



11693 Final Project

Biomedical Question Answering System

Team 6

Rui Wang, Victor Zhao,

Carol Cheng, Hua Tang, Yan Zhao



Outline

- Overview
- Architecture
- Performance
- Improvement
- Sprint Review

Overview - Objective

- Build a biomedical question answering system
- Answer yes/no questions

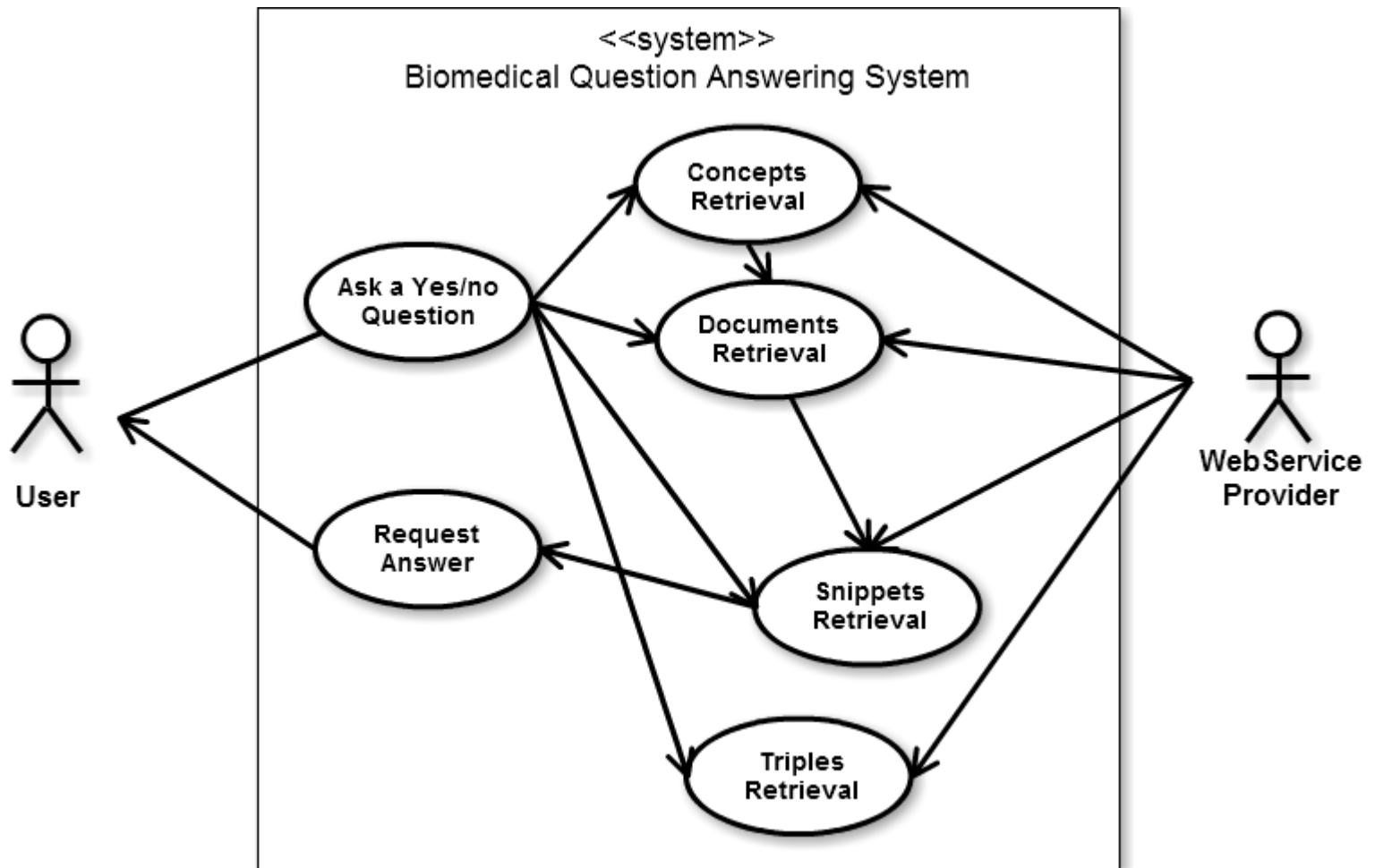
Question:

Is Mammaprint approved by the United States Food and Drug Administration?

