

# Sprint 1 Presentation

Team 2: Xi LIU, Lei XIAO, Yifu WANG, Xiaoxu LU, Yan HE

## Project Management Overview

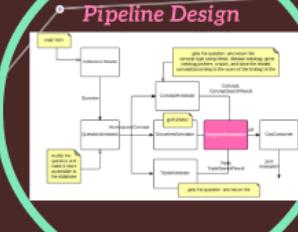
- Burn-down Charts
- Retrospective
- Problems and Solutions

## Project Overview

- **Vision:** Solve yes/no questions with high precision
- **Requirement Analysis:** information flow and evaluation approaches
- **Pipeline Draft**

## Task Checklist

- Read questions
- Convert questions to terms
- Build queries with terms
- Retrieve query results
- Extract snippets from returned documents
- Rank the result by relevance
- Evaluate the overall performance and hierarchical performance



## Pipeline Design



## Achievements

- Enhanced Pipeline Design
  - Question Annotation
  - Snippet Annotation
  - Hierarchical Evaluation
- Updated Documentation
- Studied Collaboration Tools

# Sprint 1 Presentation

Team 2: Xi LIU, Lei XIAO, Yifu WANG, Xiaoxu LU, Yan HE

## Project Management Overview

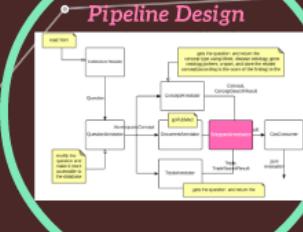
- Burn-down Charts
- Retrospective
- Problems and Solutions

## Project Overview

- **Vision:** Solve yes/no questions with high precision
- **Requirement Analysis:** information flow and evaluation approaches
- **Pipeline Draft**

## Task Checklist

- Read questions
- Convert questions to terms
- Build queries with terms
- Retrieve query results
- Extract snippets from returned documents
- Rank the result by relevance
- Evaluate the overall performance and hierarchical performance



## Pipeline Design



## Achievements

- Enhanced Pipeline Design
  - Question Annotation
  - Snippet Annotation
  - Hierarchical Evaluation
- Updated Documentation
- Studied Collaboration Tools

# Project Overview

- **Vision:** Solve yes/no questions with high precision
- **Requirement Analysis:** information flow and evaluation approaches
- **Pipeline Draft**

# *Task Checklist*

- Read questions
- Convert questions to terms
- Build queries with terms
- Retrieve query results
- Extract snippets from returned documents
- Rank the result by relevance
- Evaluate the overall performance and hierarchical performance

