

Auto-FAQ

Nagasai Chandra,
Teja Kalvakolanu,
Michael Fekadu

Table of Contents

01

Background

02

Implementation

03

Evaluation

04

Results

05

Conclusion

01

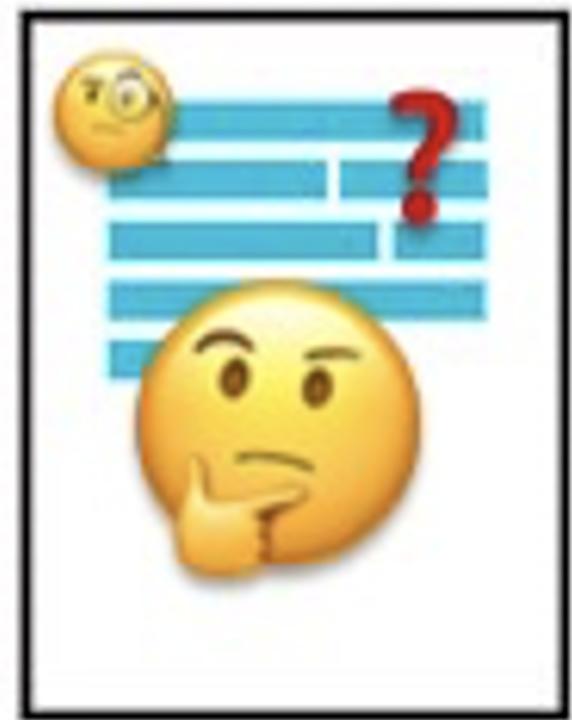
Background

What is the whole idea?









Summarization
gives us ***salience***
within a
paragraph

PageRank gives us
salient sentences

*from the
summaries*

So if we **generate**
questions from
those ***salient***
sentences

We suppose
that gives us
FAQs

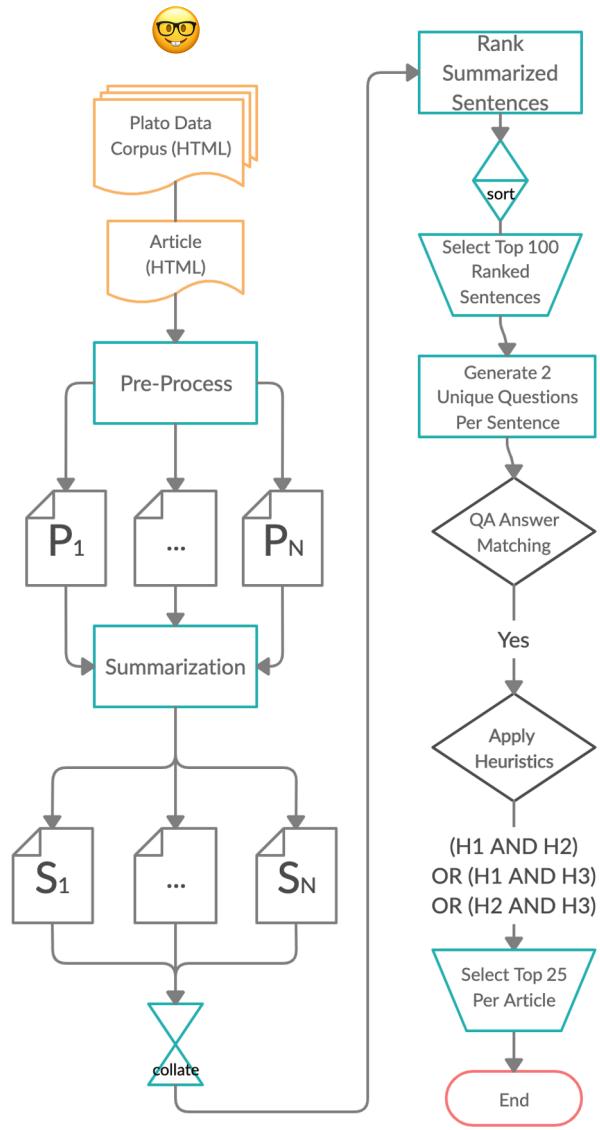
We suppose
that gives us
FAQs

* note

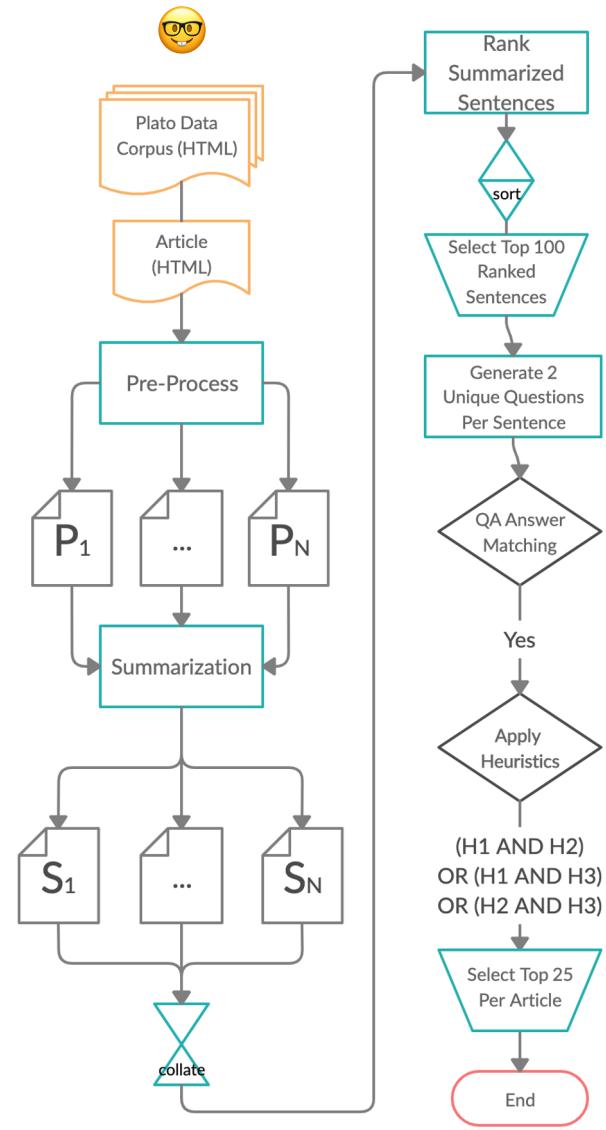
We assume that
salience is a good
proxy for frequent
Qs by readers

