

A Wrong Answer or a Wrong Question?

An Intricate Relationship between Question Reformulation and Answer Selection in Conversational Question Answering

Svitlana Vakulenko, Shayne Longpre, Zhucheng Tu, Raviteja Anantha | Workshop on Search-Oriented Conversational AI (SCAI) 2020 | November 19, 2020

Project

Question Rewriting for Conversational QA

Take-away

Question rewrites are useful for QA model evaluation!

Task

Conversational Question Answering

Q: Where is Xi'an?

A: Shaanxi, China

Q: What is its GDP?

A: 95 Billion USD

Q: What is the share in the province GDP?

A: 41.8%

Task

Conversational Question Answering

Q: Where is Xi'an?

A: Shaanxi, China

Q: What is **its** GDP?

A: 95 Billion USD

Q: What is **the share in the province** GDP?

A: 41.8%

Anaphora

Ellipsis

Approach

Question Rewriting

Q: Where is Xi'an?

A: Shaanxi, China

Q: What is **its** GDP?

A: 95 Billion USD

Q: What is **the share in the province** GDP?

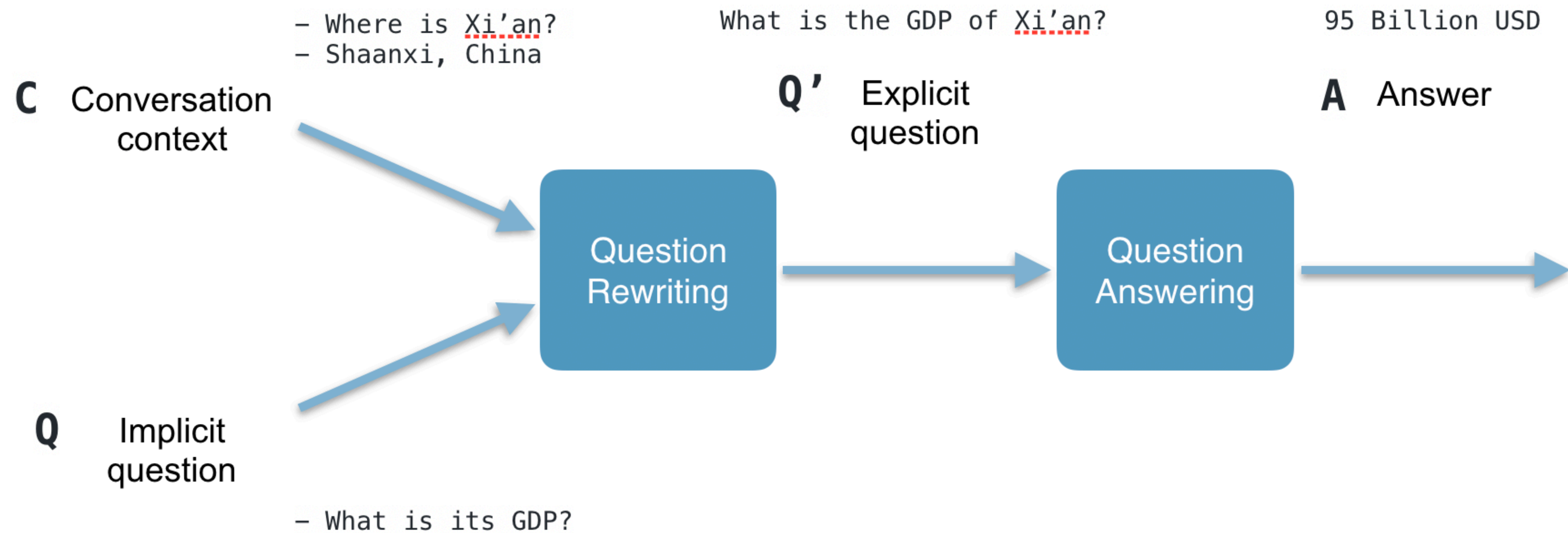
A: 41.8%

QR: What is **Xi'an's** GDP?

QR: What is **the share of Xi'an in the Shaanxi province** GDP?

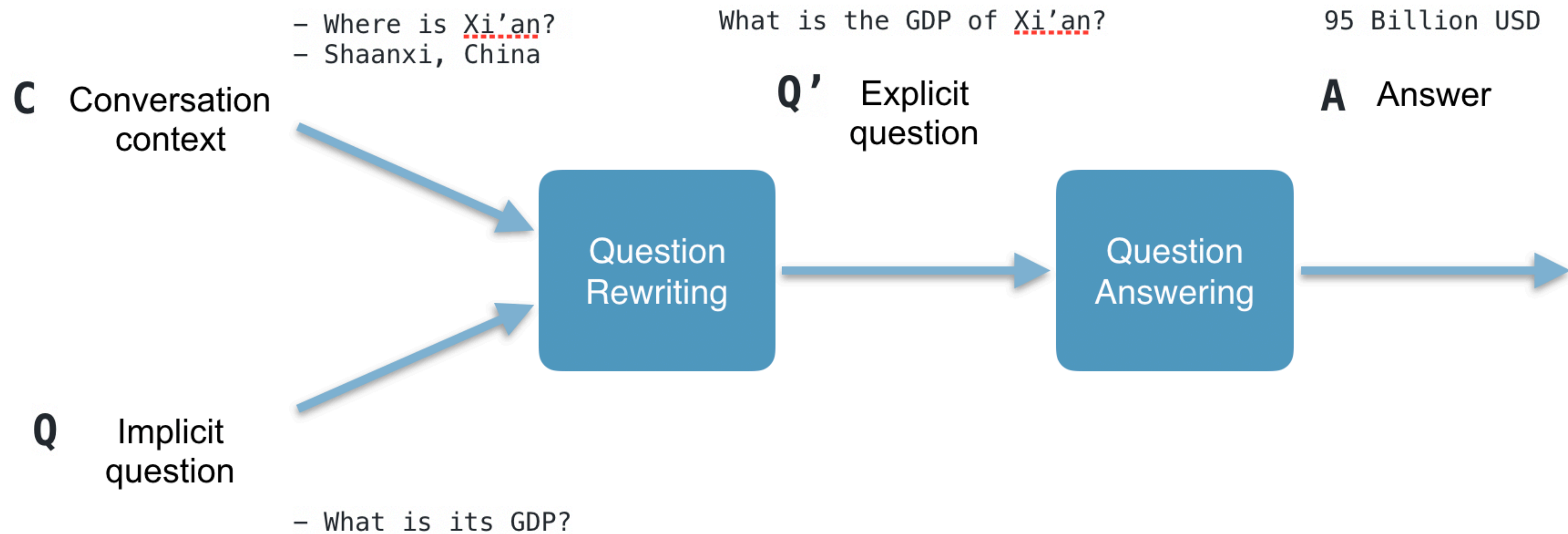
Approach

Question Rewriting



Evaluation

Question Rewriting



Error Analysis

Question Rewrites

Original: What is **its** GDP?