# Achievements

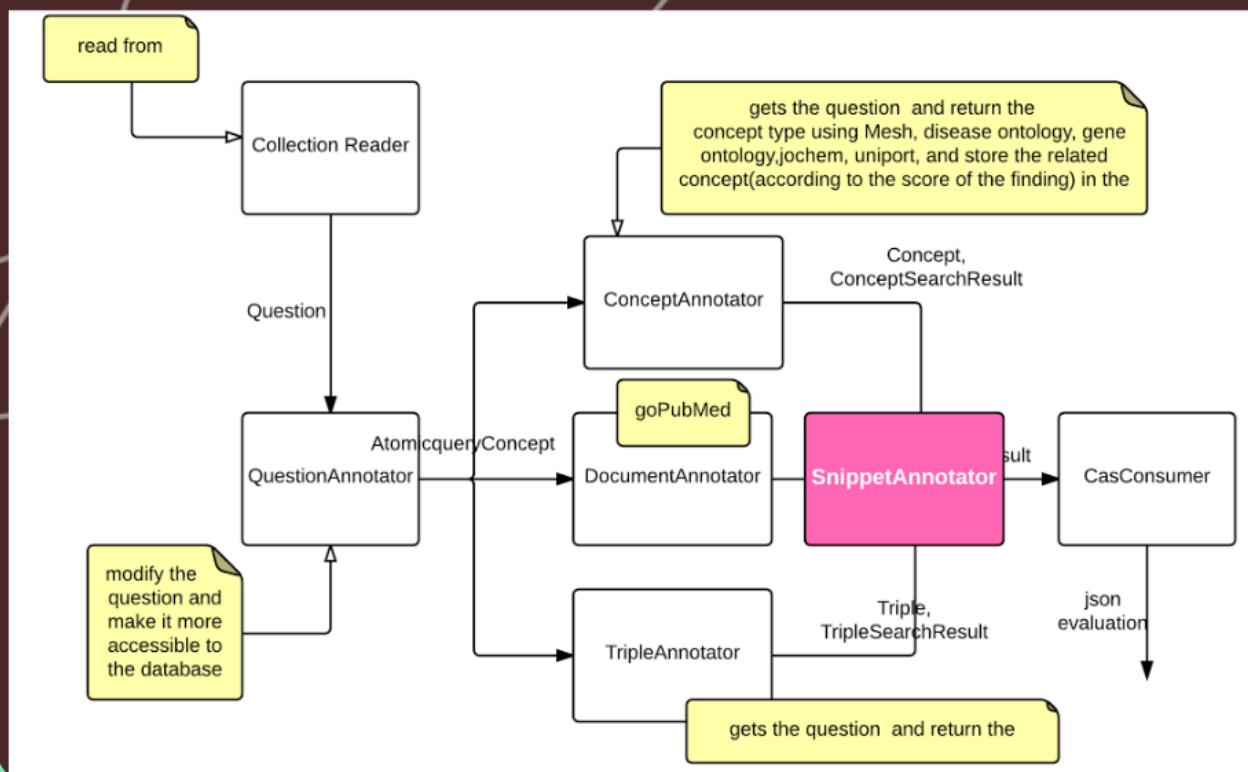
## Enhanced Pipeline Design

- Question Annotation
- Snippet Annotation
- Hierarchical Evaluation

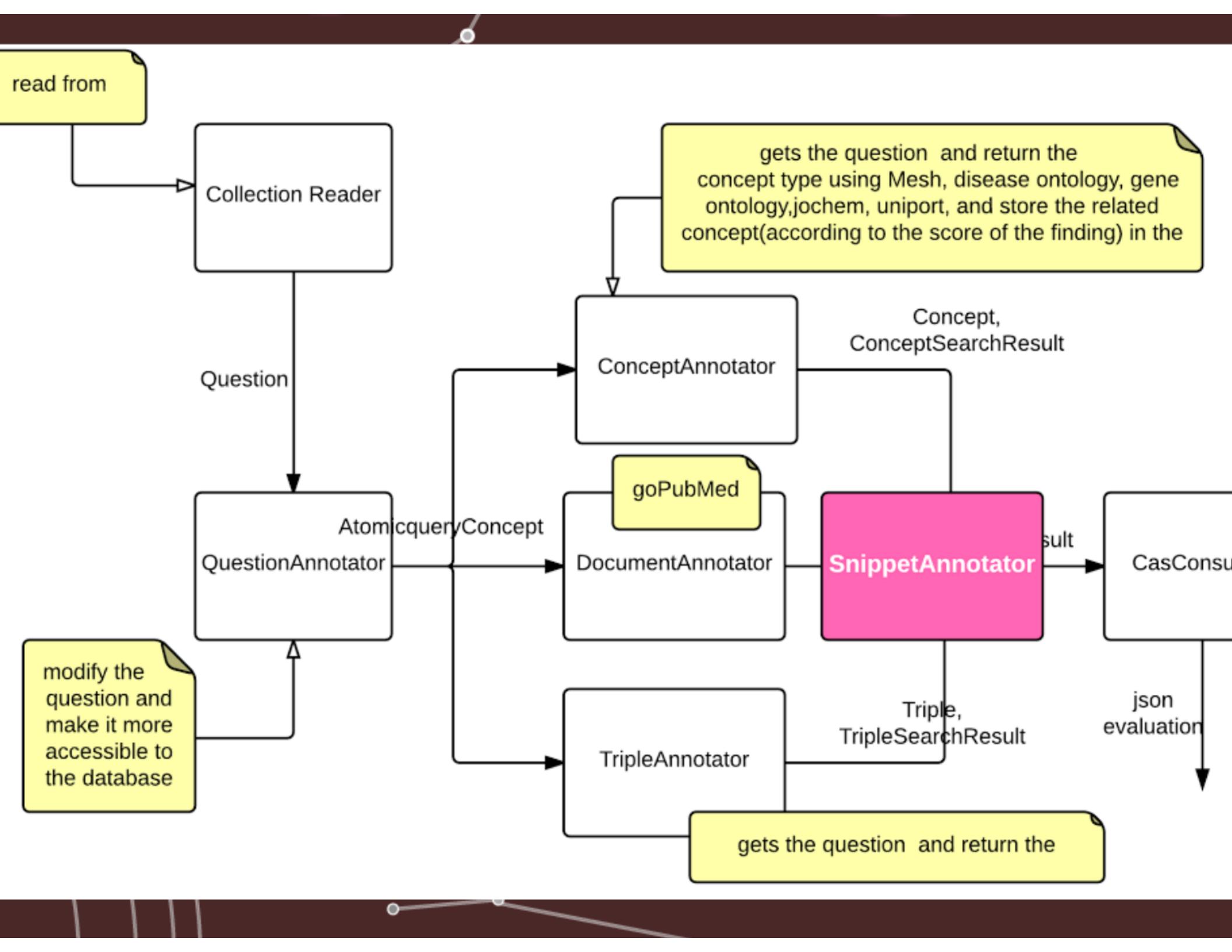
## Updated Documentation

## Studied Collaboration Tools

# Pipeline Design



Evaluator



## **Question Annotation**

### **Extract key terms**

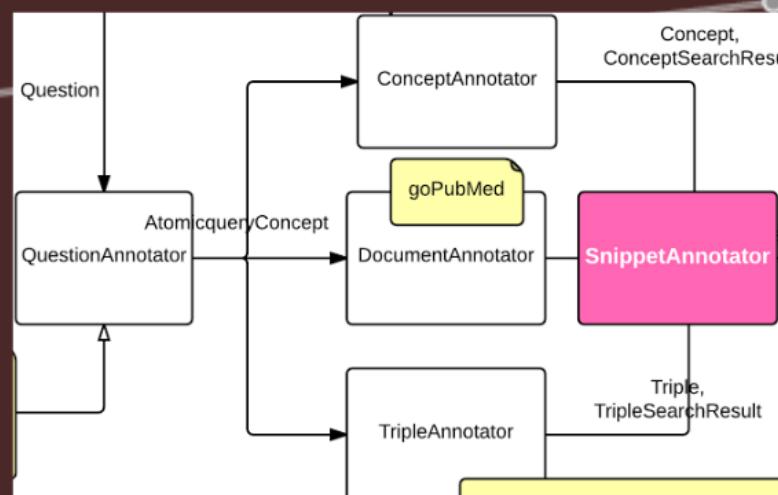
- Word stemming:  
StanfordLemmatizer
- Stop word and punctuation  
removal: Regular Expression
- Tokenization: OpenNLP
- Biological words: Lingpipe

### **Build queries**

- Complex Query Concept with  
Operators

## *Snippet Annotation*

- Get information from JSON sections
- Calculate offset of extractions
- Calculate TFIDF scores



## *Evaluator*

Precision, recall, F-score, MAP, GMAP

Evaluate Performance at Each Stage  
Document, Concept, Triple, Snippet...

