

Technical Disclosure Commons

Defensive Publications Series

November 2019

CRITICAL REQUEST FOR PROPOSALS (RFPs) REQUIREMENT INSIGHT ENGINE (caRRIE)

HP INC

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

INC, HP, "CRITICAL REQUEST FOR PROPOSALS (RFPs) REQUIREMENT INSIGHT ENGINE (caRRIE)",
Technical Disclosure Commons, (November 14, 2019)
https://www.tdcommons.org/dpubs_series/2688



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Critical Request for Proposals (RFPs) Requirement Insight Engine (CaRRIE)

Abstract

According to projections from IDC, 80% of worldwide data will be unstructured by 2025. For many large companies including, its reached the critical mass already. Unstructured data, such as RFPs, creates a unique challenge for organizations wishing to extract insight and use it to make strategic decisions [1]. We created a stacked machine learning algorithm based on Natural Language Processing (NLP) to extract insights from different RFP documents, we named the algorithm CaRRIE for Critical RFPs Requirement Insight Engine. CaRRIE is able to analyze how similarly-matched an RFP is to different an organization service offerings thus identifying gaps in it's current portfolio of service. CaRRIE will help an organization understands whether its services would fit the request in the RFP documents hence they can align their current services to the request hence win more deals. CaRRIE was built using a classifier and if we input any services, we can see a score breakdown for each word and an overall average score for that services. Higher scores are produced when the services is closely similar to the RFP document and lower scores means that RFP document requires services that the organization doesn't have. Having equipped with these insights, an organization will understand the customer needs better.

Problem statement

Traditional unstructured data stores are composed of text documentation and other file types. The new wave of unstructured data is increasingly coming from sources outside of the organization, such as RFP documents. A RFP is a document that lists out all the requirements and needs of a project. Companies create a RFP for upcoming projects, as a form of proposal to potential service providers. These providers bid to win the contract, based on the requirements of the RFP. The illustration shown in Fig 1.

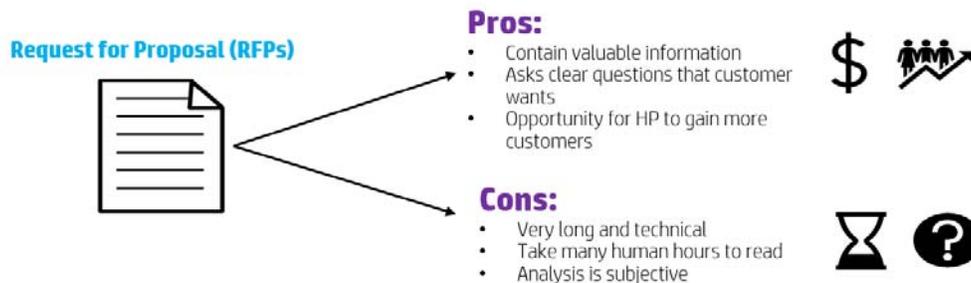


Fig 1. Request for Proposal Documents.

To be able to gain insight into these types of data, organizations will assign a person/a team to read and summarize it manually. Finally based on the summarization, an organization will find alignment with their current portfolios. Reading, generating insight and then responding to an RFP takes time and resources. These processes to generate insight will take a lot of time. The results of summarization are strongly subjective, depending on the experience and knowledge of the person assign to read the RFP.

Our solution

CaRRIE can do all the above fast and automatically. She saves times and provides more analysis insight with very less human interaction. CaRRIE can get more insight from RFP documents and quickly determine whether an organization can provide a viable service that meets customer requirements. In addition to that, she can provide insight to new area that an organization should focus on R&D strategic development in order to improve qualification on new client opportunities. The high level view of how CaRRIE works is shown in Fig 2.