QR: What is **Xians** GDP?

Human: What is **Xi'an's** GDP?

Error Analysis

Question Rewrites

Original: What is **its** GDP?

A: 2 Million EUR



QR: What is **Xians** GDP?

A: 95 Billion USD



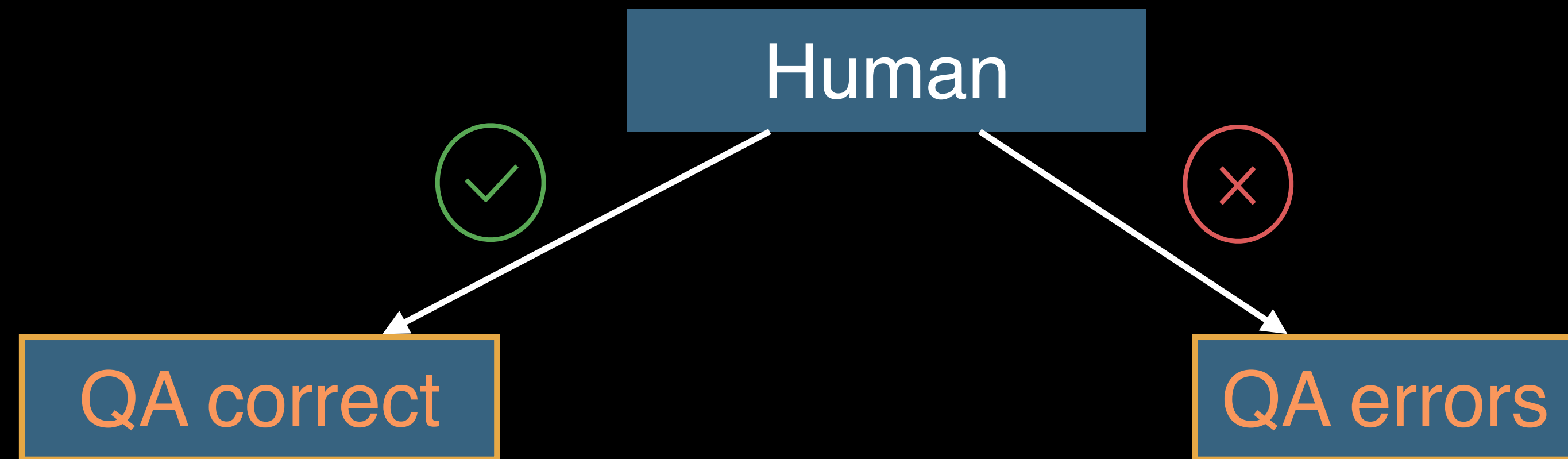
Human: What is **Xi'an's** GDP?