Evaluate Overall Performance

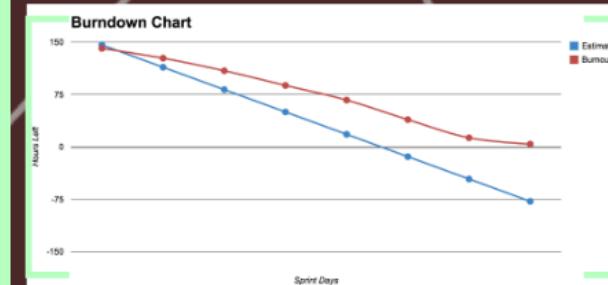
```
Evaluation
-----document-----
doc ans size: 2 | gold ans size:15
AVG Precision: 0.25
positive: 1
-----concept-----
concept ans size: 37 | gold ans size:3
AVG Precision: 0.0
positive: 0
-----triple-----
triple ans size: 100 | gold ans size:0
positive: 0
-----concept result size(in consumer):17
prominent sequence consensus polyadenylation site
```

# *Project Management Overview*

- Burn-down Charts
- Retrospective
- Problems and Solutions

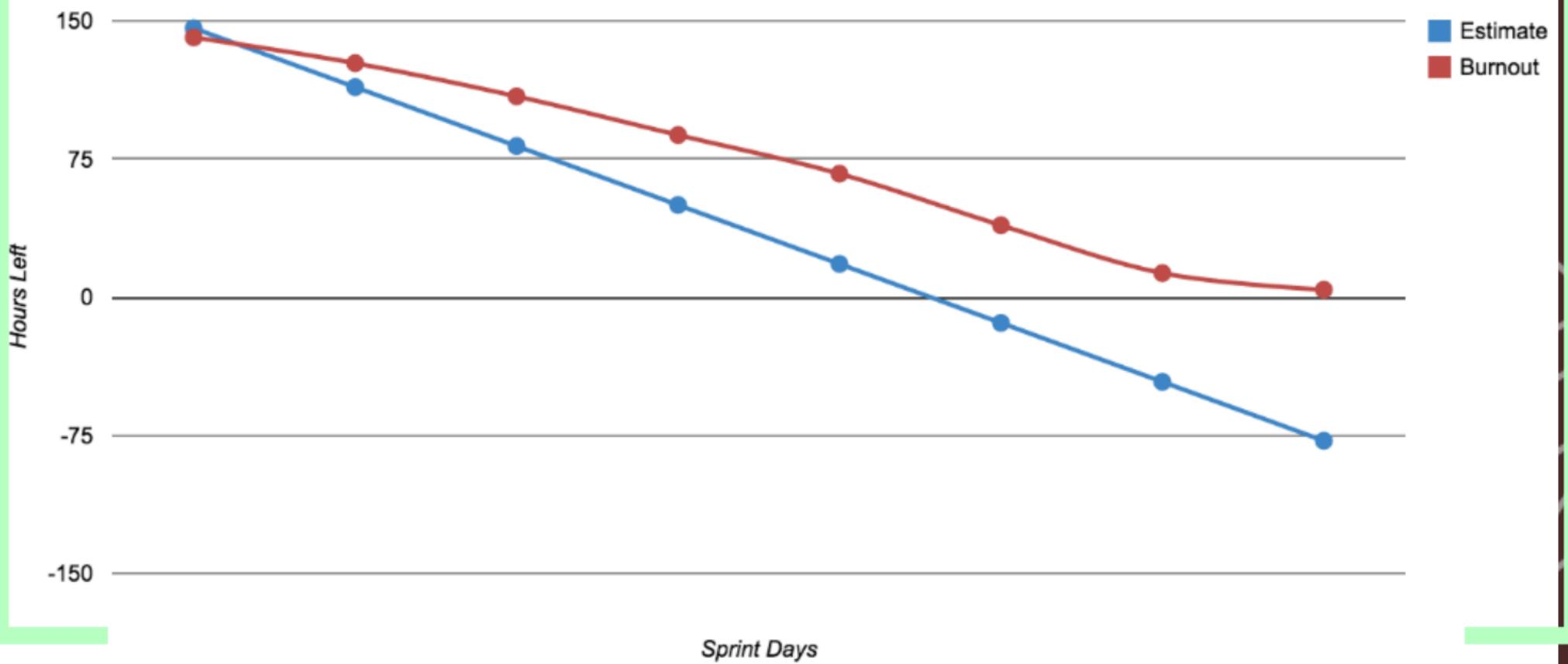
# Group Burn-down Chart

Task	Time (est)	Time (act)	Time (left)	1	2	3	4	5	6	7
Planning	25	26	0	10	8	8				
Team documentation	12	12	0	4	4	4				
Pipeline Design	18	12	6	4	4	4				
Type System Analysis	8	6	2			3	3			
Annotator Design	12	12	0			8	4			
Build Pipeline	8	8	2			4	2			
Question Annotation	8	8	0				4	4		
Document Annotation	8	8	0				4	4		
Concept Annotation	8	8	0				4	4		
Triple Annotation	8	8	0				4	4		
Retrieval Evaluation	12	10	2				6	4		
Question Post-process	8	8	0				4	4		
Answer Extraction	0	0	0							
Answer Evaluation	0	0	0							
Error Analysis	0	0	0							
Hierarchical Analysis	0	0	0							
Performance Enhancement	0	0	0							
System Evaluation	0	0	0							
Documentation	8	12	2	2	2	2	2	2		
Feedback	3	3	0							
<b>TOTAL</b>	<b>146</b>	<b>141</b>	<b>14</b>	<b>32</b>	<b>32</b>	<b>32</b>	<b>32</b>	<b>32</b>	<b>32</b>	<b>0</b>
Daily burnout	-78									