02

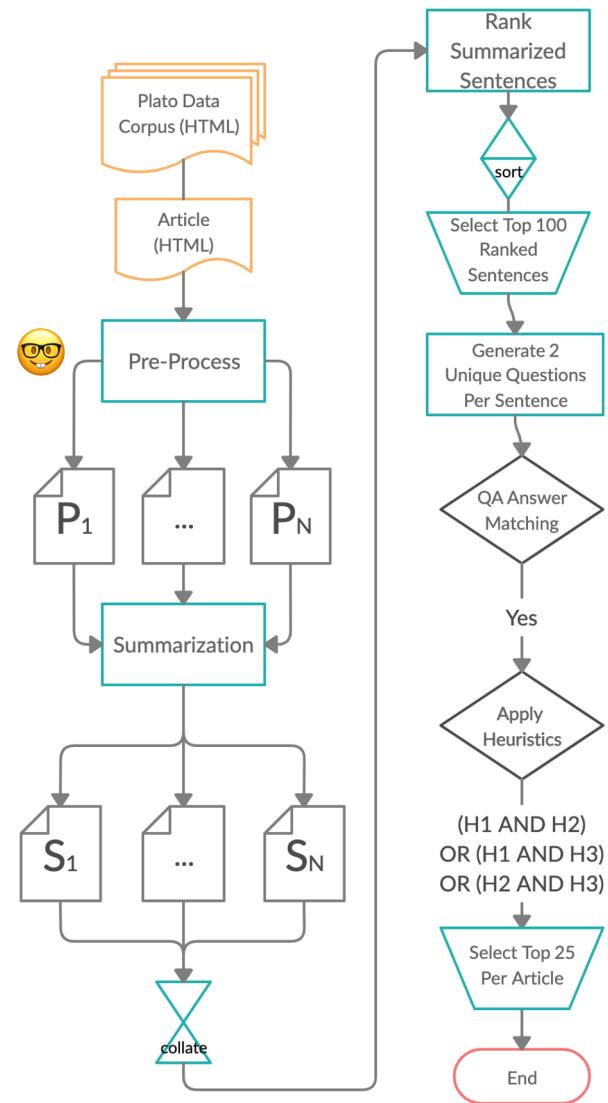
Implementation



Thank you
Dr.
Khosmood
for your
help with
scraping
the data

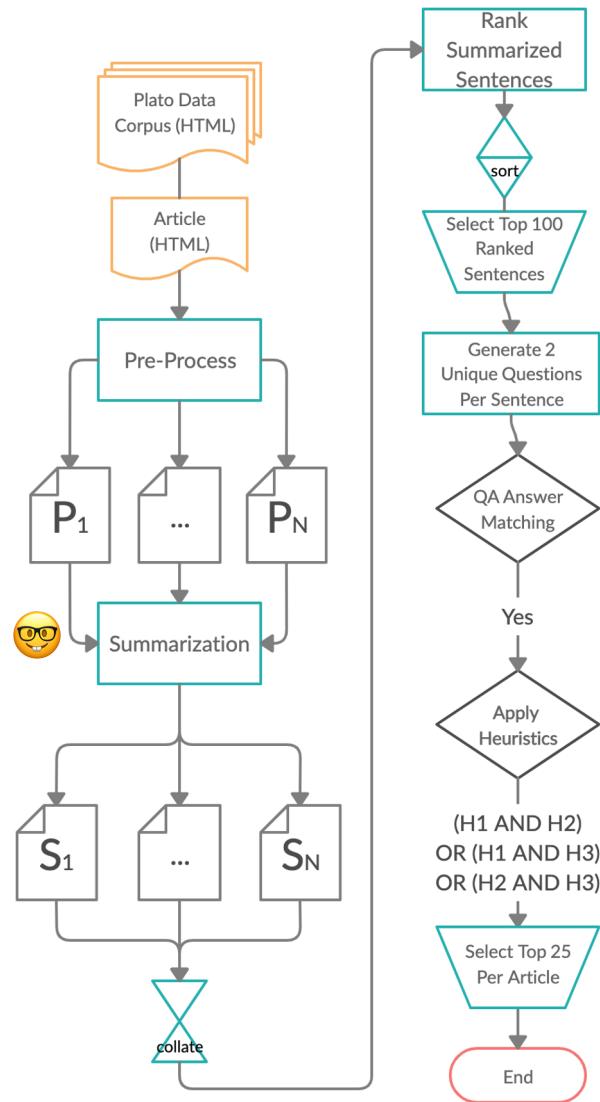


We processed the data.