A: 95 Billion USD



Method

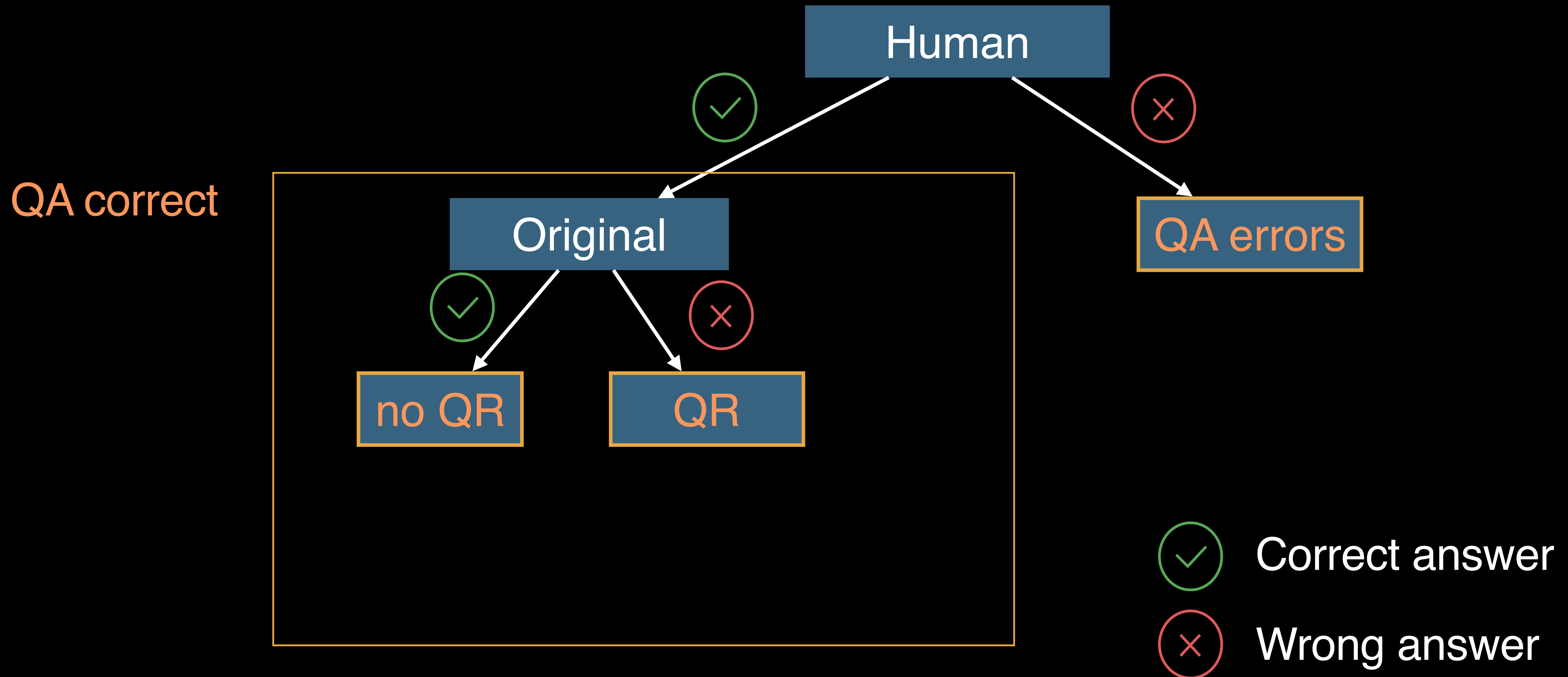
Error types hierarchy



- ✓ Correct answer
- ✗ Wrong answer

Method

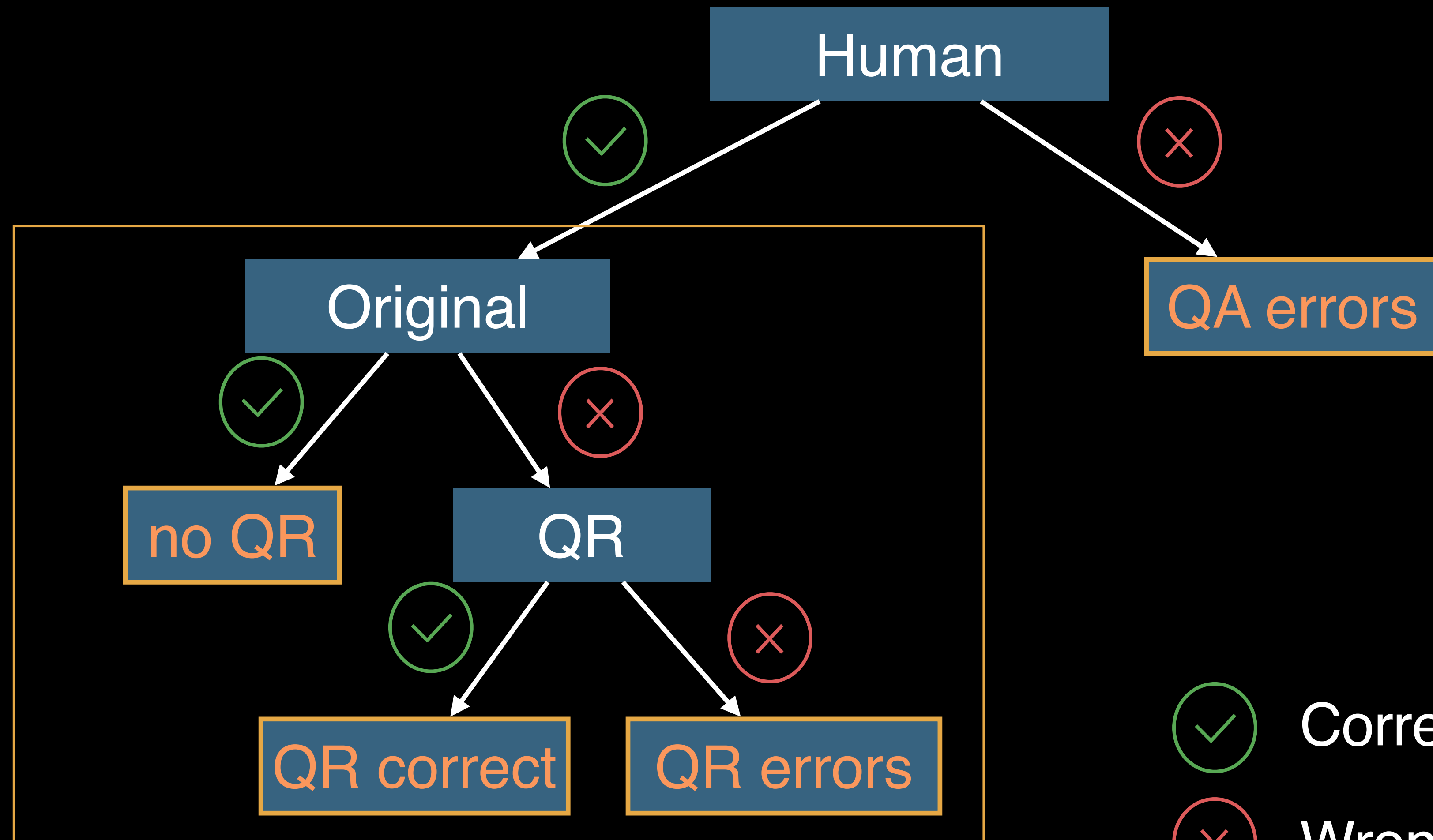