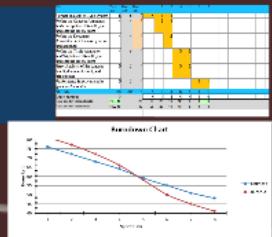


## Burndown Chart

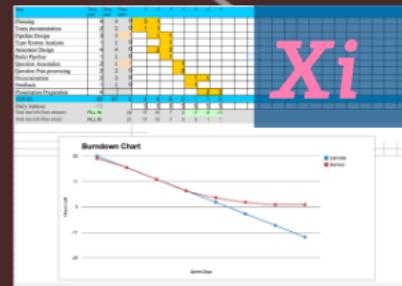


# *Individual Burn-down Charts*

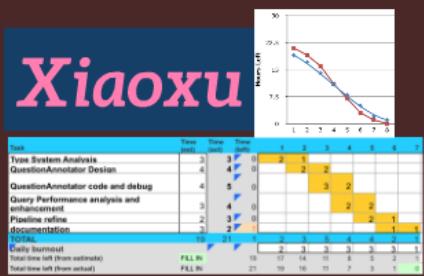
*Lei*



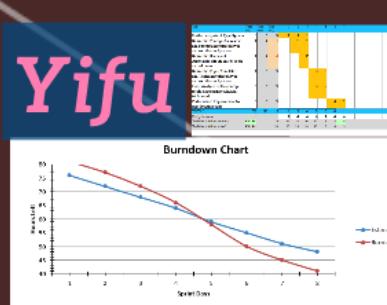
*Xi*



*Xiaoxu*



*Yifu*



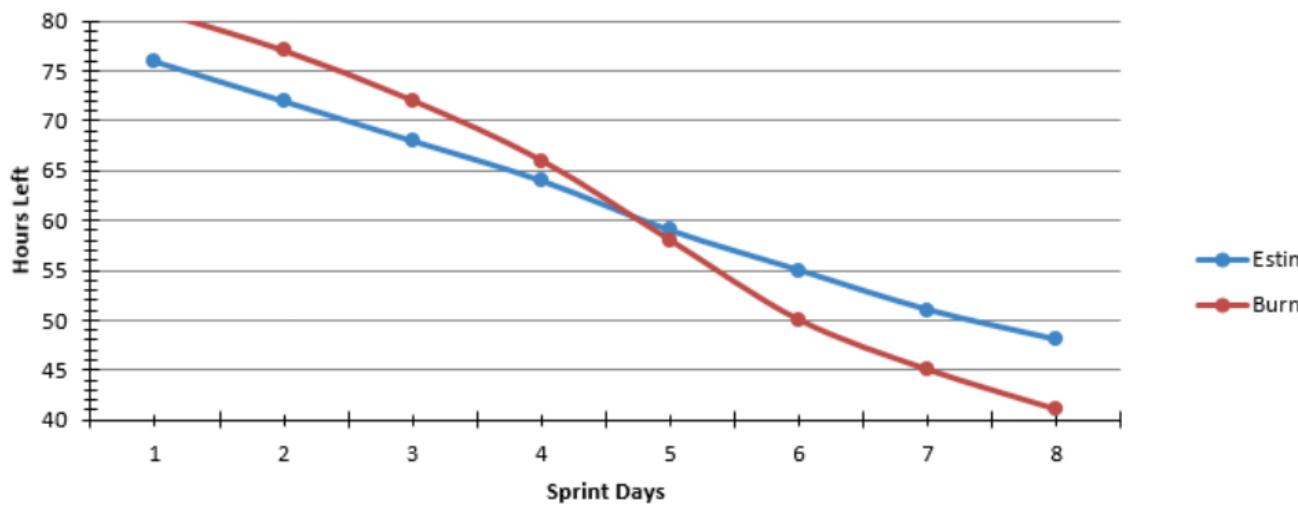
*Yan*



# Lei

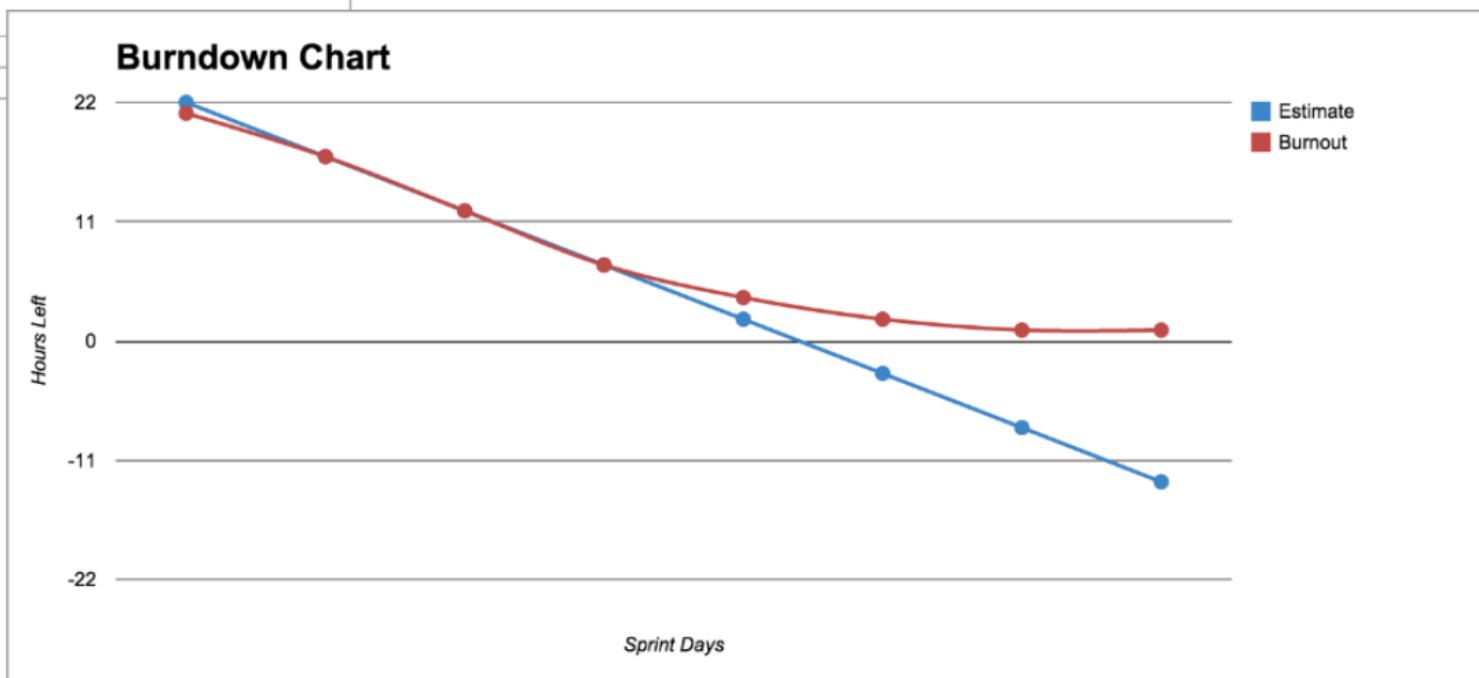
Task	Time (est)	Time (act)	Time (left)	1	2	3	4	5	6	7
Further analysis of Type System	6	6	0	4	1	1				
Refine the Concept Annotator: add conceptSearchResult type and sort the cas by score	4	4	1		2	1				
Refine the Document Annotator: sort the cas by score and add rank	4	4	2			2				
Refine the Triple Annotator: add TripleSearchResult type and sort the cas by score	4	6	0				3	3		
Error Analysis of the concept result, document result, and triple result	5	5	0				3	2		
Performance Improvement by questionAnnotator	5	5	0					4	1	
<b>TOTAL</b>	<b>28</b>	<b>30</b>	<b>3</b>	<b>4</b>	<b>3</b>	<b>4</b>	<b>6</b>	<b>5</b>	<b>4</b>	<b>1</b>
Daily burnout	0			4	4	4	4	4	4	4
Total time left (from estimate)	FILL IN			28	24	20	16	12	8	4
Total time left (from actual)	FILL IN			30	26	23	19	13	8	4

