Baseball-AI
Winning Rate Computing

Ji Am Chung*, Young Jae Byun†, Seung Myeon Park‡, Jun Jeon§
*Department of Information System
Hanyang University
Seoul, Korea
Email: hummernk@gmail.com
†Department of Information System
Hanyang University
Seoul, Korea
Email: qusdudwowo@gmail.com
‡Department of Information System
Hanyang University
Seoul, Korea
Email: antimoto@nate.com
§Department of Information System
Hanyang University
Seoul, Korea
Email: jeonjun2@gmail.com

***Abstract*— Today baseball is becoming more and more popular and getting new fans. Those who just have put their first step into baseball - find it hard to follow the rules and to understand the changes in the situation. Unlike our baseball league KBO, MLB already has many meaningful methods for in-game analyzing with real time statistics. So we are going to adopt those ideas into KBO league. Winning Probability Added, which is also known as WPA, is one of good factors which allows easy understanding of the situation and the impacts that pitcher or batter makes each time. Our project aims to show Winning Rate and WPA factor in real-time, with the algorithm that fits KBO situation.**

1. Introduction

Baseball, even though the myriad scandals, attracts more and more fans and is becoming more nationwide sports. There is a major problem that, though baseball fans are inflowing, its hard for beginners to understand the rules and catch situation of the match. Though baseball is much more a number-oriented sport than any others such as soccer or basketball, the classical stats do not fully reflect the match stream and evaluate the value of the players properly. However, in MLB, a lot of sabermetricians have already quantified vague situations and values in numerical way, and KBO tends to follow it. (i.e Babip, OPS, etc..). Korean baseball websites nowadays use the same WPA algorithm which had been used in earlier days of MLB, so the differences between two leagues are not considered. We wanted to solve this problem by creating our own WPA algorithm based on KBO, which not only calculates the stats suitable for KBO, but also analyzes winning possibilities of each team and the impacts each players make. We also focus on helping beginners to understand baseball. We will make KBO winning rate DB and we are going to put that into our program. The program represents ongoing situation of the match and each teams winning possibilities simultaneously. Also, by showing quantified stats such as batters WPA or pitchers WAR , it shows how much powerful the player is(or has been) throughout the game, seasons or his entire career. Whether baseball match is underway or not, the program will show player rank categorized by players, teams, or positions according to users request. It is not just Baseball statistic calculator, but also somewhat like websites such as Statiz or KBReport. The software indicates different types of statistical analysis, and shows them in visualized ways. We thought that current WPA algorithm used in korean web sites does not fit korean baseball situation. That is why we have started this project.

This project is composed of 4 steps.

1) Crawl the data of last 10 seasons of KBO from baseball statistics websites.
2) Compute the data to make some statistics.
3) Apply the statistics to WPA algorithm.
4) Real-time data capturing and showing the winning rate and WPA.

May 2, 2016

2. Requirements

2.1. Data Handling

2.1.1. Crawling.

- Get every single raw data of KBO from baseball webpage to construct the root database.
- Crawiling source : http://www.koreabaseball.com

2.1.2. Capturing.

- Get real-time data when the match is underway.

2.1.3. Real-Time Mirroring.

- Program should immediately renew the database according to the result of the match.

2.1.4. Computing.

- Calculate the numerical data to make some meaningful statistics.
- Every single data has different weight. e.g.) Hits at 1st inning have different value from those at 9th.
- Calculate numerical values including WPA.

2.1.5. Data Storage.

- Save every single stats data.
- Divide players into two tables. One table is for players who is in active service, the other for retired.
- Table for players who is in active service needs to be updated constantly, and the other table doesnt.
- User who wants to conceal the program from screen can do that by clicking window minimization button.

2.2. Function

2.2.1. EXCEL Compatibility.

- User can export datas of specific player or stats to MS Excel files.
- User can import fixed form of MS Excel file of specific game result to compute changes of KBO algorithm winning rate shown as image file which can also exported as jpeg, gif, png, or bmp.
- User can import fixed form of MS Excel file of specific league(fantasy or amatuer) data to compute WAR stats. This data can be exported as EXCEL file.

2.2.2. On-Board Posting(abandoned).

- Someone who wants to post any idea or thoughts can share what they have.
- Make another Q&A board so as to hel beginners solve their curiosity.