1. Extracted articles from HTML
1. Organized into sections & paragraphs

We ran
summarization
on each
paragraph

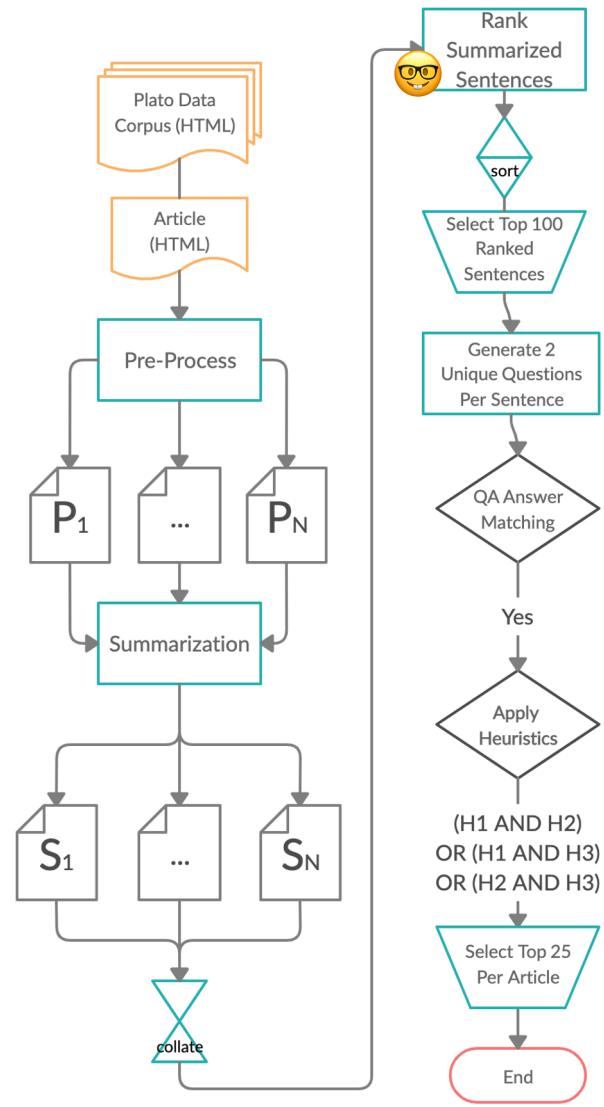


Used HuggingFace
pipeline for
summarization

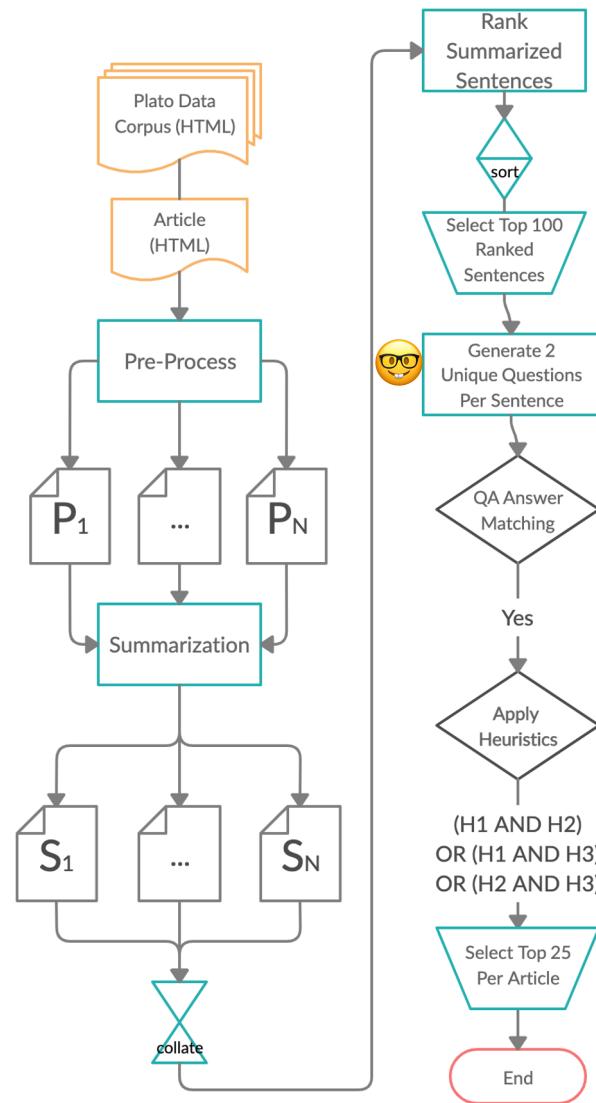
- **Paper:** Liu et al., [Text Summarization with Pretrained Encoders](#).
- **Dataset:** CNN/DailyMail
- **Rouge-1 score** of 43.25

We ran **TextRank**

across all
summarized sentences



We ran question generation

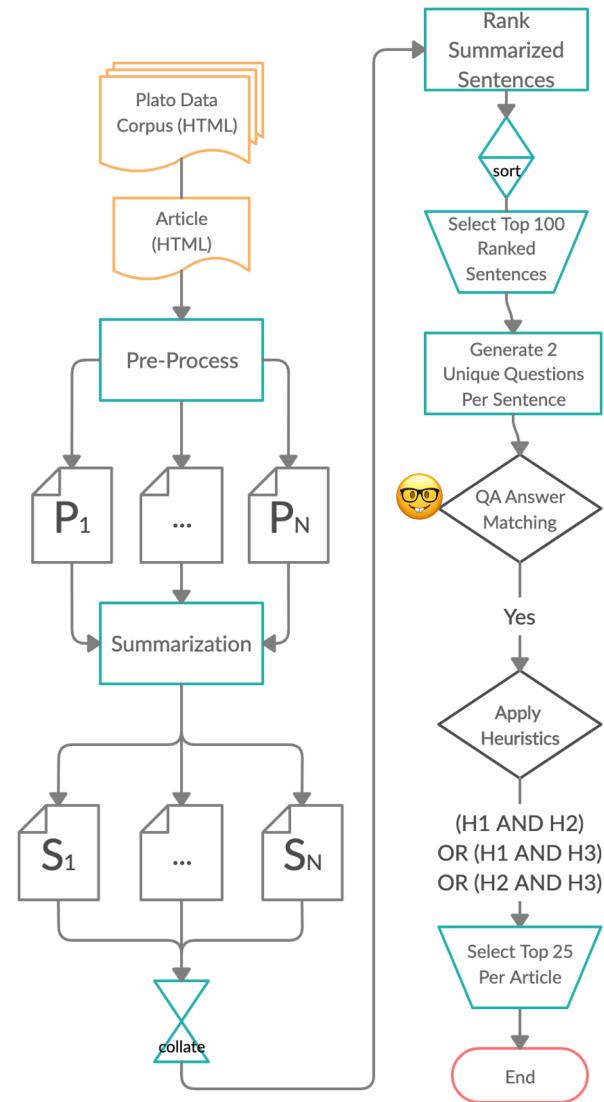


Used pretrained model

- **Paper:** Kriangchaivech, Question Generation by Transformers
- **Dataset:** trained on inverse SQuAD consisting of 100,000+ questions posed on Wikipedia articles.
- **WER** was used as metric to compare btw SQuAD questions and model-generated questions
- Their results showed that generated questions are mostly grammatically correct and plausible

Heuristic 0

QA should match QG answer



Question:

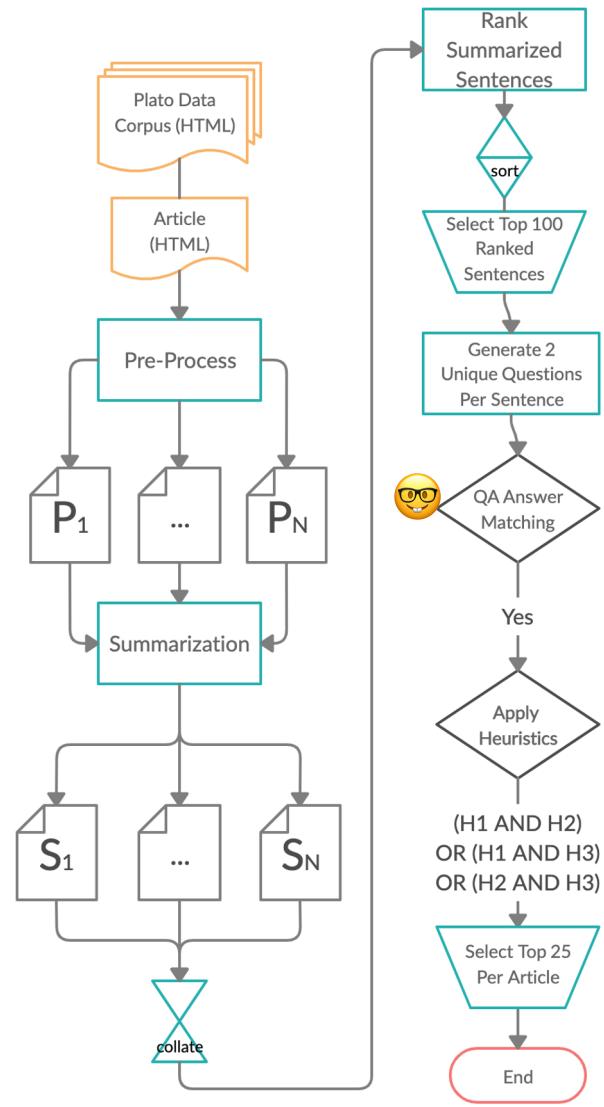
- QG output

Context:

- Sentence given to QG

Heuristic 0

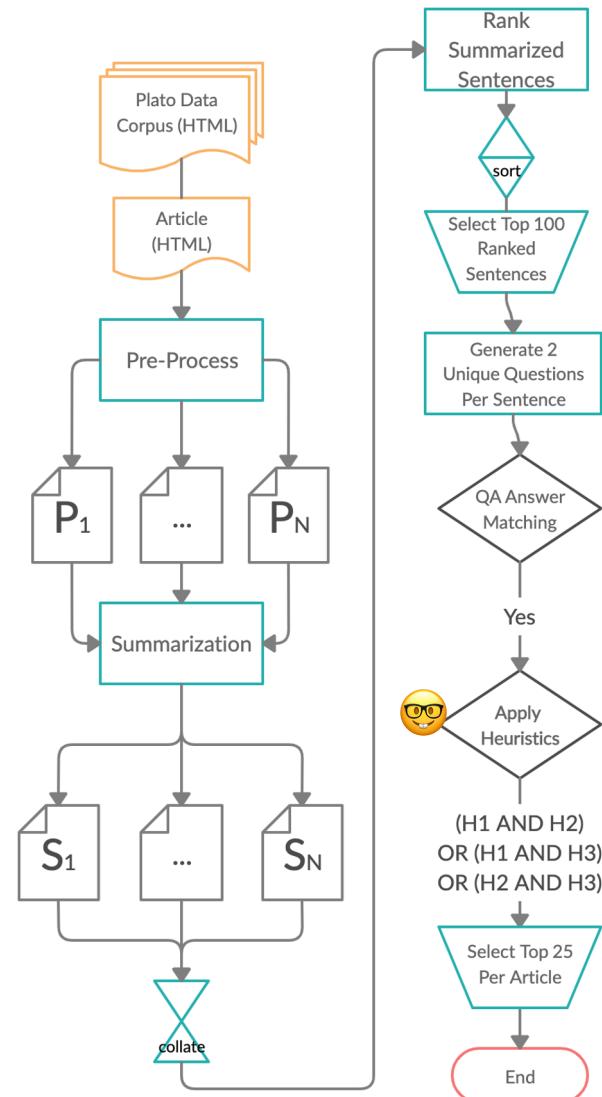
QA should match QG answer



- We used the QA system trained on SQuAD in
 - **Paper:** Devlin et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
 - **Py Package:** HuggingFace
- They achieved an average **F1 score of 88** with their evaluation.

Heuristic 1

QG grammar is good



GrammarBot API

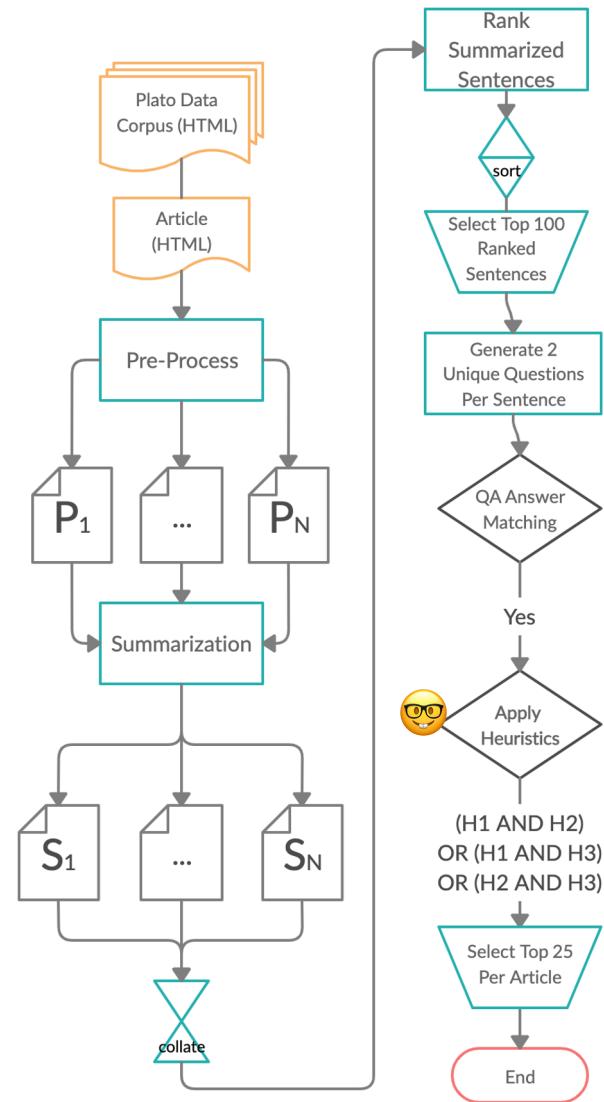
Example bad grammar:

What type of cancer
is a major
component of the
causal basis of?