Burndown Chart

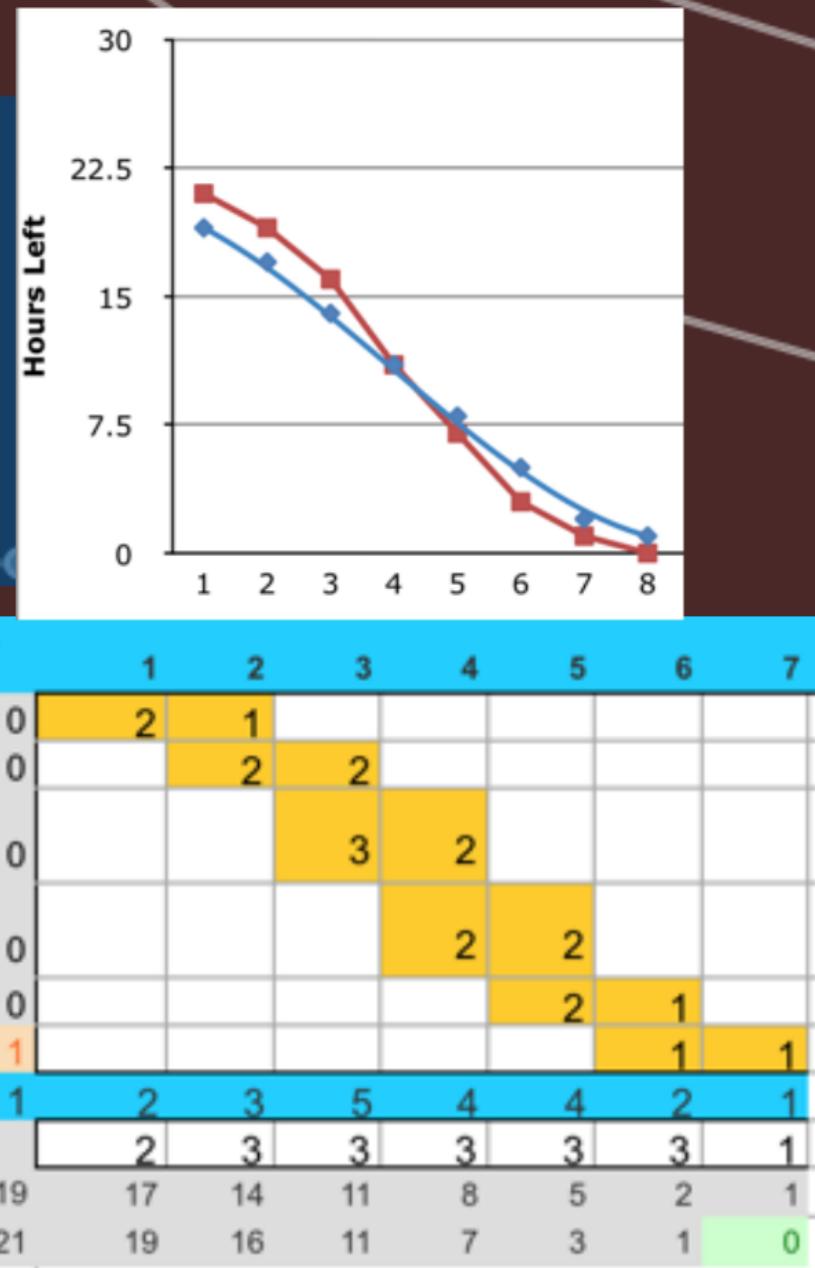


Task	Time (est)	Time (act)	Time (left)	1	2	3	4	5	6	7
Planning	4	4	0	3	1					
Team documentation	2	2	0	1	1					
Pipeline Design	3	3	1		1	1				
Type System Analysis	1	1	0			1				
Annotator Design	4	4	0	2	2					
Build Pipeline	1	1	0		1					
Question Annotation	2	1	1			1				
Question Post-processing	2	2	0		2					
Documentation	2	2	0			1	1			
Feedback	1	1	0		1					
Presentation Preparation	4					2	3			
<b>TOTAL</b>	<b>22</b>	<b>21</b>	<b>2</b>	<b>4</b>	<b>5</b>	<b>5</b>	<b>3</b>	<b>1</b>	<b>1</b>	<b>0</b>
Daily burnout	-13			5	5	5	5	5	5	5
Total time left (from estimate)	FILL IN			22	17	12	7	2	-3	-8
Total time left (from actual)	FILL IN			21	17	12	7	4	2	1