Answer:

Yes

Overview - Use Case



Overview - Milestones

Milestone 1

- Concepts, documents, and triples retrieval

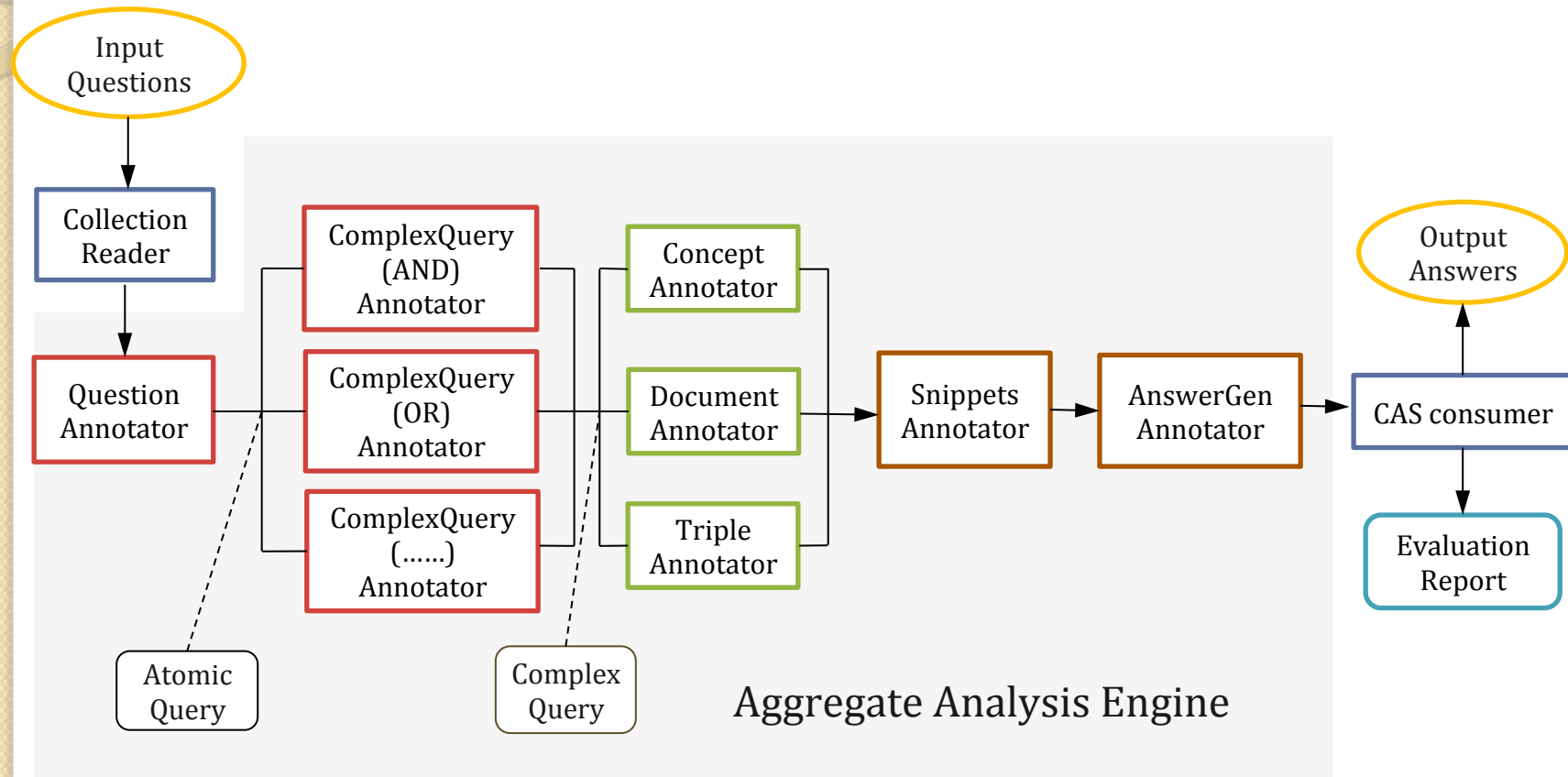
Milestone 2

- Snippets retrieval

Milestone 3

- Exact answer generation

Architecture



Query Extraction

- ✓ Question preprocessing
 - Remove punctuation
 - Remove stop words
 - Do stemming

Question

Is Mammaprint approved by the United States Food and Drug Administration?