What are there
sixteen such as?

Heuristic 2

QG has only one WH-word



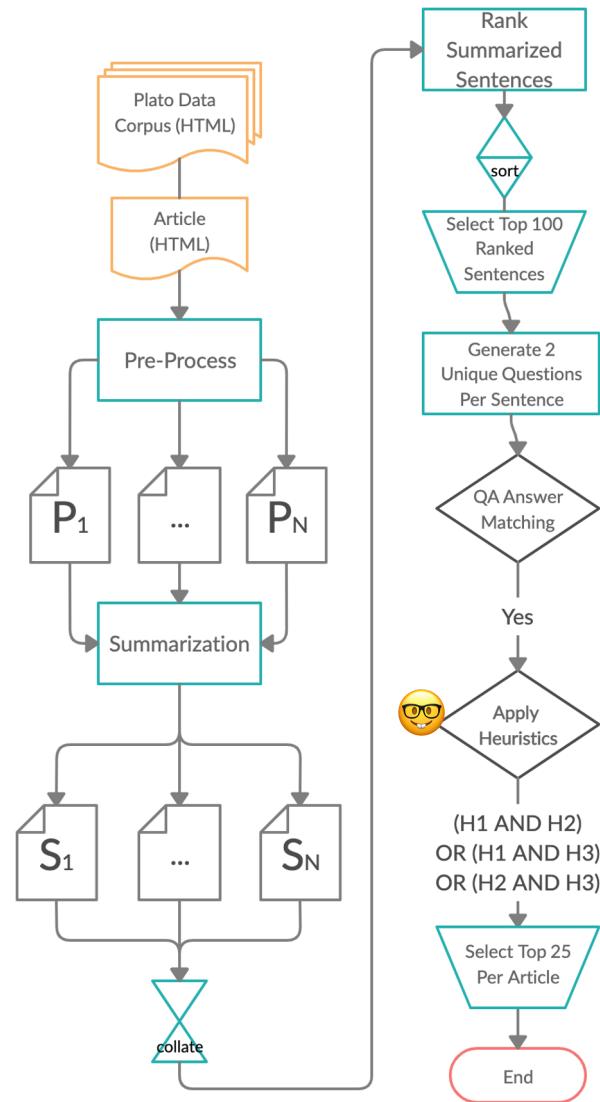
Example bad
QG:

“What is the
meaning of
what?”

↖_(ツ)_↗

Heuristic 3

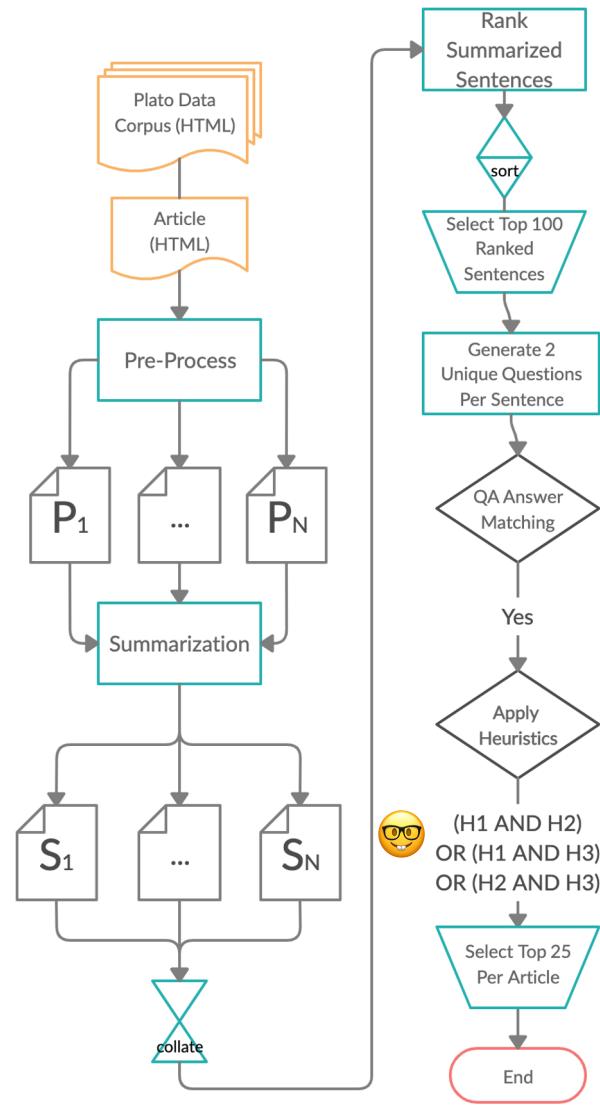
Question
should contain
Named Entity



Example
preferred QG:

“What did
Aristotle do on
Plato’s birthday?”

Combined heuristics by at least two are True



select top 25 per article

03

Evaluation

Hello, anonymous-hippo-36792

This is a survey about question generation ([question answering](#))

Every time you visit this website, a new question & paragraph will appear for evaluation.

use the same browser every time please 😊 because that keeps your pseudonym the same.

Consider this article (it's about philosophy):

[Evolutionary Epistemology](#)

Don't worry, you don't have to read it. Just skim and get the gist of it 😊

Consider this question:

qid 1014

▼ click to expand generated questions

What kind of phenomenon has evolutionary epistemology been concerned with?

Now, the following should answer the question:

sec_id 1.0 par_id 7.0

▼ click to expand paragraph

toggle LaTeX

Every scientific enterprise requires formal and semi-formal models which allow the quantitative characterization of its objects of study. The attempt to transform the philosophical study of knowledge into a scientific discipline which approaches knowledge as a biological phenomenon is no different. [much of the evolutionary epistemology literature has been concerned with how to conceive of knowledge as a natural phenomenon, what difference this would make to our understanding of our place in the world, and with answering objections to the project.](#) There are, as well, a number of more technical projects which attempt to provide the theoretical tools necessary for a naturalistic worldview.