Xi



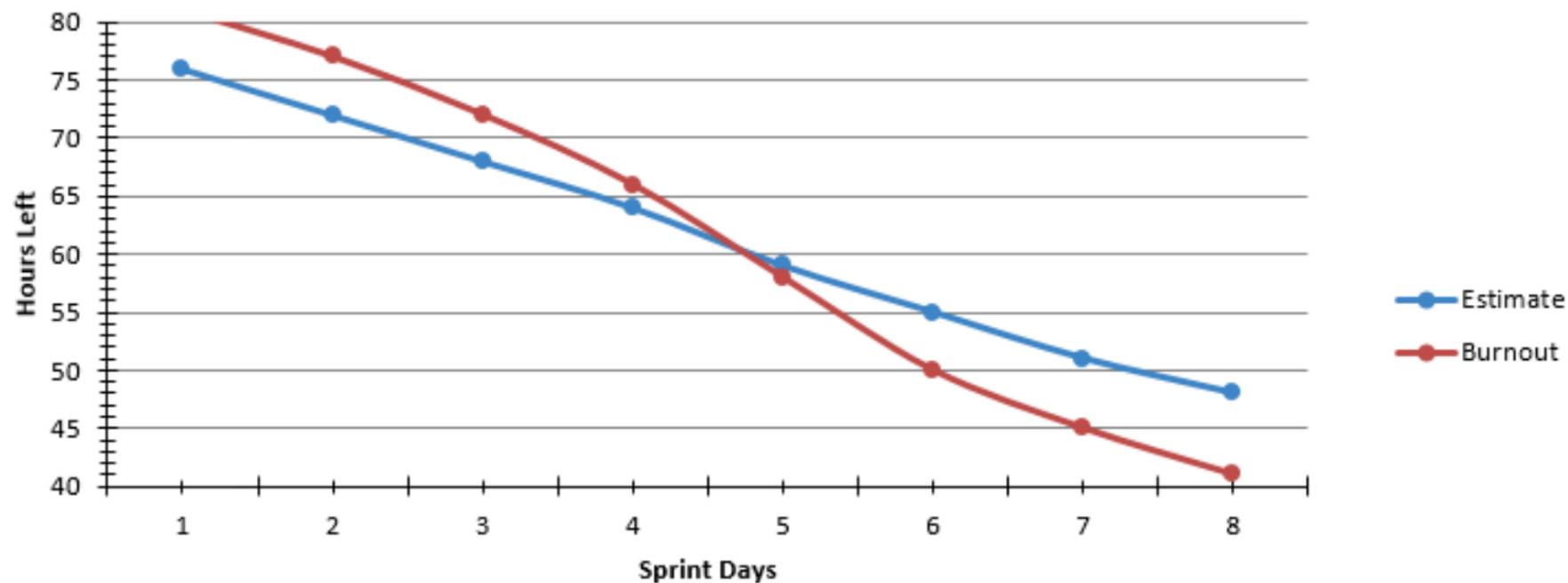
# Xiaoxu



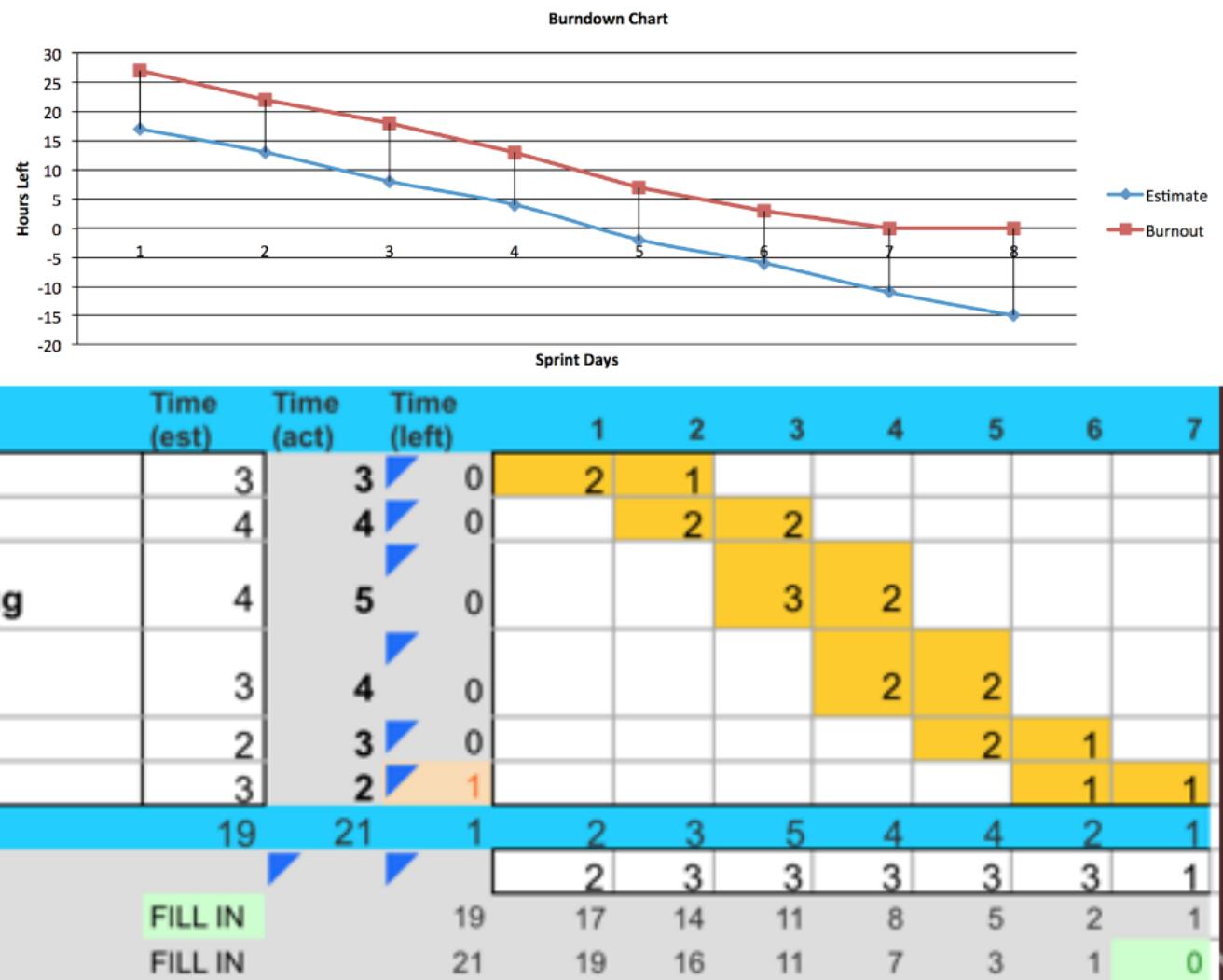
# Yifu

Task	Time (est)	Time (act)	Time (left)	1	2	3	4	5	6	7
Further analysis of Type System	6	6	0	4	1	1				
Refine the Concept Annotator: add conceptSearchResult type and sort the cas by score	4	4	1		2	1				
Refine the Document Annotator: sort the cas by score and add rank	4	4	2			2				
Refine the Triple Annotator: add TripleSearchResult type and sort the cas by score	4	6	0				3	3		
Error Analysis of the concept result, document result, and triple result	5	5	0				3	2		
Performance Improvement by questionAnnotator	5	5	0					4	1	
<b>TOTAL</b>	<b>28</b>	<b>30</b>	<b>3</b>	<b>4</b>	<b>3</b>	<b>4</b>	<b>6</b>	<b>5</b>	<b>4</b>	<b>1</b>
Daily burnout	0			4	4	4	4	4	4	4
Total time left (from estimate)	FILL IN			28	24	20	16	12	8	4
Total time left (from actual)	FILL IN			30	26	23	19	13	8	4

