Supercharging Expert Networks

Designing a High-Performance Matching Engine to Leverage GLG's Network of Technology and Healthcare Experts to Wide Ranging Customer Requests

Project GLG | Team MAS

Mark

Aiman

Spencer

Motivation



GLG's Value Proposition

A Diverse Expert Network with 2 Core Competencies



Driving Value to GLG Clients



Reduce Client Response Time

Drive **customer satisfaction** by connecting with Experts quickly



Increase Matching Accuracy

Improve **customer retention** by getting the match right the first time.



Improve Client Retention

Deliver **higher quality matches** to drive customer satisfaction in as little time as possible



Increase Customer Lifetime Value

Customers that are more satisfied with their prior experiences with GLG are likely to stay longer and be repeat customers

Finding a Needle in a Haystack

Inbound: Client Requests

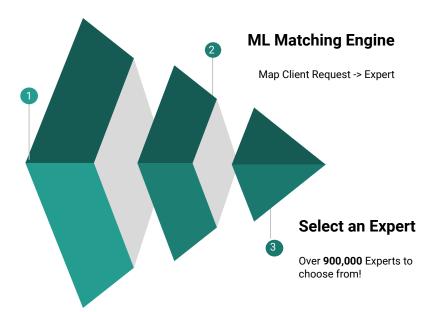
A <u>diverse</u> range of subjects to choose from!

"What are the unmet needs of global B2B customers in the organic materials segment?"

Japanese Industrial Conglomerate

"How large is the existing market for robotic process automation (RPA)? Is there room for it to grow? What does the typical RPA customer look like? Which players deliver the most ROI in this space?"

Hedge Fund



Visions: Towards A New Engine

Current Engine



Precise Matches

We combine the human insight of our team with an Al-driven matching platform to find the right people.

- Keyword Search
- Human Review
- Contact Client with Chosen Expert

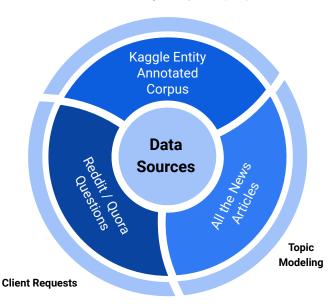
Proposed ML Engine

- Learn Topics from text
- Learn **Entity Tags** from text

- Topics + Entities
- Submit a Client Query
- Human Review
- Customer Facing Dashboard
- Serve Matched Topics + Entities

Training Data

Named-Entity-Recognition (NER)

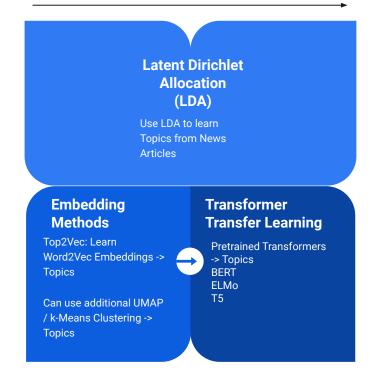


Increasing Complexity

Approaches to Explore

Training Data Named-Entity-Recognition (NER) Data All the News Sources Topic Modeling

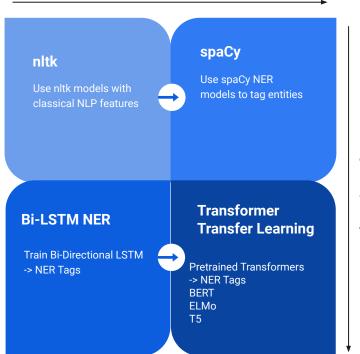
Increasing Complexity



Approaches to Explore

Training Data Named-Entity-Recognition (NER) Kaggle Entity **Annotaated** Corpus Data **Sources** Topic Modeling **Client Requests**

Increasing Complexity



Increasing Complexity

System Design



Customer

Enquirer in search of the right expert to address their request



Customer Query

Unstructured text requests (test data to be taken from news dataset and past case study examples from GLG's website)



Monitor classification, NER accuracy via (hand-labeled portion of) incoming requests

Improvement

Expert

AWS Flastic Beanstalk Instance

Streamlit / Flask Dashboard

Web App UI

Predicted Topics and Entities Displayed on Web Interface Display Dendrogram

3 REST APIs

2 Deployed ML Models

REST APIs to serve responses from each ML Model and to construct the Dendrogram over time

Dendrogram Pipeline

Continuously updating a visual, hierarchical representation of customer requests via dendrogram

AWS S3 Storage

Data Acquisition

Stream of incoming text queries

Data Pre-Processing

Tokenization Lemmatization Stopword Removal

Trained ML Models

Models with sufficient performance will be readied for deployment via API

2 Parallel Development **Pipelines**

Topic + NER Pipeline Dendrogram Pipeline

Topic Modeling Pipeline

- Categorize incoming requests according to trained Topic Model
- Identify and extract Named Entities as an aid to better request classification

Preliminary Mockup

Input Query Here

Try: "How large is the existing market for robotic process automation (RPA) in Japan?"

Or: "What are the unmet needs of global B2B customers in the organic materials segment?"

Topic / Area of Expertise:

Industrial Manufacturing

Named Entities in Query:

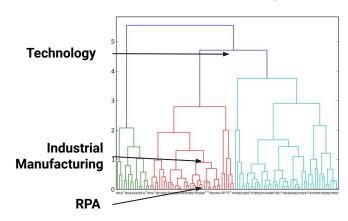
Japan - GPE (Geopolitical Entity - Country, City, States)

Top Experts:

Experience combining Japan and Industrial Manufacturing Technology

- Name 1, Contact Info
- Name 2, Contact Info
- Name 3, Contact Info

Hierarchical Dendrogram



Ethical Considerations

Data Privacy

- Tool will function using the minimum amount of customer data necessary to fulfill the task.
 No session metadata to be stored.
- Purpose and aggregate-level function of tool not designed to benefit from personal data
- User prompt to state that the request should contain "No PII (Personal, Name, Address, SSN or other financial information)".
- By using this tool, user consents to long-term retention of request for model-training purposes.

Bias

Primary sources of potential bias:

- Cognitive Bias in the labeling training dataset
- Linguistic Bias inherent in the training data due to the time period from which articles were written, formal styles of language present in news articles, etc.
- Representation Bias in the training data where news coverage may lean towards celebrities vs. normal citizens.
- Gender Imbalance where men may be featured more prominently over women, etc.

Timeline



Q+A