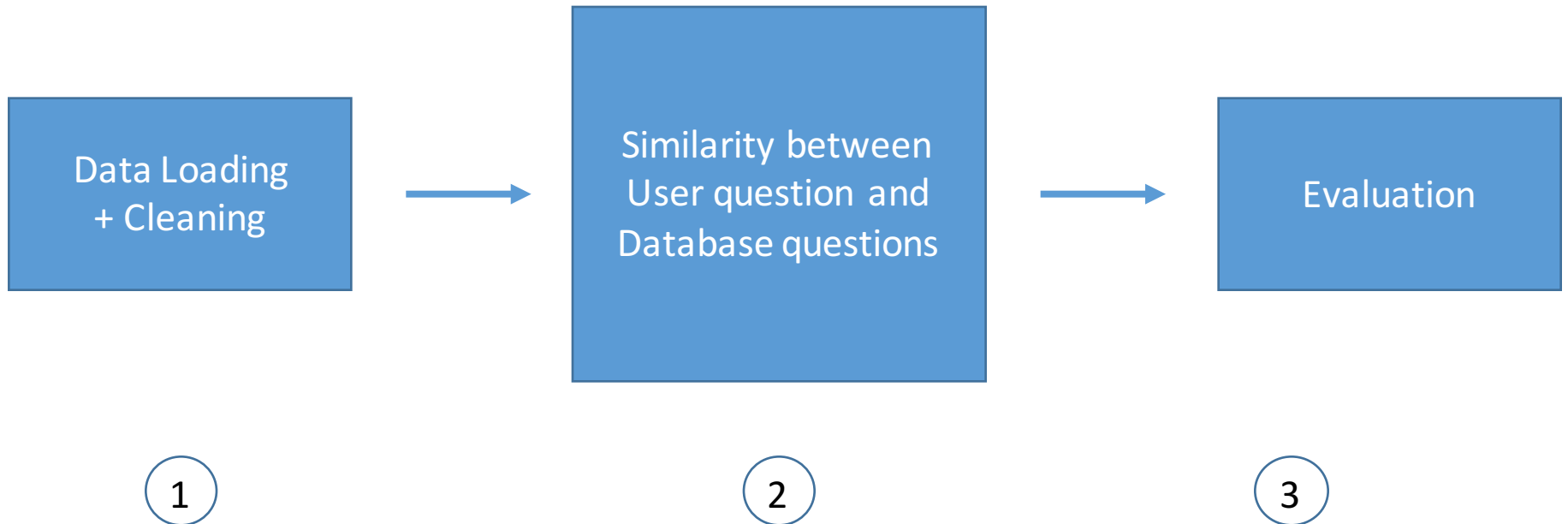# DATACHAT_

Urjit Patel
Vincent Chabot

Zelros (zelros.com) : chatbot for conversational BI

"What was the average sales amount per vendor last month"

Framework

Similarity #1

Continuous Bag of Words

What is the number of cars?

	What	<254...6457>
+	is	<083...2134>
+	the	<324...1564>
+	number	<123...3563>
+	of	<765...9776>
+	cars	<986...6457>

<353...5434>

What is the number of users??

	What	<254...6457>
+	is	<083...2134>
+	the	<324...1564>
+	number	<123...3563>
+	of	<765...9776>
+	user	<123...5432>

<353...3456>

Score = cosine similarity = 0.8

Similarity #2

Sentence2vec (Skip-thought)

What is the number of cars?

What is the number of users??

<254...6457>

<254...6754>

Score = cosine similarity = 0.75
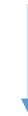
Similarity #3

Max words pairwise similarity

What is the number of cars?

Part Of Speech
Filtering

[(number, noun), (cars,noun)]

What is the number of users??

Part Of Speech
Filtering

[(number, noun), (user,noun)]

Score = max cos_sim(word1 in sentence 1, word in sentence 2)
    + max cos_sim(word1 in sentence 1, word in sentence 2)

Improvements

- Weighted average with inverse document frequencies

- Multiple choice when unsure

Current accuracy : 0.85

Model : Max word pairwise similarity + google w2v embeddings + POS filtering + idf + multiple choices