## Burndown Chart



# Yan



# Retrospectives

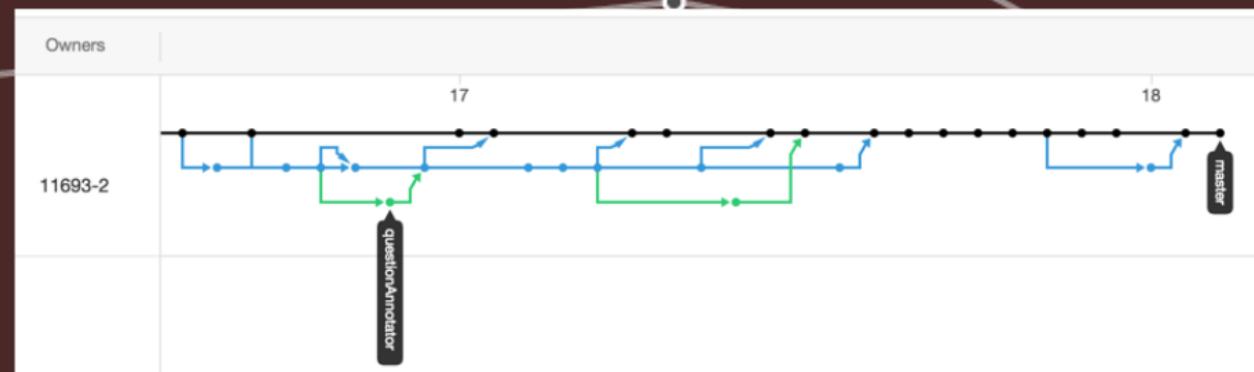
Team Member	Set the Stage (how did you feel this sprint)	What Happened this sprint	Why did those things happen	What To Do to keep up what worked well and prevent what didn't
Xi Liu	Good planning matters	In sprint 0, we drafted the system design, and further discussed the scenarios. We made detailed documentations, and gathered resources and reading materials, which make sprint 1 run smoothly. My teammates are very supportive and have a strong sense of ownership and responsibility. I left for an on-site interview on Wednesday, after having prepared presentations. They knew each part very well and did a great job. We had problems with github and uima environment settings.	Careful planning, thorough research, and detailed documentations. Not familiar with the tools.	Keep maintaining the documentations and logs. Solve the environmental problem by interacting with command lines and revising source codes.
Yifu Wang	Keep on track	Understand the whole architecture and type system. Learn team collaboration on github, build evaluator and update CASconsumer for result analysis. Having some problems about version control on github.	Use a little bit more time on learning part due to limited understanding of Intelligence System and github.	Keep unit testing on every function, Commit project frequently and merge wisely. Make every committed version runnable.
Lei Xiao	Follow the milestone closely	We modified the conceptAnnotator, documentAnnotator, and TripleAnnotator according to the rank of the result. And we modified the question text to improve the performance of the system, such as deleting the stopwords, and the punctuation.	This time we understand the requirement more clearly, and understand the relationship of the type system more detailed and precisely.	Keep analysing the codes and type system, and learn to use the github more precisely.
Yan He	Keep learning and communicating	In the last sprint we tried to understand what should our system do and how are we gonna build it up. Thus, in this sprint, we reconsider some components of the system and try to figure out some new ideas to improve the final performance of the evaluation beforehand in order to avoid us spend much more time to solve the problem later.	Since we have not thought it through about how should we extract the target documents and sentences in the first stage, we realized that if we didn't solve this problem first, the final evaluation performance can not be quite good, therefore we did some refinement about our former ideas and methods.	Keep having several group meetings every week and communicate with each other's problem solving tips and working progress.
Xiaoxu Lu	Progress a bit per day and gain more understanding about what we are going to do	Work on pipeline design refinement, employ more types from given type system to previous work, help develop question annotator. Got a lot of rejected for github, yet not familiar with branch issues, still needs a lot of practice and deep understanding.	Collaborative coding in a project is so important and we kind of wasted a lot of time on dealing with github issues.	Keep employing more useful types in the given type system to fulfill the bioQA requirement.

# Retrospectives

Team Member	Set the Stage (how did you feel this sprint)	What Happened this sprint	Why did those things happen	What To Do to keep up what worked well and prevent what didn't
Xi Liu	Good planning matters	In the sprint, I've drafted the system design, and further discussed the scenarios. We made detailed documentations, and gathered resources and reading materials, which were very useful for our work. My teammates are very supportive and have a strong sense of ownership and responsibility. I left for an on-site interview on Wednesday, after having prepared presentation. They never stopped working and did a great job. We had problems with github and some environment settings.	Careful planning, thorough research, and detailed documentations. Not familiar with the tools.	Keep maintaining the documentations and logs. Solve the environmental problem by interacting with command lines and revising source codes.
Yifu Wang	Keep on track	Understand the whole architecture and type system. Learn team collaboration on github, build a pipeline for evaluation, and CAConsumer for result analysis. Having some problems about version control on github.	Use a little bit more time on learning part due to limited understanding of Intelligence System and github.	Keep unit testing on every function. Commit project frequently and merge wisely. Make every committed version runnable.
Lei Xiao	Follow the milestone closely	We modified the conceptAnotator, documentAnotator, and TripleAnotator according to requirements. After some modification, the generator fail to improve the performance of the system, such as deleting the stopword, and the punctuation.	This time we understand the requirement more clearly, and understand the relationship of the type system more detailed and precisely.	Keep analysing the codes and type system, and learn to using the github more precisely.
Yan He	Keep learning and communicating	In the last sprint we tried to understand what should we do to make sure how to correctly build it up. Thus, in this sprint, we reconsider some components of the system and try to figure out some new ideas to improve the final performance of the evaluation beforehand in pipeline. In this sprint much more time to solve the problem was spent.	Since we have not think it through about how should we extract the target documents and sentences in the first stage, we realized that if we didn't solve this problem first, the final evaluation performance can not be quite good, therefore we did some refine about our former ideas and methods.	Keep having several group meetings every week and communicate with each others' problem solving tips and working progress.
Xiaoxu Lu	Progress a lot per day and gain more understanding about what we are going to do	Work on pipeline design refinement, employ more type from given type system to previous work, help develop question annotator. Got a lot if rejected for github, yet not familiar with branch. Result needs a lot of practice and deep understanding.	Collaborative coding in a project is so important and we kind of wasted a lot of time on dealing with github issues.	Keep employing more useful type in the given type system to fulfill the bioData requirement.