How was the grammar of the question? *

1 = question grammar is BAD

5 = question grammar is GOOD



Is the question meaningful? *

yes

no

maybe

How well did the paragraph answer the question? *



Why maybe? *

*only required
if "maybe"*



Evaluation Method

FAQs of 4 articles were rated

Total around 300 ratings with at least 3 people rating each question

Criteria:

- Grammar of the question rated (1-5),
- Whether the question makes sense in relation to article(Yes, NO, Maybe). We asked Explanation for Maybe.
- Whether the passage Answers the Question: Ans

04

Results

Bad Coreferencing, good questions?

What does he do with psycho - physical parallelism?

Now, the following should answer the question:

sec_id 2.0 par_id 9.0

▼ click to expand paragraph

toggle LaTeX

In 1932 [1996], von Neumann suggested that the entangled state of the object and the instrument collapses to a determinate state whenever a measurement takes place. This measurement process (a type 1-process as he called it) could not be described by quantum mechanics. He argues that during a measurement the actual observer gets a subjective perception of what is going on that has a non-physical nature. **however, he holds on to psycho-physical parallelism as a scientific principle, which he interprets such that there exists a physical correlate to any extra-physical process of the subjective experience.**

Too Generic Questions?

The screenshot shows a web browser window with a light blue header bar containing standard OS X-style icons (red, yellow, green circles, back, forward, etc.) and the URL 582eval.glitch.me. The main content area has a white background and is divided into two sections by a horizontal line.

Consider this question:

qid 294

▼ click to expand generated questions

What has many causes , and many effects?

Now, the following should answer the question:

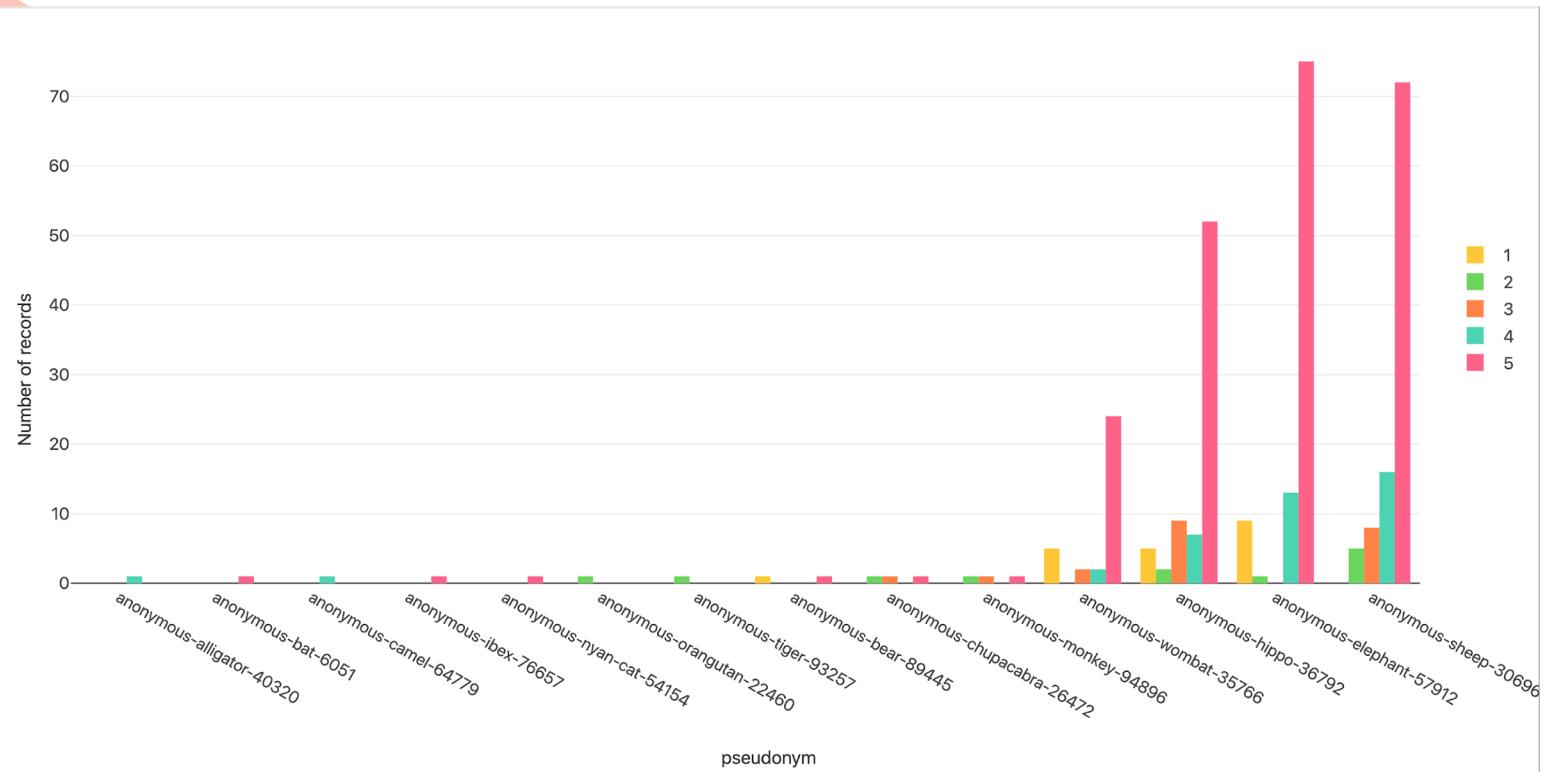
sec_id 5.0 par_id 3.0

▼ click to expand paragraph

toggle LaTeX

The history of attempts to identify defining features of cancer, let alone arrive at a unified theory, has floundered. Either cancer is defined so vaguely as to include non-pathological states, or focused so narrowly on a specific class of causes. **such a variety of definitions is due to the fact that cancer has many causes, and many effects.** It involves many different types of dysregulation, at a variety. of temporal and spatial scales. It is no small challenge to identify defines cancer, and arrive at unified "theory" of disease etiology.