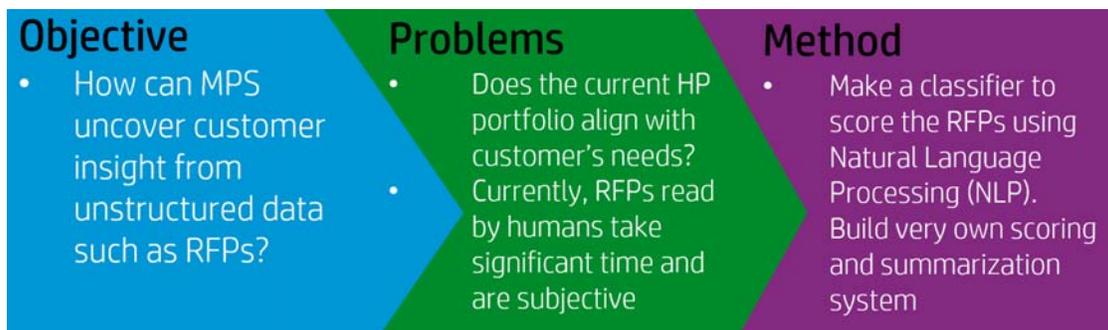


Fig 2. The CaRRIE Solutions

CaRRIE can also predict if an organization can win or lose the deal, using a classification algorithm. Using existing winning and losing RFP data, CaRRIE can be trained to understand why we win a deal or lost a deal. To be able to make CaRRIE classifier better at predicting future results, a lot of training data needed.



Fig 3. The CaRRIE algorithm flow.

In Fig 3. is shown how CaRRIE algorithm works. We start by taking the RFP documents and then ‘cleaned’ all the text data. A lot of these documents have irrelevant numbers, words that will mess up the scoring/classifier machine learning part. We also removed all extra white spaces, make everything lowercase. After we have considerably cleaned text data, we tokenized and stemmed each word in the document.

Tokenizing is the task of chopping the sentences into pieces, called tokens, at the same time throwing away certain characters, such as punctuation [2]. Here is an example of tokenization: input = “lend me your ears”, output = [lend, me, your, ears]. Stemming usually refers to a crude heuristic process that

chops off the ends of words and often includes the removal of derivational affixes[2], for example, input = “car, cars, car’s, cars’ “, output = “car”.

The next step, we apply tf-idf (term frequency and inverse document frequency) technique to create feature-vectors from each text document. This method is intended to reflect how important a word is to a document in a collection [3]. This technique is calculating frequency of each word in a document and inversely weighted with the frequency of the word appear in every document [3]. After each text transform into feature vectors, the algorithm will be able to learn from it and find patterns in the documents.

Using this tf-idf classifier we can input products and receive a score based on how the classifier is trained on the RFP. The score ranges from 0 to 1, and a score that is closer to 1 represents a better product fit for the customer request based on what they’ve said. This is extremely useful as one can input any products/services into the classifier and receive instantaneous feedback on if that product might fit what the customer is requesting from its bidders.

To build on top of the scoring system, we also wanted to add in an insight component to our algorithm. Not all products are going to score high, some are going to receive a low score. The question then becomes, how can we get further insight into things that an organization might need to work on to get more customers? Thus, we developed the summarization technique. The program divides up the RFP document into sections, then iteratively scores each section. After the scoring is complete the program collects all the sections that are high scoring and low scoring and creates an important summary for each category. This summarization summarizes the important parts of the document, so for the high scoring sections we can see what are the important parts that are scoring really well, and for the low scoring document we get insight into what areas we can improve on.

Results

We are going to split the result into three part, scoring, summarization and further exploration.

Scoring part

In Fig. 4, we shown CaRRIE scoring results for a couple services.

```
In [798]: get_scores(winning_rfp_0013)
Scores for product: Print Security Advisory Service
servic - 0.5750055535525305
secur - 0.6092837102637222
print - 0.5460237850048482
Average Score for Product: 0.5767710162737002
-----
Scores for product: Print Security Advisory Retainer Service
servic - 0.5750055535525305
secur - 0.6092837102637222
print - 0.5460237850048482
Average Score for Product: 0.5767710162737002
-----
Scores for product: Managed Environment Assessment
manag - 0.6246432054780712
environ - 0.7809102162906827
Average Score for Product: 0.702776750884377
-----
Scores for product: Eco Printing Assessment Environment
print - 0.6424915029716878
environ - 0.7662928086633606
Average Score for Product: 0.7043921558175241
-----
```

Fig 4. The CaRRIE scoring results.