AtomicQueryConcept

Mammaprint

approve

United

States

Food

Drug

Administration

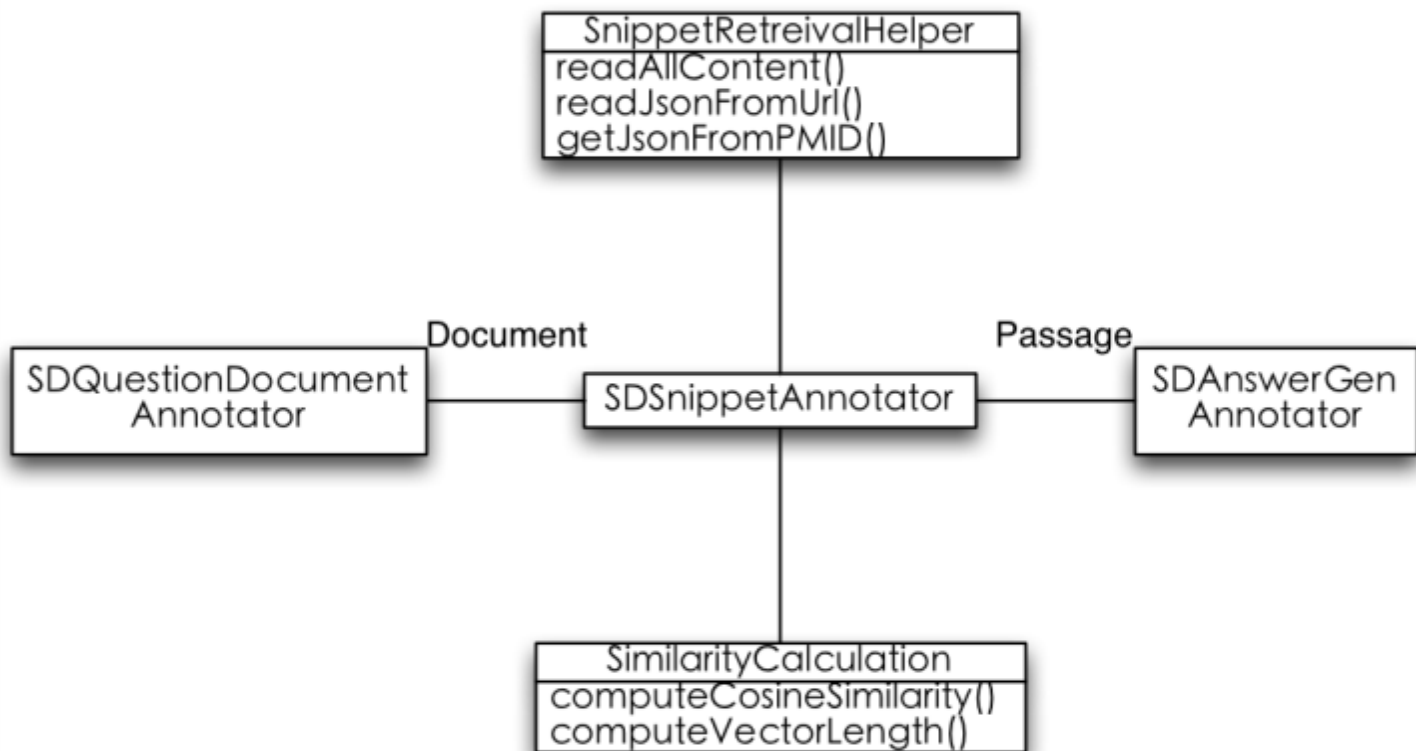
ComplexQueryConcept

Mammaprint AND approve AND United AND States AND Food AND Drug AND Administration?

Snippet Retrieval

Two helper classes

Collect articles by the given documents



Calculate the similarity

Performance

Milestone 1

	Concept	Document	Triple
MAP	0.4699	0.0737	0.0000
GMAP	0.0574	0.0026	0.0010
Total: 29 Questions			