Error types hierarchy



Method

Error types hierarchy

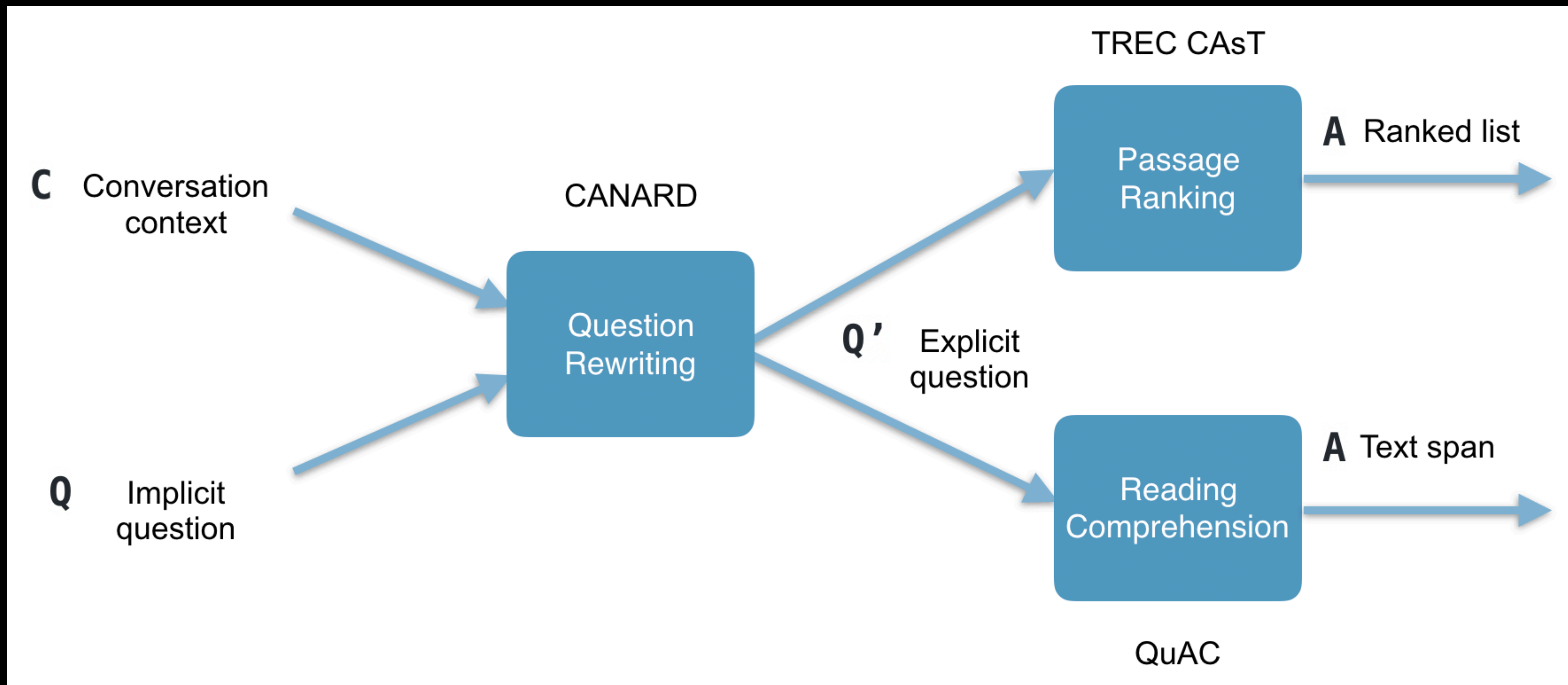
QA correct



✓ Correct answer
✗ Wrong answer

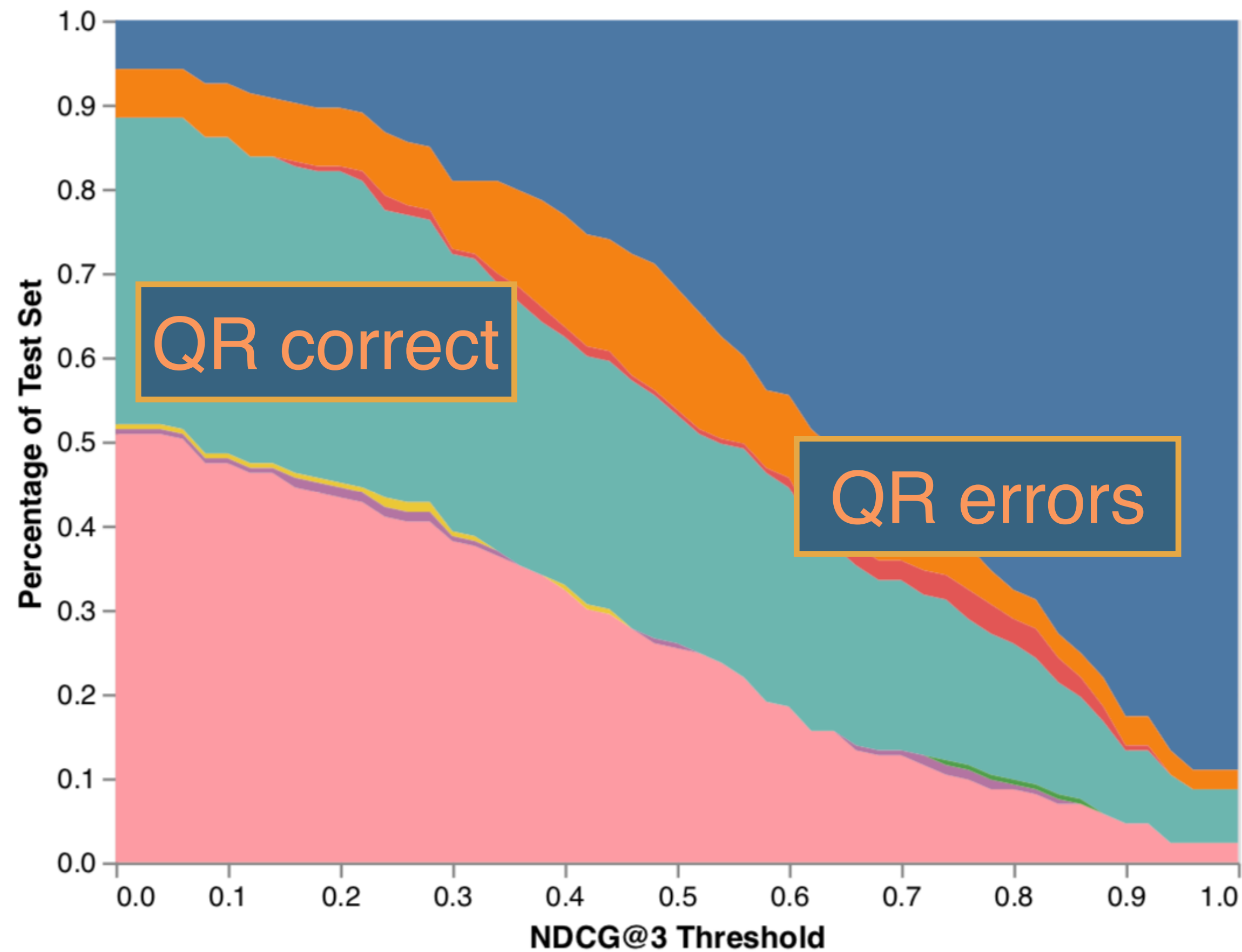
Experiments