2.2.3. Board Log-in & Sign-out(abandoned).

- Log-in to or Sign-out from Board.
- User who logged-in the board can upload their post or reply to other users

2.2.4. Stats Visualization.

- Show current state of game in a table.
- Show current winning average of each team
- Show current WPA stats of players
- Show player's photograph

2.3. User Interface

2.3.1. Window Minimization & Window Maximization.

- User who wants to see the program widely can do that by clicking window maximization button

2.3.2. Program Turn On & Turn Off.

- User can turn on the program by clicking desktop icon
- If user try to power on the program even if that is already turned on, terminate existing program and launch the program again
- User can turn off the program by clicking x button at the top-right corner of the program

2.3.3. Mouse Click Event.

- Provide user with three options [To Home, Window Minimization, Termination] when user right-click any area within program.

2.3.4. Player Stat Pop-Up.

- When user clicks certain player, program shows his profile by generating a new pop-up
- If player is a pitcher, pop up list of the first string who has not been on the match yet
- If player is taking the field, pop up his profile as batter
- Pitcher pop-up profile stats list : ERA(Earned Run Average) for applicable season, WPA, WAR, WHIP for last 5 matches, (KBB 9), hyperlink connected to NAVER article about him
- Batter pop-up profile stats list : BA(Batting Average), WAR, WPA, OPS for last 5 matches, BABIP for applicable season, hyperlink connected to NAVER article about him
- The number of pop-up cannot be over two

2.3.5. Player Ranking.

- Sort players by team, position, date and game with WPA stats

2.3.6. Data Searching.

- Searching option constitutes of match schedule, player and stats and player
- If option match schedule is chosen, program shows match schedule as a calendar
- If user click one of date, there are three cases. First one is past match, so program shows match log.

Second one is on-going match, so program directs user to the match. And the last one is coming match, so program shows every details of the match including players, referees, park, appointed first thrower, weather forecast
- If option player and stat is chosen, program shows the applicable stats by entire players, team, position, monthly separately.
- If option player is chosen, program shows every single stat of applicable player

2.3.7. Get Information real-time.

- User can choose the way one gets some information(pop-up or push window)
- Pop-up is a kind of window, so when user have it on the screen, one cannot click main program
- Information could be as follows
- Agreed Decision : User can get information about agreed decision and its details
- Cancellation in case of rain : User can get information when the match is cancelled in case of rain by getting a pop-up or push window
- Player Substitutution : In case of substituting player, User can get information why the player was substituted with other, and information about that other player
- When option is pop-up, user can have additional function which is Multi-View. By doing so, user can watch several matches simultaneously
- If there have multiple pop-ups, eliminate pop-up windows sequentially after checking them

2.3.8. Error.

- Error alert

3. Development Environment

3.1. Choice of Software Development Platform

3.1.1. Which platform and why? (e.g. , Windows, Linux, Web, or etc.).

- We adopted Windows, because there are some merit when we choose Windows. First of all, the percentage of all Windows user is almost 85%. Since we want to emphasize on majority, we chose Windows. Second, because of encoding compatibility. For more convenience, its much better to share encoding method in OS, web server, database.

3.1.2. Which programming language and why?.

- We used both python and Java. At first we used only python 2.7 to crawl html source, to parse tag data and manipulate date with python DB driver. But pyhon 2.7 had problem of unicode encoding. Because the web site that we wanted to crawl was encoded with 'utf-8', but python doesn't support it. So we have to think another IDE. Two candidates was python 3.5, who support 'utf-8' and Java with a friendly programming language. With a contempletation, we decided that crawl, parsing and DB manipulation by using python 3.5 and the others, GUI, etc. by Java. That's because we had not enough time to learn a new Jave crawling, parsing code, but only to have time to transplant python 2.7 code into python 3.5. And we were not sure whether python 3.5 is an effective tool for GUI making, so we adopted Java IDE Netbeans which support easy GUI method.

3.1.3. Provide a cost estimation for your built. (including any purchase of software/hardware).

- human labor - 0
  Our group constitute of four members who all do this group term project voluntarily. So human labor cost is zero.
- software cost - 0
  Our group will make program with open source API, which costs zero for academic purpose.
- hardware cost - 0
  Our group uses our existing laptop to simulate or implement our program. There is no need to buy other hardware.