CaRRIE will score between 0 to 1 for any products or services inputted to a RFP document. The output score is an average for each products or services. CaRRIE will also provide score for each individual word in the

product/service descriptions. Scores close to 1 means product/service is closely related to what customer is asking in RFP. Scores close to 0 means that product/services are never requested by the customer in their RFPs. Equipped with this information, an organization can configure which product/service could fit the customer’s needs in a matter of seconds.

Summarization part

In Fig 5, we show the approach to summarize a RFP documents. The basic idea is very similar to a person looking for a coin in a football field using a metal detector. He will use his detector to guide him to the coin by going through each part of the field. With the same principle, CaRRIE will use the scoring part as “the metal detector” for scanning the entire document section by section and score each section. Once we have the score, we rank them and then we summarize the highest and the lowest sections to get insight.

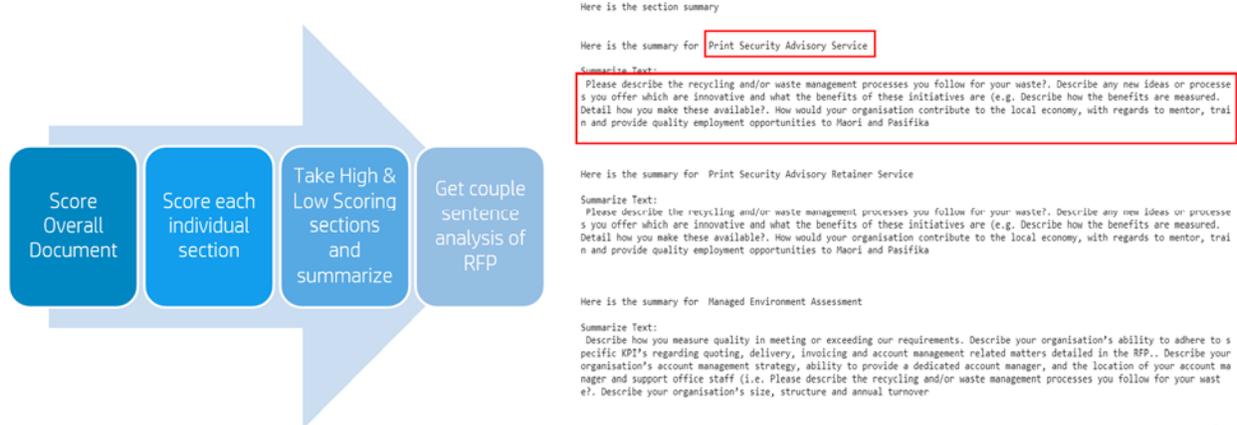


Fig 5. The CaRRIE scoring results.

As you can see In Fig 5, we took a couple services that the organization provides and use the above technique to get into the “right” section and then summarize it.

Further exploration

CaRRIE also capable of finding pattern within winning RFPs as well as in losing RFPs. This is crucial to truly understand what the main drive are behind the winning and the losing ones. Hence an organization can utilize these information to be used in the future.