Question Rewriting

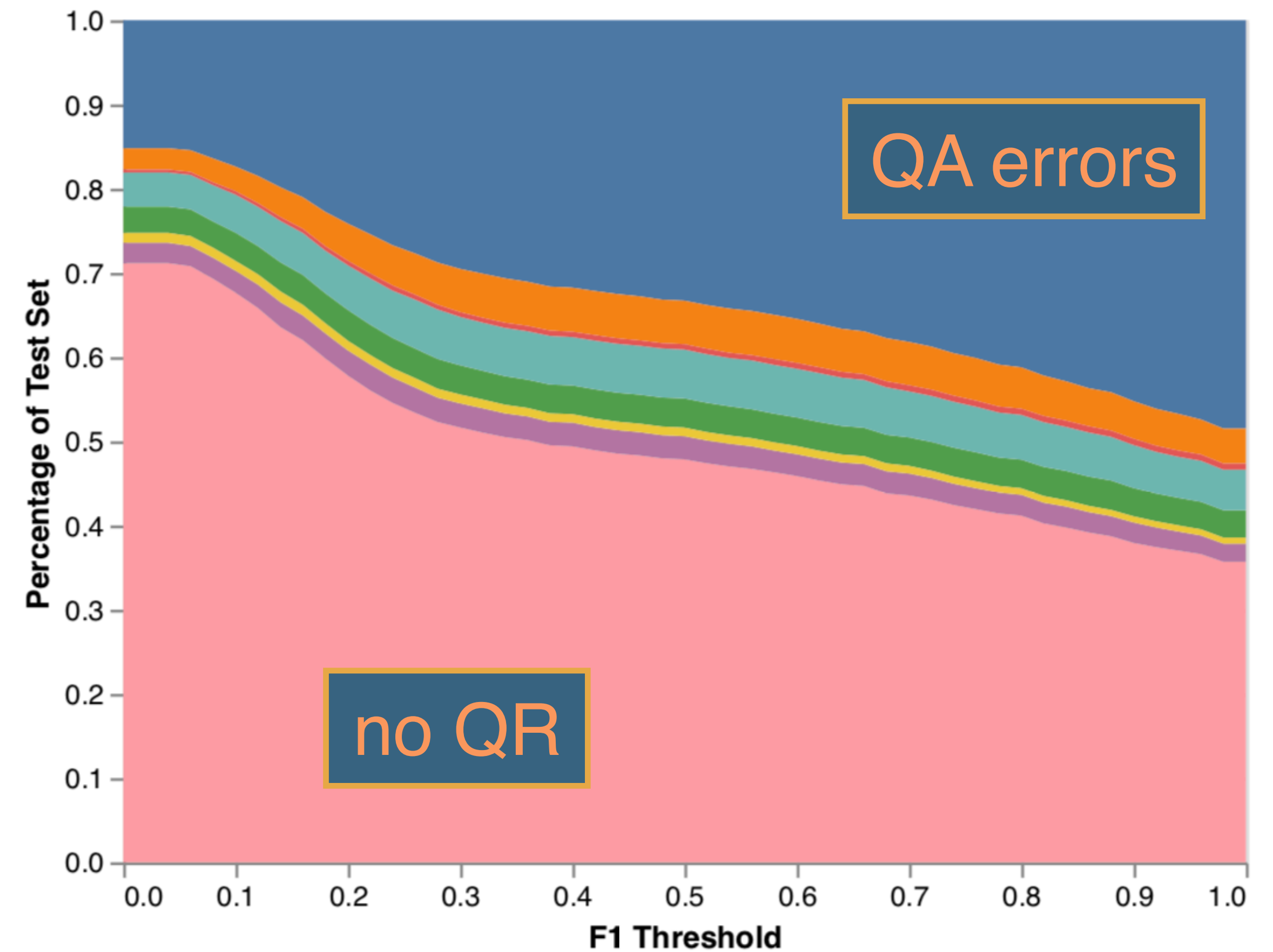


Results

Error comparison



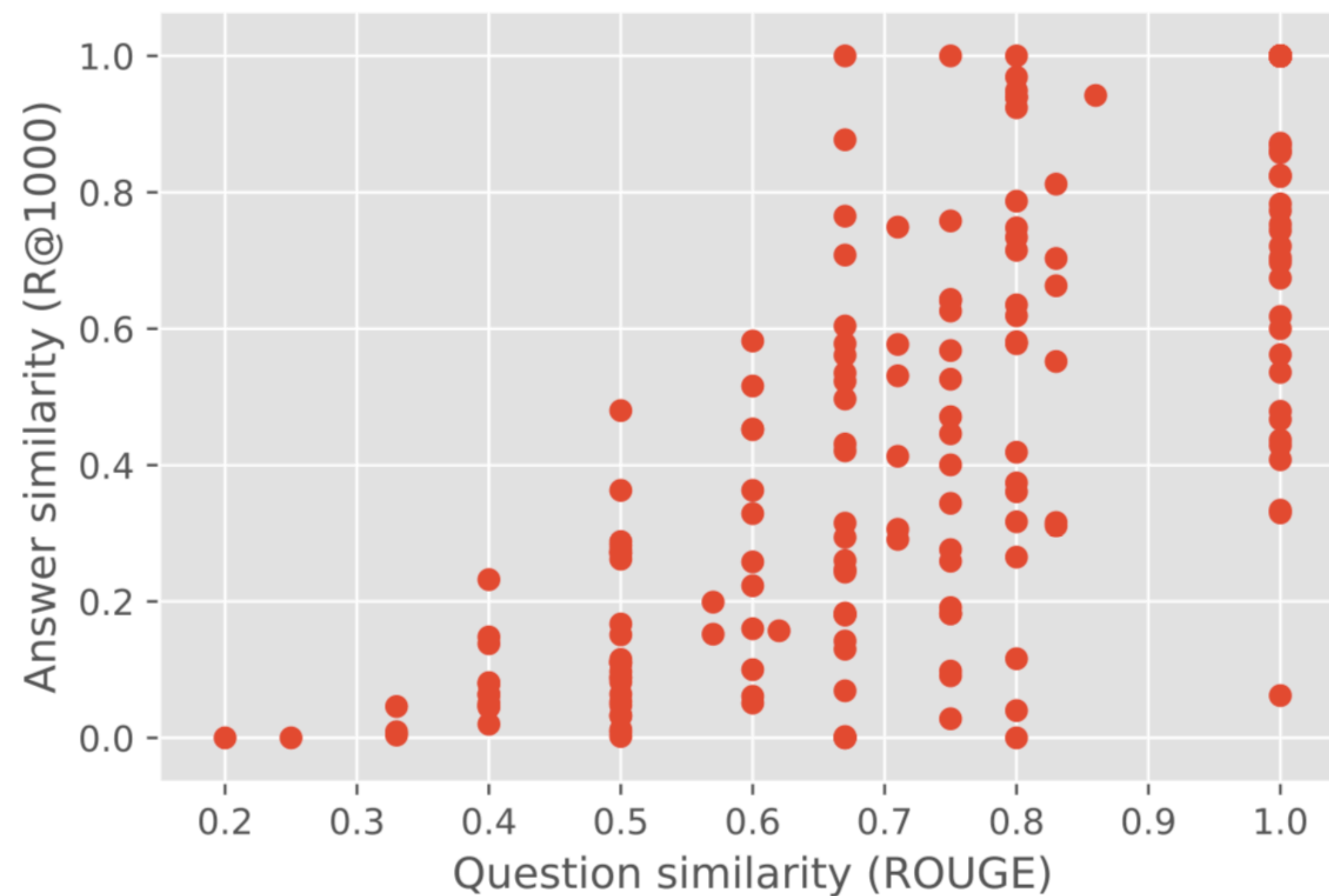
Passage Ranking



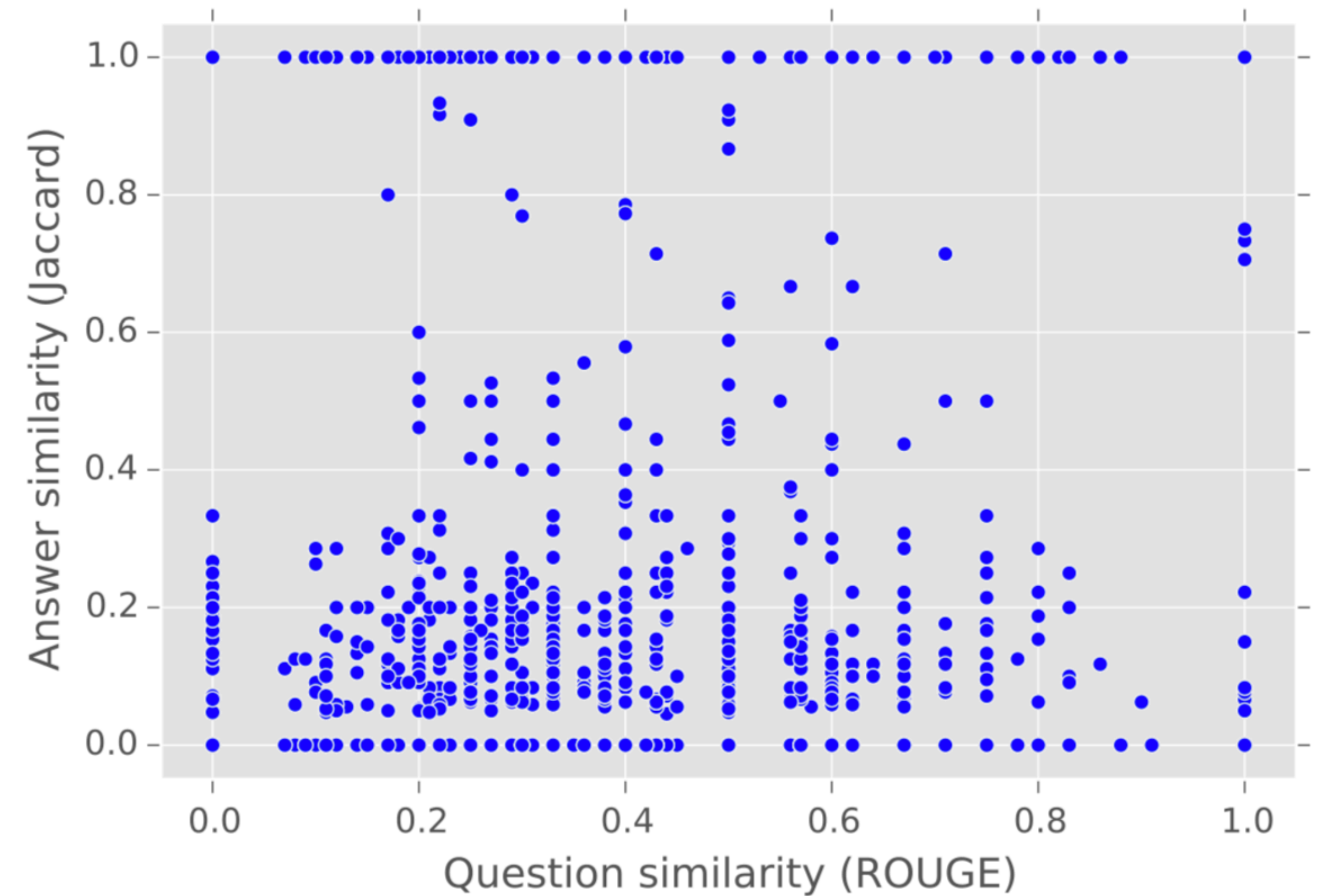