Good Planning  
Continuous Learning  
Deep Understanding  
Consistant Improving  
Efficient Communicating

## *Problems and Solutions*



- Commit project frequently and merge wisely.
- Make every committed version runnable.
- Group meeting and communication on wechat

# Product Backlog

ID	Importance	Story	Notes	How to demo	Points
1.	20	planning	Planning the project timeline and get everyone involved	Documentation work been accomplished	80
2.	20	Team documentation	Document writing about what we are doing and what we are planning to do	Documentation work been accomplished	15
3.	20	Pipeline design	Design the pipeline on UIMA structure based on previous home homework	Documentation work been accomplished	90
4.	40	Type System Analysis	Read and try to understand the type system given from the archetype	Documentation work been accomplished	90
5.	40	Annotator design	Design the annotators needed for process the text form given input, based on the pipeline design	Documentation work been accomplished	95
6.	30	Build pipeline	Build the java class and xml descriptor for the component of the pipeline, reader, consumer and annotators involved	Stdout to test if pipeline runs and get major functions print something during the running phase	95
7.	50	Question Annotation	Buildd the annotator for the question, employ stemming, keywords retrieval or NER to generate better query for given question	Stdout to test if the question annotation works, if query generated reasonable for further information retrieval	95
8.	50	Document Annotation	retrieval of the related document	Stdout to test if the document annotation works, employ gold standard to evaluate the result	95
9.	50	Concept Annotation	retrieval of the related concept	Stdout to test if the concept annotation works, employ gold standard to evaluate the result	95
10.	50	Triple Annotation	retrieval of the related triple	Stdout to test if the triple annotation works, employ gold standard to evaluate the result	95
11.	60	Snippet retrieval	retrieval of the snippet for answer extraction based on what the document, concept and triple has generated	Stdout to test if the snippet retrieval works, employ gold standard to evaluate the result	99
12.	50	Answer extraction	Extract the answer from the snippet for the query we built over the question	Stdout to test if the answer is generated and employ gold standard to test and compare	99
13.	30	Document Evaluation	Evaluation on document answer and golden document answer, calculate performance scores like F-score, MAP	Print and evaluate retrieval result	80

14.	30	Concept Evaluation	Evaluation on concept answer and golden document answer, calculate performance scores like F-score, MAP	Print and evaluate retrieval result	80
15.	30	Triple Evaluation	Evaluate the annotator in the following aspects: response time, precision, recall, MAP, GMAP	Print and evaluate retrieval result	80
16.	50	Snippet Evaluation	Evaluate the annotator in the following aspects: response time, precision, recall, MAP, GMAP.	Print and evaluate retrieval result	80
17.	50	Answer Evaluation	Evaluate the annotator in the following aspects: response time, precision, recall, MAP, GMAP.	Print and evaluate retrieval result	80
18.	30	Error Analysis	Analysis on why some methods can not give us wanted result by looking at the answer structure	Print and evaluate answer information Manual Comparison and experiments	99
19.	30	Hierarchical Analysis	We will evaluate the annotations for: every stage, intermediate retrieval results, final result, and overall performance, trade-offs	Print and evaluate answer information	70
20.	30	Performance enhancement	Enhance the system performance by utilizing NLP API, building better queries, and experimenting answer selecting strategies	Print and evaluate answer information	70
21.	20	System evaluation	Evaluate the system in the following aspects: response time, precision, recall, MAP, GMAP. We will evaluate the annotations for: every stage, intermediate retrieval results, final result, and overall performance, trade-offs	Print evaluation information	70

# Story Wall

Backlog	Sprint	In Development	Ready for Testing	Done
<p>Answer Extraction</p> <p>Answer Evaluation</p> <p>Hierarchical Analysis</p> <p>Team Documentation</p> <p>Documentation</p> <p>Feedback</p>	<p>System Evaluation</p>	<p>Snippet Retrieve</p>	<p>Error Analysis</p> <p>Type System Analysis</p> <p>Performance Enhancement</p>	<p>Result Evaluation</p> <p>Planing</p> <p>Pipeline Design</p> <p>Concept Annotation</p> <p>Annotator Design</p> <p>Question Annotation</p> <p>Triple Annotation</p> <p>Build Pipeline</p> <p>Document Annotation</p>

# Sprint 1 Presentation

Team 2: Xi LIU, Lei XIAO, Yifu WANG, Xiaoxu LU, Yan HE

## Project Management Overview

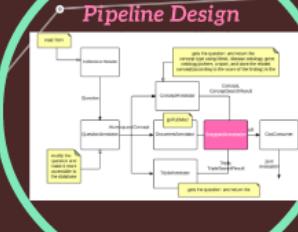
- Burn-down Charts
- Retrospective
- Problems and Solutions

## Project Overview

- **Vision:** Solve yes/no questions with high precision
- **Requirement Analysis:** information flow and evaluation approaches
- **Pipeline Draft**

## Task Checklist

- Read questions
- Convert questions to terms
- Build queries with terms
- Retrieve query results
- Extract snippets from returned documents
- Rank the result by relevance
- Evaluate the overall performance and hierarchical performance



## Pipeline Design



## Achievements

- Enhanced Pipeline Design
  - Question Annotation
  - Snippet Annotation
  - Hierarchical Evaluation
- Updated Documentation
- Studied Collaboration Tools