acquisition

Notice Type:
<mark>Special Notice</mark>

Posted Date:
August 27, 2015

Response Date:
-

Archiving Policy:
Automatic, on specified date

Archive Date:
September 3, 2015

Original Set Aside:
N/A

Set Aside:
N/A

Classification Code:
84 -- Clothing, individual equipment & insignia

NAICS Code:
315 -- Apparel Manufacturing/315990 -- Apparel Accessories and Other Apparel Manufacturing

Synopsis:
Added: Aug 27, 2015 1:18 pm
<mark>CANCELLATION NOTICE</mark> OF RFQ HTC711-15-Q-J002 BRAND NAME CRYE PRECISION PPE FOR JECC/JPASE.

THIS RFQ IS CANCELLED

IF YOU HAVE ANY QUESTIONS PLEASE FEEL FREE TO CONTACT THE CONTRACTING OFFICER (arlene.y.davis.civ@mail.mil)

Contracting Office Address:
508 Scott Drive
Bldg 1900
Scott AFB, Illinois 62225
United States

Primary Point of Contact.:
Arlene Y. Davis
arlene.y.davis.civ@mail.mil
Phone: 7578369889

Opportunity History

- **Original Synopsis**
  Aug 27, 2015
  1:18 pm

The challenge

❖ Second Attempt: hard coded rule

➢ If the notice title, synopsis or filenames of attachments contain certain keywords (e.g. "cancel", "FAR PART 6.3", "sole source", etc. classify accordingly)

➢ Issue: boilerplate text/legalese

8. "Notice to Offeror(s)/Supplier(s): Funds are not presently available for this effort. No award will be made under this solicitation until funds are available. The Government reserves the right to cancel this solicitation, either before or after the closing date. In the event the Government cancels this solicitation, the Government has no obligation to reimburse an offeror for any costs." If funding is not available by 30 Sep 15 the solicitation will be cancelled. An award prior to 30 Sep 15 would require the contractor to begin performance within 90 days of award date in order to meet Bona Fide Need.

From a presolicitation

All future information regarding this acquisition, including the solicitation and any amendments, will be distributed solely through the FedBizOpps website (www.fbo.gov). Copies of the solicitation document and its related documents, as appropriate, will be posted at this website. Interested parties are responsible for monitoring the FedBizOpps website to assure that they have the most up-to-date information regarding this acquisition. The Government will not reimburse

From another presolicitation (we were stemming words so "solely" would become sole)

Machine Learning Approach

❖ How does a human look at this notice?
  ➢ Keywords in the title are more important than keywords in the synopsis
  ➢ Cancellation/Award/J&A announcement related keywords usually appear at the start of the synopsis rather than at the end
  ➢ The overall length of the synopsis gives a clue about notice type (cancellations and justification and approval [J&A] notices are usually short)
  ➢ The number of attachments gives a clue about the notice type
  ➢ Synopses for award notices and J&A notices often include names of companies or dollar figures
  ➢ The number of notices gives a clue to the project status (many modification notices suggest that it is an ongoing acquisition effort, while few notices mean either cancellation or J&A)
  ➢ The time between notices also gives a clue to the project status

Machine Learning Approach

❖ Possible machine learning approaches (given relevant features)
   ➢ Logistic Regression
   ➢ * Discriminant Analysis (QDA, LDA)
   ➢ Decision Tree
   ➢ MaxEnt classifier
   ➢ ** Neural networks

* not suitable for a mix categorical and continuous predictors

** no familiarity with this type of model

Machine Learning Approach

❖ Take a simple random sample of 10,000 projects and manually classify them

❖ Hold back 2,500 projects for testing

❖ Hold back another 2,500 projects for cross-validation (model selection)

❖ Train data using the rest of the 5,000 points

Data Engineering

❖ Notices are collected from several sources and stored in a MongoDB "notices" collection
  ➢ We are only interested in notices from www.fbo.gov (about 10GB of notices)
  ➢ However collection also contains notices from www.usaspending.gov (300+ GB of notices)
  ➢ Initial MongoDB architecture slow to query/process due to large size of collection
  ➢ Data was not being backed up
❖ Notices from www.fbo.gov are collected by scraping the new notices list every 10 minutes
  ➢ MongoDB "tasks" collection contains a list of tasks to be run, along with details such as "last time run successfully", "how often to run", etc.
  ➢ A PHP (Laravel) script goes through items in this collection, and runs the tasks based on certain rules (e.g. run the fbo.gov scraper if (now() - lastRun > interval)
  ➢ The fbo.gov scraper scrapes the first x pages of the new notice listing page (usually 10), and sends a list of notices to a message queueing service
  ➢ Another Laravel script polls the queue for a list of notices. It then checks if the notice already exists in the collection. If not it scrapes the new notice (transforming the html into a MongoDB document)
  ➢ Process was already set up before I joined GovTribe. Might have been easier to just use a cron scheduled task
  ➢ Scraping left some ambiguities. For example, the "Set Aside Type" field for a notice could be "None", "N/A", the field is present but has no value, the field is not present, the field is present and has an actual value