Histogram of Grammar Ratings Per Evaluator



Mostly the 3 of us evaluating, but some friends helped too

Correlation between Grammar and Answerability

Group by grammar rating

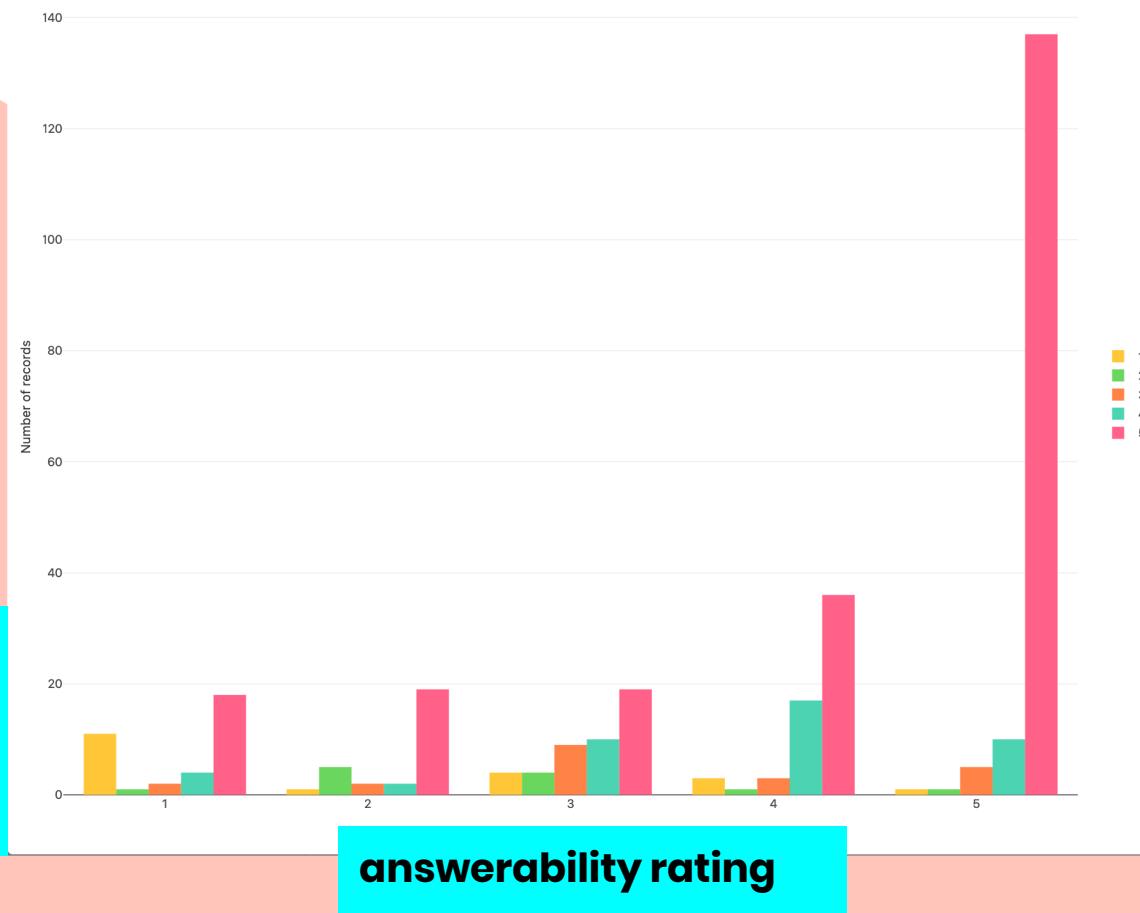
How was the grammar of the question?

1 = question grammar is BAD

5 = question grammar is GOOD



How well did the paragraph answer the question?



Group by grammar rating

How was the grammar of the question?