Milestone 2 – Snippet retrieval

The 11's sentence is: This has been proposed to be responsible for improvement of certain autoimmune pathologies such as rheumatoid arthritis during pregnancy
Similarity is:0.2236
VVVVVVVVVVV
The stopArticle is : 1. Introduction Many connective tissue diseases have abnormal immune system activity with inflammation in tissues as a result of an immune system that
The query is:be rheumatoid arthritis more common in man or woman
The 12's sentence is: Similarly women have higher levels of prolactin and growth hormones compared to males
Similarity is:0.0
VVVVVVVVVVV
The stopArticle is : 1. Introduction Many connective tissue diseases have abnormal immune system activity with inflammation in tissues as a result of an immune system that
The query is:be rheumatoid arthritis more common in man or woman
The 13's sentence is: These pituitary hormones also enhance autoimmunity [1]
Similarity is:0.0
VVVVVVVVVVV
The stopArticle is : 1. Introduction Many connective tissue diseases have abnormal immune system activity with inflammation in tissues as a result of an immune system that
The query is:be rheumatoid arthritis more common in man or woman
The 14's sentence is: The effect of hormonal factor in inflammatory eye disease is complex and is not uniform among all diseases; male patients with Behçet's syndrome have
Similarity is:0.0572
VVVVVVVVVVV
The stopArticle is : 1. Introduction Many connective tissue diseases have abnormal immune system activity with inflammation in tissues as a result of an immune system that
The query is:be rheumatoid arthritis more common in man or woman
The 15's sentence is: Nonhormonal gender differences such as environmental exposures, drugs (more commonly used by one gender than the other), or infectious organisms can
Similarity is:0.063
VVVVVVVVVVV
The stopArticle is : 1. Introduction Many connective tissue diseases have abnormal immune system activity with inflammation in tissues as a result of an immune system that
The query is:be rheumatoid arthritis more common in man or woman
The 16's sentence is: Autoimmune diseases affect the eye in different ways
Similarity is:0.1111
VVVVVVVVVVV
The stopArticle is : 1. Introduction Many connective tissue diseases have abnormal immune system activity with inflammation in tissues as a result of an immune system that
The query is:be rheumatoid arthritis more common in man or woman
The 17's sentence is: We will summarize the ocular manifestations of the most encountered connective tissue diseases and vasculitides
Similarity is:0.0
maxId is11

Performance

Without eliminating stop words

	Concept	Document	Triple
No concept score threshold			
MAP	0.4309	0.1117	0.0000
GMAP	0.1227	0.0035	0.0010
Concept score > 0.1			
MAP	0.4511	0.0743	0.0000
GMAP	0.0462	0.0023	0.0010
Concept score > 0.15			
MAP	0.4569	0.0743	0.0000
GMAP	0.0399	0.0023	0.0010
Concept score > 0.2			
MAP	0.3046	0.0743	0.0000
GMAP	0.0124	0.0023	0.0010
Total: 29 Questions			

Performance

With eliminating stop words

	Concept	Document	Triple
No concept score threshold			
MAP	0.4309	0.1117	0.0000
GMAP	0.1227	0.0034	0.0010
Concept score > 0.1			
MAP	0.5043	0.1117	0.0000
GMAP	0.0882	0.0035	0.0010
Concept score > 0.15			
MAP	0.4847	0.1117	0.0000
GMAP	0.0502	0.0035	0.0010
Concept score > 0.2			
MAP	0.4502	0.1117	0.0000
GMAP	0.0396	0.0035	0.0010
Total: 29 Questions			

Performance

Summary:

- With eliminating stop words has better performance.
- Different concept score threshold result to different performances.

Hence, we finally choose eliminating stop words and concept score threshold = 0.1.

Concept score > 0.1			
MAP	0.5043	0.1117	0.0000
GMAP	0.0882	0.0035	0.0010

Improvement

- Problems in our system
 - Low document MAP
 - Triple MAP = 0

Improvement

- Query Extension: Regenerate updated query text for document searching
 - Advanced combination strategy for received concepts from different complex query concept annotators.
 - Adding concepts into query text, with different operators.

Improvement

- Document result ranking: re-rank received documents.
- Original received documents does not arrange in reasonable ordering.
- Try different ranking strategy and get documents in reasonable ranking.



Product Backlog



Sprint Burndown Chart



Story Wall

Whole Project Retrospective

Team Member	Set the Stage (how did you feel this project)	What Happened in this project	Why did those things happen	What To Do to keep up what worked well and prevent what didn't
Rui Wang	A good chance to communicate and work together	Help teammates to be familiar with the whole pipeline. Clarify the structure of the system and the strategy used to get better performance	Aggressive learner. Try to utilize lots of new concepts and techniques.	Learn how to cooperate with other guys via github, and gain experience on how to merge code and sync repo.
Victor Zhao	Improve a lot	Mainly in charge of building the whole structure of the system, create the snippets retrieval section to help get the exact answer	Try to learn new knowledge and improve all the time	Employ time management strategies to my own life. Be more proactive.
Carol Cheng	Challenging and interesting	Mainly in charge of the documents of the whole project, help to manage sprints. Create complex query concepts from atomic query concepts.	Be very patient on work and proactive on learning details of the project	Interact and communicate with teammates more often.
Yan Zhao	Learn a lot of things	Help to build the pipeline of the whole system, help to improve the performance of the snippets. In charge of the consumer implementation to evaluate the answer retrieved	Try to learn new knowledge of the whole system	Work harder on the task, cooperate more with other teammates and learn to manage the time.
Hua Tang	Exciting	Help to manage sprints and prepare documents. Add complex query processing in Concept Annotator, Document Annotator and Triple Annotator. And extra finishing processing in question into atomic query.	Has great responsibility to finish own work	Begin work earlier