Reading Comprehension

Results

Question-answer similarity



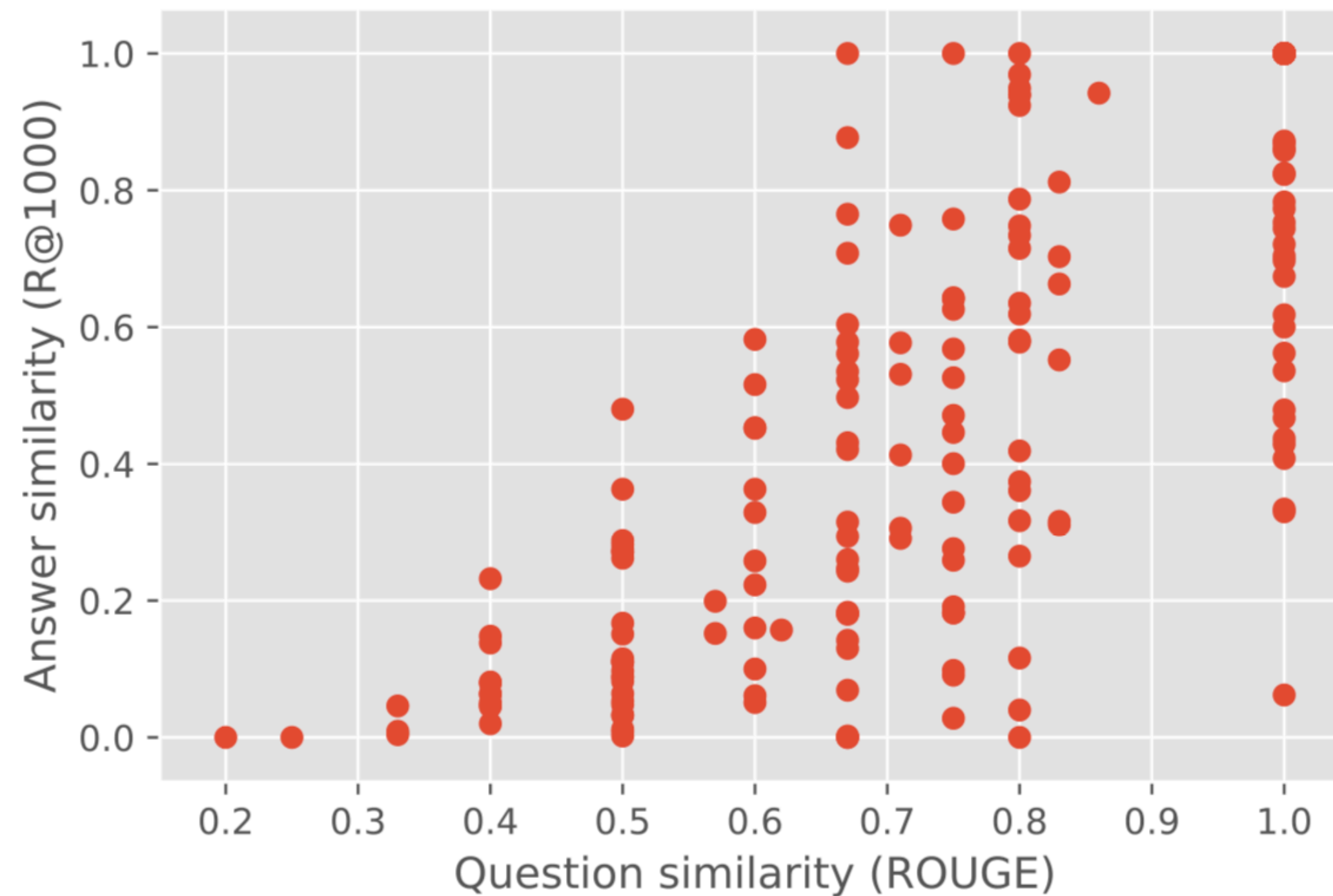
Passage Ranking



Reading Comprehension

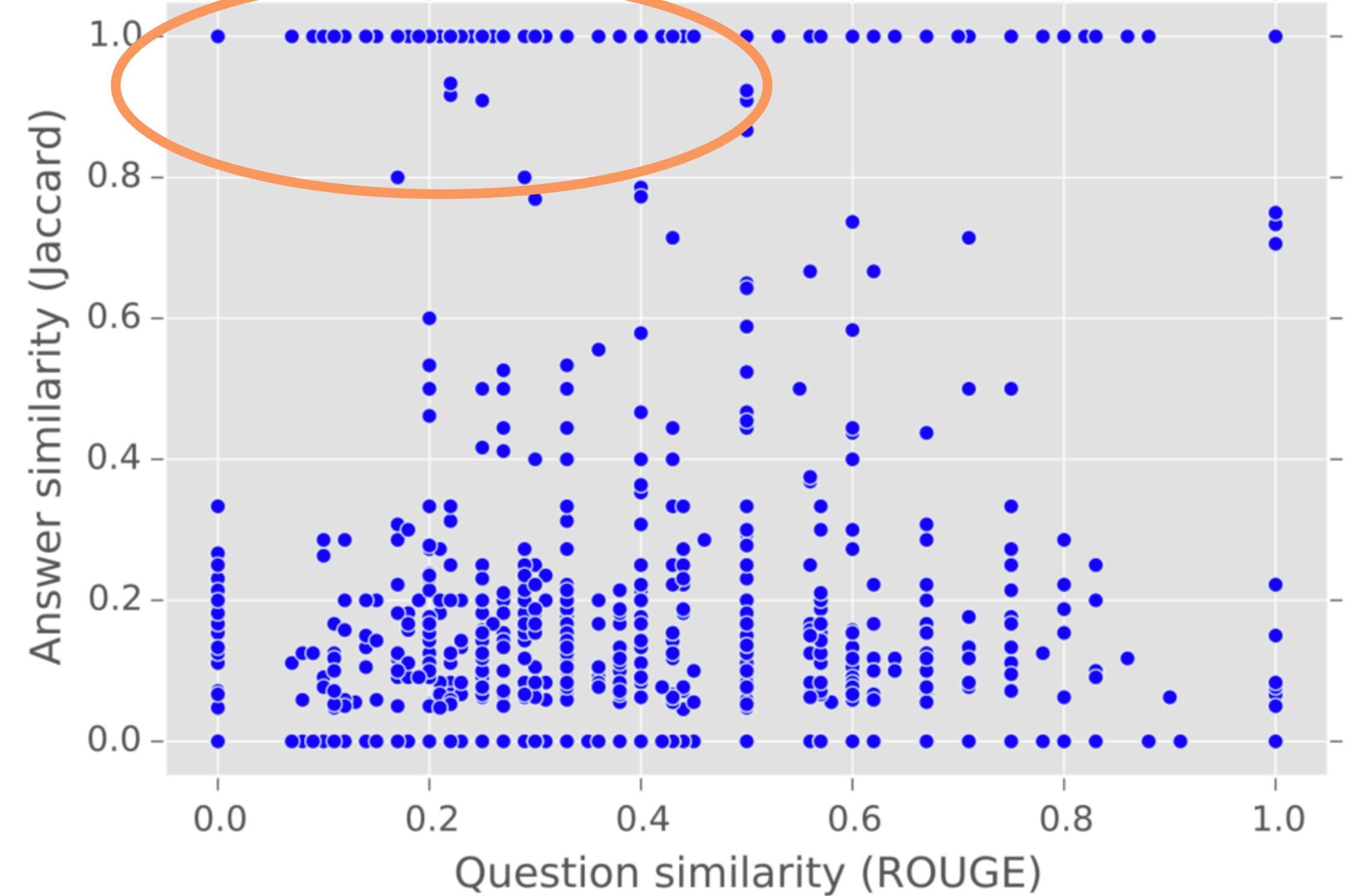
Results

