AnswerBot - Automated Generation of Answer Summary to Developers' Technical Questions

Bowen Xu, David Lo

Motivation

Problem Solving Process

Problem	Query	Relevant Documents	Target Information	Solved

Survey

Redundant Information

Irrelevant / Low-quality Information
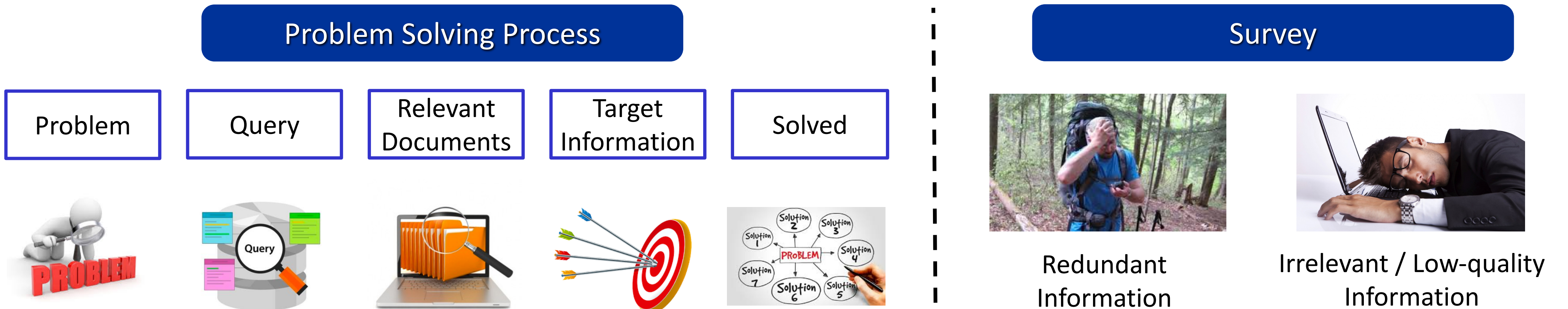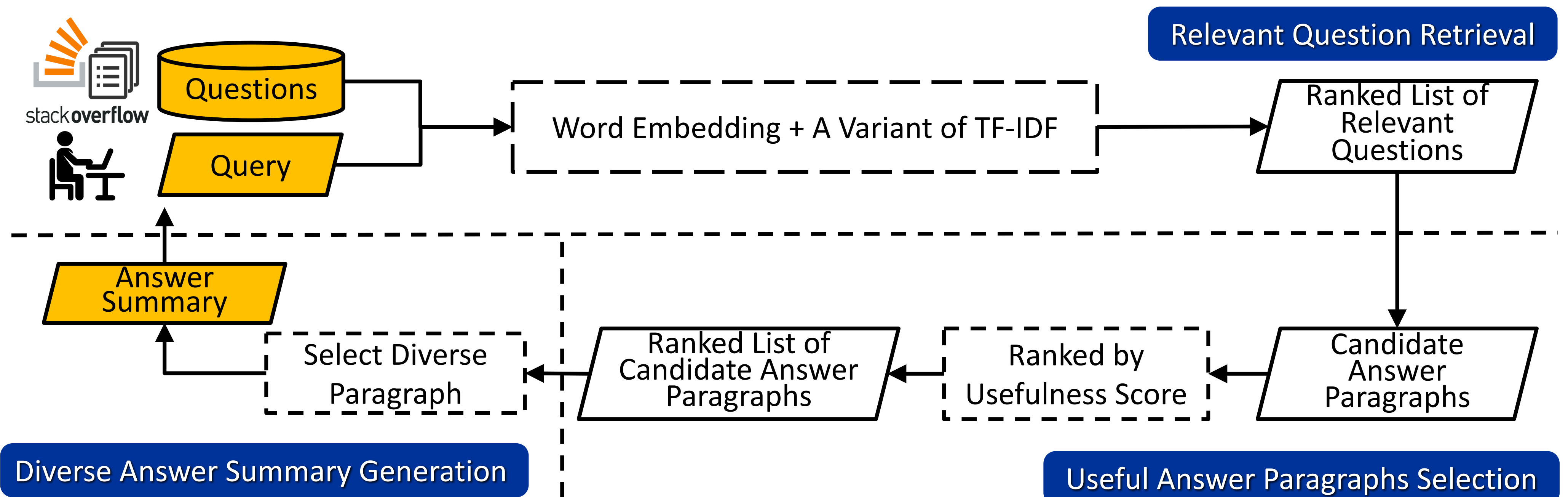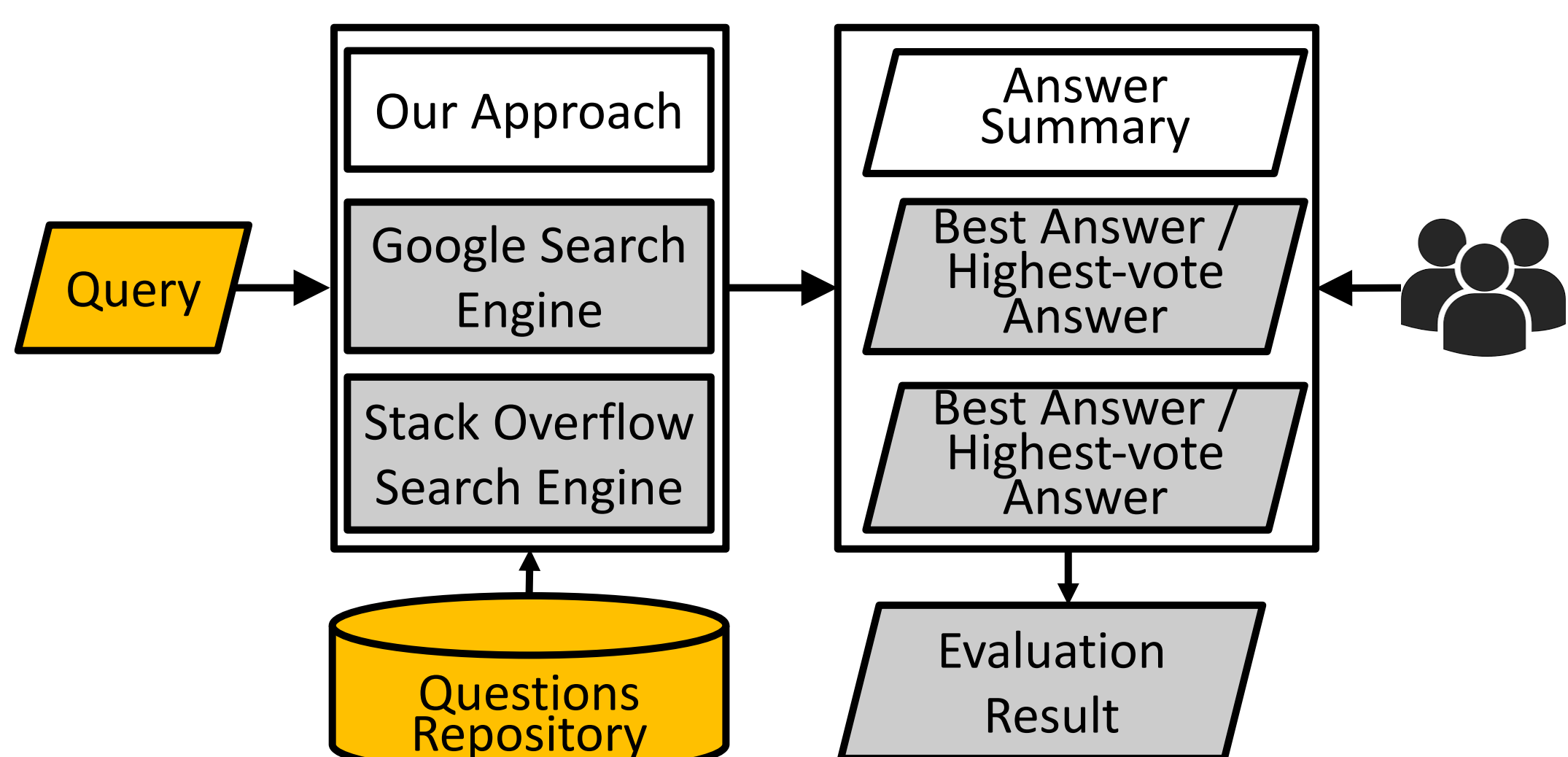
> Developers really need an effective technique to generate an answer summary with diverse, relevant and high-quality information from answer posts.

Overview of Approach

Relevant Question Retrieval

stack**overflow**

Questions → Query → Word Embedding + A Variant of TF-IDF → Ranked List of Relevant Questions

Answer Summary ← Select Diverse Paragraph ← Ranked List of Candidate Answer Paragraphs ← Ranked by Usefulness Score ← Candidate Answer Paragraphs

Diverse Answer Summary Generation

Useful Answer Paragraphs Selection

Evaluation & Result

Query → [Our Approach / Google Search Engine / Stack Overflow Search Engine] → [Answer Summary / Best Answer / Highest-vote Answer / Best Answer / Highest-vote Answer] → Evaluation Result

Questions Repository

	Relevance	Usefulness	Diversity
Our Approach	3.450	3.720	3.830
Google (Baseline)	3.440	3.480*	2.930***
Stack Overflow (Baseline)	2.576***	2.712***	2.305***

***p<0.001, **p<0.01, *p<0.05

> Our approach achieves best performance.

Conclusion

> Best or highest-vote answers may not cover as diverse information as the developers need.
> It is worth reading more answer posts to summarize more complete information.