*

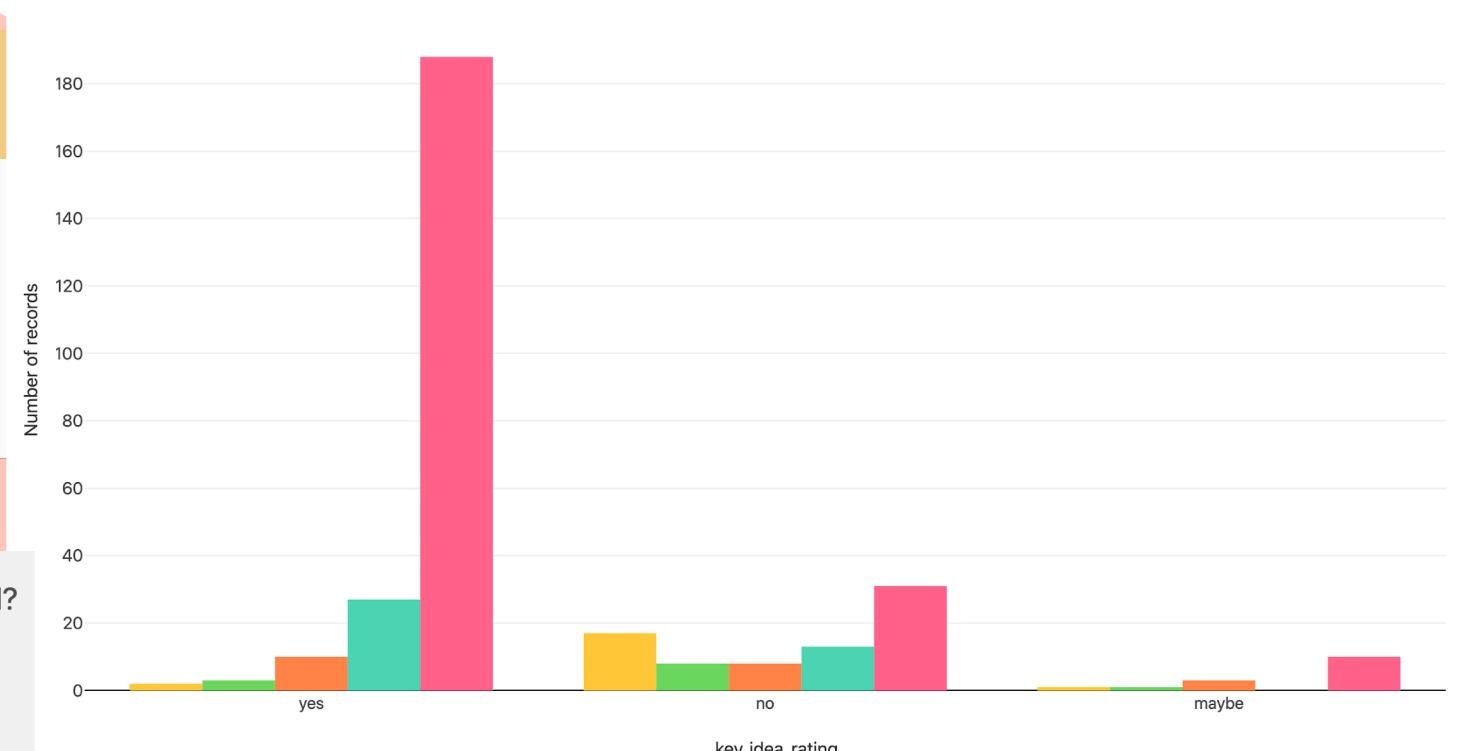
1 = question grammar is BAD

5 = question grammar is GOOD



Is the question meaningful?

- yes
- no
- maybe



Results based on Document

TITLE (FROM FILTEREDQUESTIONS)		Count	grammar_rating	answerability_rating	has_yes	has_no	has_maybe
▶	Cancer	79	▼ Avg 4.06	▼ Avg 3.41	▼ Sum 50	▼ Sum 26	▼ Sum 3
▶	Copenhagen Interpretation of Quantum Mechanics	85	▼ Avg 4.54	▼ Avg 3.89	▼ Sum 62	▼ Sum 20	▼ Sum 3
▶	Cusanus, Nicolaus [Nicolas of Cusa]	81	▼ Avg 4.33	▼ Avg 4.12	▼ Sum 62	▼ Sum 14	▼ Sum 5
▶	Evolutionary Epistemology	76	▼ Avg 4.59	▼ Avg 3.84	▼ Sum 55	▼ Sum 17	▼ Sum 4

05

Conclusion

Good Questions

What is the wall of paradise?

Cusanus leads us through a series of reflections on seeing and on the face of God. God is located beyond both imaginative exercise and conceptual understanding.

nicholas symbolizes our approach to this beyond by encouraging us to enter "into a certain secret and hidden silence wherein there is no knowledge or concept of a face," characterizing it as an "obscuring mist, haze, darkness or ignorance." he invokes the coincidence of opposites from on learned ignorance and proposes his second central metaphor: the wall of paradise.

Are we more familiar with the phrase " the not other " in negative ways of stating self - identity?

In De li Non Aliud/On the Not Other Nicholas returns to the ancient Platonic categories of the One and the Other in order to re-construe in novel language what Christians believe about the dependence of creatures on their Creator. Several of Cusanus' later works use verbal coinages or Latin neologisms as "names" or characterizations of God that are original with him, though they have earlier echoes in Christian Neoplatonism. **in this dialogue he uses the expression "the not other" as a substantive for god as the divine not-other, even though we are more familiar with the phrase in negative ways of stating self-identity.**

Good Questions

What is one way we can speculate about the non - finite presence - in - absence?

we are able to speculate about the case of a non-finite presence-in-absence by starting with but moving beyond the limits of the presence and absence we are familiar with in the realm of limited things. Now we turn to the presence of the unfamiliar infinite One as what is finally required, even if not obviously experienced, to keep the creaturely image present and real. In this case, we may think that absence becomes what is metaphorical. But God's presence is hardly like that of one physical thing to another. It is ineluctably and literally necessary to explain the reality of anything and everything.

Bad Questions

What does he do with psycho - physical parallelism?

In 1932 [1996], von Neumann suggested that the entangled state of the object and the instrument collapses to a determinate state whenever a measurement takes place. This measurement process (a type 1-process as he called it) could not be described by quantum mechanics. He argues that during a measurement the actual observer gets a subjective perception of what is going on that has a non-physical nature. **however, he holds on to psycho-physical parallelism as a scientific principle, which he interprets such that there exists a physical correlate to any extra-physical process of the subjective experience.**

What is the term for fitness in cancer?

Cancer is a product of breakdown in mechanisms that permitted the emergence of multicellularity. The mechanisms that protect us from "revolt from within" are not error-free, and over time will fail. **on this view, cancer itself may be viewed as an evolutionary process, where these "adaptations" enable short term "fitness," or relative success at survival and reproduction.** Several scientists have developed theoretical models of this process, linking it to empirical data, e.g., on emergence of chemotherapy resistance.

Overall

This again links with coreference resolution

The knowledge base plays a huge role .

Trained on SQuAD and Used for Plato

Performed well with facts based articles
compared to others

Futurework

Train with the Plato Tagged data

Include coreference resolution

Division of Work

Michael: Data Processing, Evaluation Infrastructure, Heuristics, Process Pipeline, Involved in all research ideas

Teja Kalvakolanu: Text Rank, Question Generation, Question Answering, Process Pipeline, involved in all research ideas

Nagasai: Summarization, Question Generation, question Answering, Process Pipeline, involved in all research ideas

Thanks!

Do you have any questions?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.



References

- [1] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Brew, J. (2019). Huggingface's transformers: State-of-the-art natural language processing. *ArXiv, abs/1910.03771*.
- [2] Malpas, J., "Donald Davidson", *The Stanford Encyclopedia of Philosophy* (Winter 2012 Edition), Edward N. Zalta (ed.),
- [3] Kriangchavech, Kettip, and Artit Wangperawong. "Question Generation by Transformers." arXiv preprint arXiv:1909.05017 (2019).
- [4] Prateek Joshi, "Introduction to text Summarization", Analytics Vidhya, 2018.
- [5] Mihalcea, R. (2004, July). Graph-based ranking algorithms for sentence extraction, applied to text summarization. In Proceedings of the ACL Interactive Poster and Demonstration Sessions (pp. 170-173). <https://www.aclweb.org/anthology/P04-3020.pdf>