Question-answer similarity



Passage Ranking

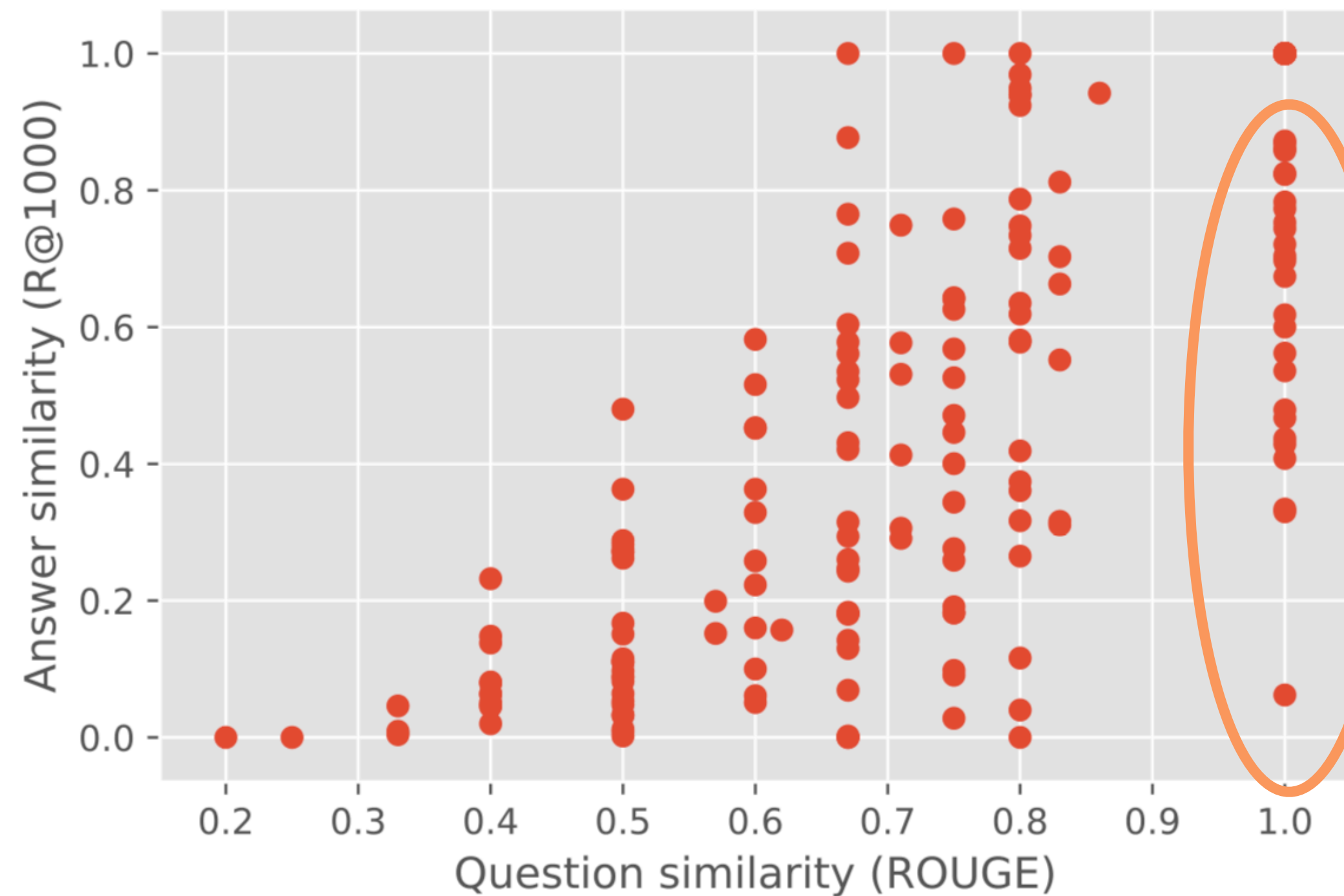
insensitive



Reading Comprehension

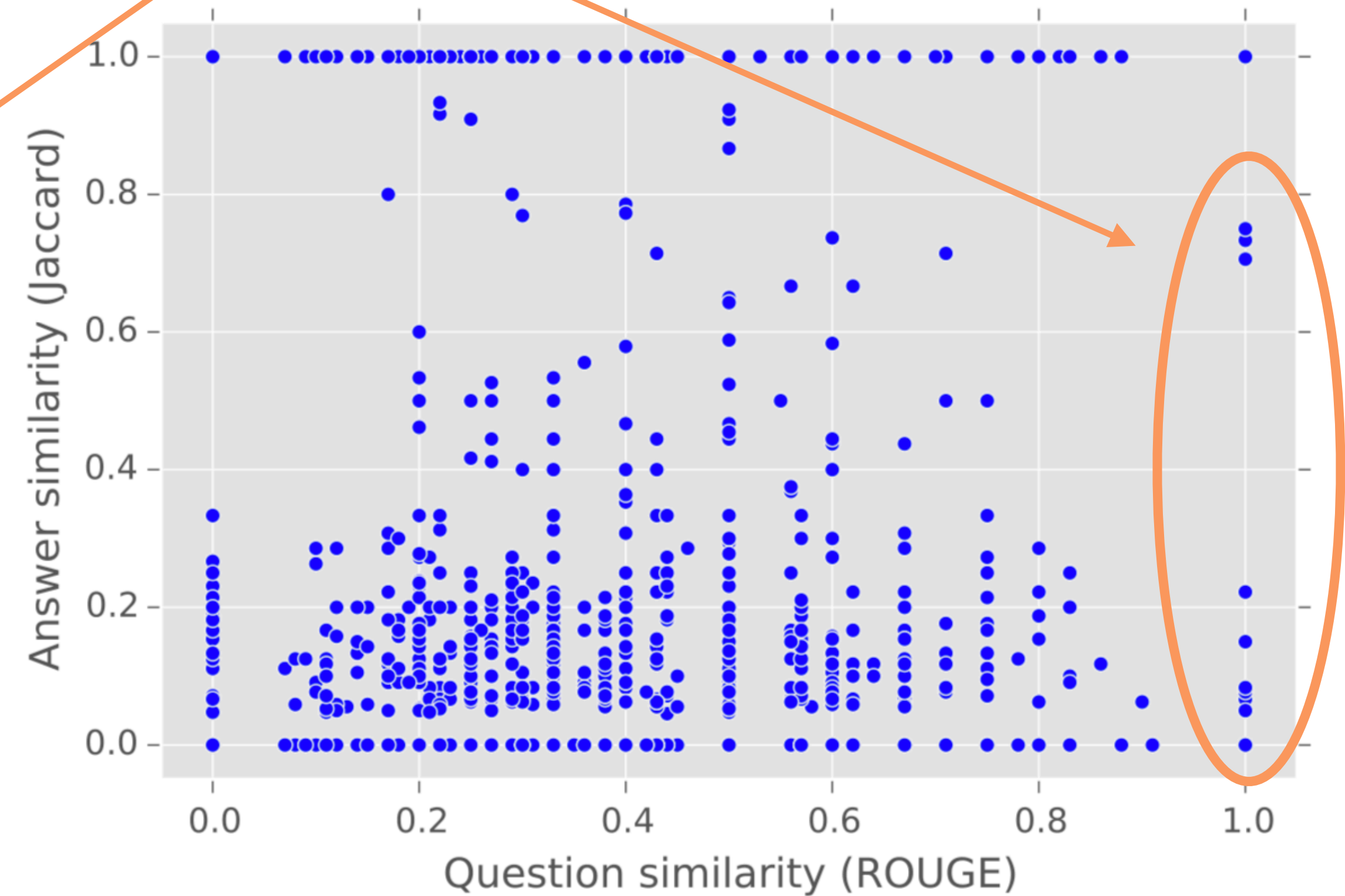
Results

Question-answer similarity



Passage Ranking

oversensitive



Reading Comprehension

Project

Question Rewriting for Conversational QA

Take-away

Question rewrites are useful for QA model evaluation!