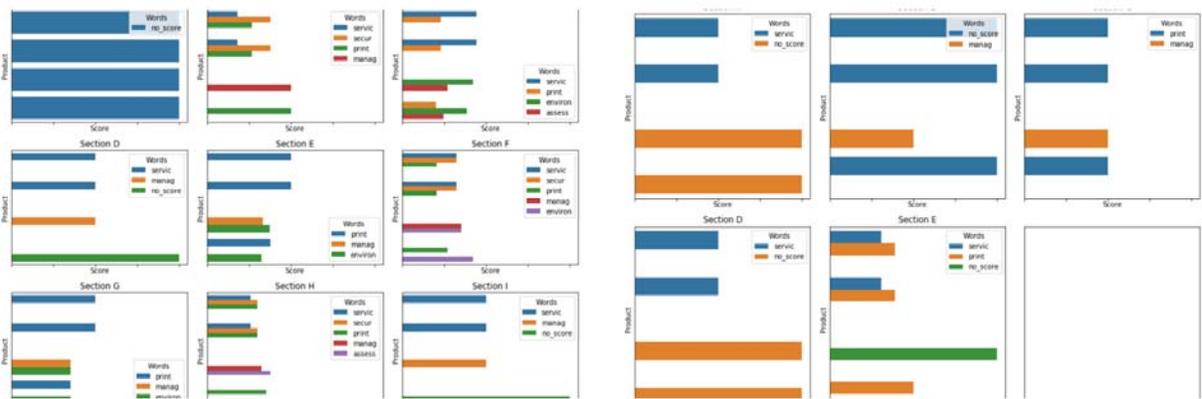


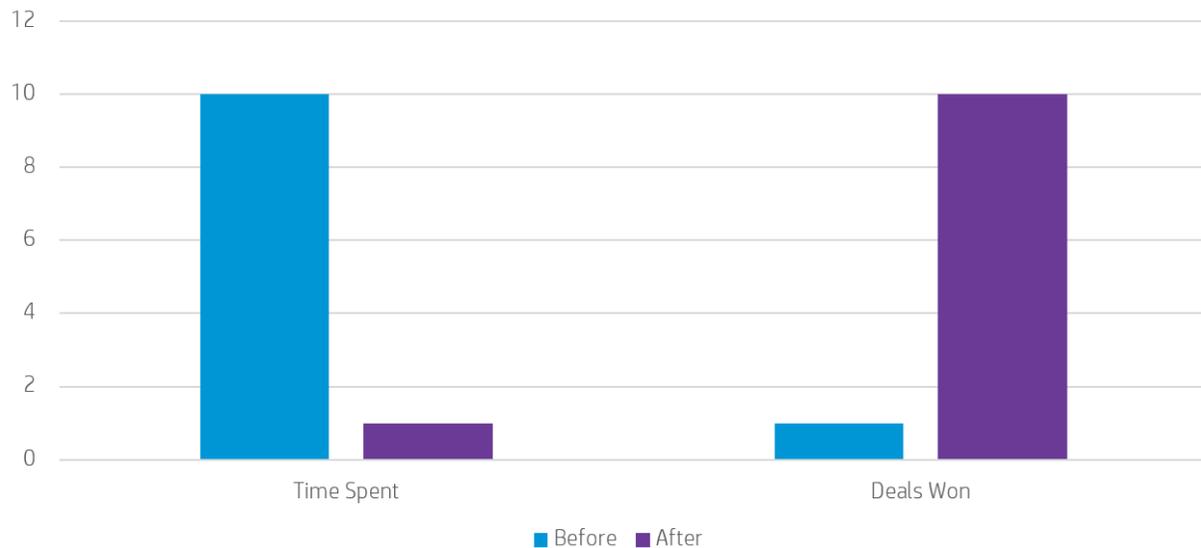
Fig 6. Left panel shows section by section scoring for winning document and the right panel shows section by section for losing document.

In Fig 6, we show how CaRRIE analyze each section in a winning document vs a losing documents. The scoring section by section show as very different sections pattern in winning vs losing documents. The winning RFP has

more scores (more colors) for different products/services. More scores mean higher correlation of RFP to an organization products/services. The losing RFP only scores on basic words/products, it is less colorful compare to winning ones.

Summary

In this section we want to summarize the advantages of CaRRIE. CaRRIE can reduce effort, time and costs by quickly capturing insight from a RFP document thus helping an organization to improve deal qualification, improve win rates and applying learnings to future service development projects. These are shown in Fig. 7 animation, it is not based on real data but the general idea is presented there.



*Graph contains no real data, just there to make a general point

Fig 7. The CaRRIE business value.

References

- [1] <https://solutionsreview.com/data-management/80-percent-of-your-data-will-be-unstructured-in-five-years/>
- [2] <https://nlp.stanford.edu/IR-book/html/htmledition/tokenization-1.html>
- [3] Rajaraman, A.; Ullman, J.D. (2011). *"Data Mining"* (PDF). *Mining of Massive Datasets*. pp. 1-17. [doi:10.1017/CB09781139058452.002](https://doi.org/10.1017/CB09781139058452.002). ISBN 978-1-139-05845-2.

Disclosed by Jai Bansal, Anthony Valadares and Anton Wiranata, HP Inc.