Data Engineering

❖ Scraped data sometimes contained encoding error
  ➢ Perhaps document is created on MS Word on a Mac, transferred to a Windows machine and then uploaded online. The encoding is not utf-8
  ➢ Synopses and titles sometime contain HTML tags and entities, line breaks (many variants of <br> tags, \n, ^b, etc.)
❖ Scraping occurs using PHP, but PHP does not have good machine learning libraries. Prefer to code in R or Python
❖ Solutions
  ➢ MongoDB : move to a clustered, sharded infrastructure to allow fast querying and index building. Manage cluster and backup data onto 10gen's MMS service
  ➢ Encoding: use publicly available code libraries to try to detect and correct encoding, or toss out non-ASCII entities if this fails
  ➢ HTML entities/natural language: use NLTK and BeautifulSoup to remove HTML, tokenize by word, filter stopwords and stem
  ➢ PHP interaction: create a python program that listens over a socket and communicate with it via JSON

Building the data set

1. Create a web page where people can easily browse through random notices and classify them into one of the cases. Save the classification in the MongoDB document
   a. Three people went through the notice collection and classified 10,000 notices over a three day period
2. Extract features using python
   a. Length of notice
   b. Do certain keywords/phrases appear in the title (sole source, cancel, request for quotes, justify, etc.)?
   c. Do certain keywords appear in the filenames of attachments?
   d. Do certain keywords appear in the first 50 words of the synopsis?
   e. ...

The Models

The Logit Model $P(X) = \dfrac{e^{\beta_0 + \beta_1\, X_1 + \ldots + \beta_p\, X_p}}{1 + e^{\beta_0 + \beta_1\, X_1 + \ldots + \beta_p\, X_p}}$, where $P(X) := P\left(Y = y \mid X\right)$

    Fit using Maximum Likelihood

The Tree Model

    Simple Tree: For each of the predictors, find the cut point that minimizes the Gini impurity. Select the predictor and cut point that minimizes gini impurity overall. Repeat for each of the new nodes, until the number of leaves in each node is less than some value. Use k-fold cross validation and cost complexity pruning to prune back the tree.

The models

Bagged Tree: Take bootstrap samples of the data and fit trees to each data. Average the estimate over all the trees to come up with a prediction.

Random Forest: Also take a bootstrap of the predictors when building the trees.

Boosted Tree: Build an initial tree on the data. Fit another tree to the residuals. Update the initial model by adding a weighted version of the tree fitted on the residuals. Fit this new model on the residuals. Repeat updating the model and fitting on the residuals, stopping using the validation set approach.

Models

Maximum Entropy

let $E[f_j(x)] = \Sigma_j \ p(x)f_j(x)$ be the empirical expected count of feature j

let $\bar{E}[f_j(x)] = \Sigma_j \ \underline{p}(x)f_j(x)$ be the modeled count of feature j

We maximize the entropy $H=-\Sigma_x p(x)*log[p(x)]$ subject to the constraint that $\Sigma_x p(x)=1$ and $\bar{E}[f_j(x)] = E[f_j(x)]$ for all features

Model Selection

The validation set approach was used to pick out among the different models

The logistic regression model achieved a misclassification rate of 8% on the validation set. Both the random forest and maximum entropy models achieved a misclassification rate of less than 4%. The random forest model was chosen because of its interpretability.

Variable Importance Charts were obtained to understand important features in the model.

Fine Tuning

- ❖ The Random Forest model achieved a misclassification rate of < 6% on the test data
- ❖ However, we encountered some notices which were being classified from "canceled" --> "open for bidding" or "awarded" --> "open for bidding". In these cases, notices were either coming to the project classifier out of order
- ❖ In these cases we chose to override the model predictions. Once a project has been cancelled or awarded, we do not try to classify its state again

Deploying

❖ The models were fitted using python (scikit.learn package) (models were initially developed using R, but deploying with python was easier)
❖ The classification tree was pickled (serialized) and uploaded to a web server
❖ A wrapper python program was created that listens for a JSON representation of a notice on a socket and responds with the classification (in JSON format)
❖ As part of the PHP notice scraping/project creation/update process, new notices are sent to this python program over the socket. The program then predicts the classification for the project and sends it back to the PHP script

Conclusion

❖ By examining some of the "meta data" associated with a project/notice, we were able to build a smarter classification system

❖ While a naive approach based solely on notice type achieved a classification rate of ~70% and a rules based approach achieved a classification rate ~80%, the machine learning approach achieved a classification rate of 94% - 96% (all rates based on test set data)