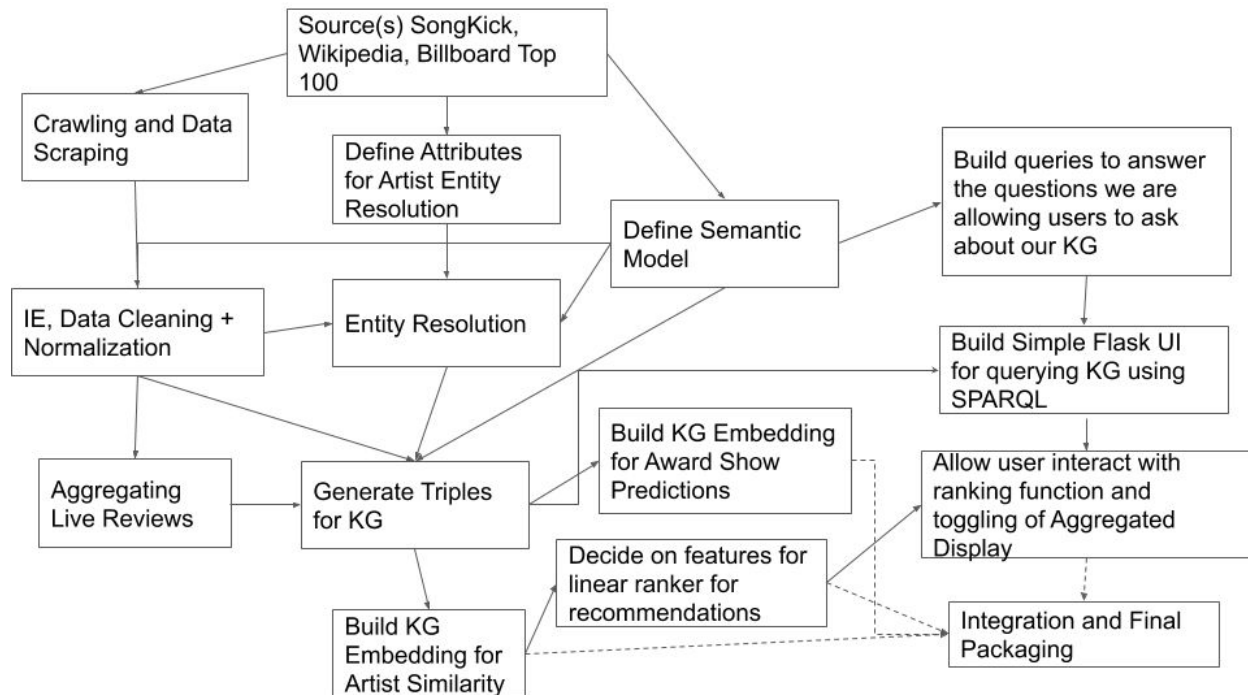


### Workflow



### How to know what done is (for certain tasks):

- 1) Entity Resolution: mark some records coming from different sources with a unique key and makes sure all records with the same key are matched by our resolution system
- 2) Build KG Embedding for Artist Similarity: We have similar artist information from Wiki and Songkick, with these labels we can:
  - a) compare distance amongst these pairs vs average distance
  - b) set thresholds on embedding distance, and see how many similar artist lie within the threshold
- 3) Build KG Embeddings for Award Show Prediction: We have historical data so can evaluate against predictions
- 4) Features for Ranker: We are taking the approach of allowing users to build their own linear rankers given a set of features. We will allow users to indicate whether recommendations received were good or bad and store the weight used for further analysis. We can learn what weight distribution creates good recommendations
- 5) Query Accuracy: Have sets of inputs and expected outputs for each query

### What have we done:

- 1) Built scraper for Songkick Artist data
- 2) Defined at least v1 of our Semantic Model

### Immediate Next Steps:

- 1) Build more scrapers and inspect output of scraping
- 2) Come up with concrete set of queries that we will allow people to ask and build query templates to answer these questions