3.1.4. Provide clear information of your development environment. (e.g., version of software, OS version, your computer resources).

- OS
  Windows 10 pro(10586.218 build)
- Language Set
  Korean
- Computer Mode l
  MSI GT60
- Processor
  Intel(R) Core(TM) i7-3600QM CPU @ 2.4GHz
- Main Memory
  8GB RAM
- Internet Connection
  IPTIME WiFi

3.2. Software in use

3.2.1. Any existing software or algorithm in use? (doing a similar task as your proposal; provide a proper reference if there is any).

- There is no similar software computing WPA, after Tom Tango introduced concepted of WPA. But, there is several websites or spreadsheet providing each statement winning rate. For MLB, there is a site called 'The Book'. It is made by Tom Tango who designed many good index including WPA. 'Statiz',

'KBReport' are sites providing many baseball stats for KBO.

Statiz

Figure 1. Statiz

KBReport

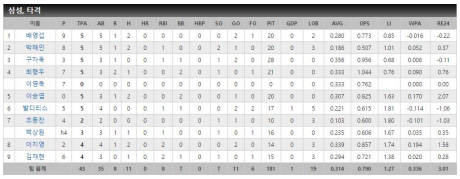

Figure 2. KBReport

TheBook

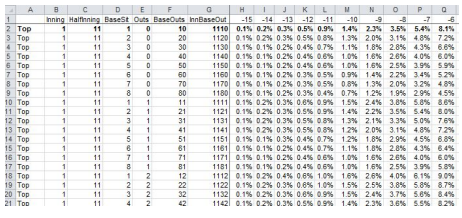

Figure 3. TheBook

3.3. Task distribution (If you want, you can provide this later at the next phase - design)

- Which member is responsible for what?

Roles	name	task description
User	Jun Jeon	test some modules
Customer	Seung Myun Park	return feedback
Software developer	Y oung Jae Byun	develop modules
Development manager	Ji Am Chung	organize architecture

4. Specification

4.1. Data Handling

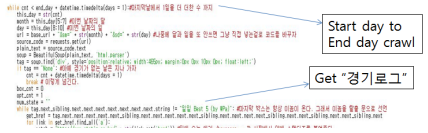

Figure 4. Crawling source

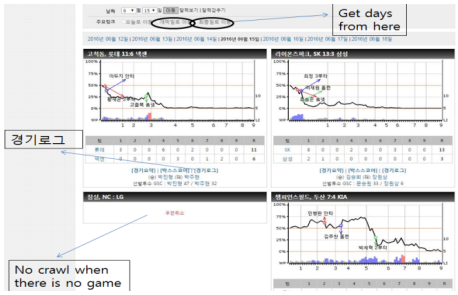

Figure 5. Parsing source

4.1.1. Crawling.

- Users can set the start date and end day that includes matches they want to crawl.
- The crawler gets data from baseball webpage to construct the database.
- Data means game box score which contains every statement in actual game.
- It cralws form the homepage: http://www.koreabaseball.com/Schedule/Game/BoxScore.aspx?lean1 of leagueId and n2 of SeriesID are arbitrary integer number, yyyymmddt1t2x of gameId is year, month, day, team1, team2, I digit ordinal number, yyyy of gyear is year.

Figure 6. crawling

4.1.2. Capturing.

- Get real-time data when match is underway.

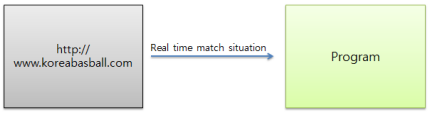

Figure 7. capturing

4.1.3. Real-Time Mirroring.

- Program should immediately renew the database according to the result of the match.

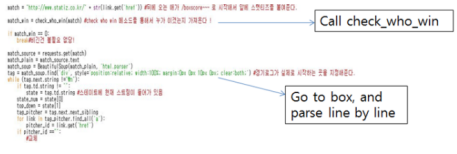

Figure 8. Crawling Real-time Code

Figure 9. Parsing Real-time Code

Figure 10. Crawling Real-time Source

- After the data is crawled, crawler checks who is winning at the specific time, ans parse the data line-by line and puts the value into the program.
- After the data is put on the program, it shows current situation of the match real-time.

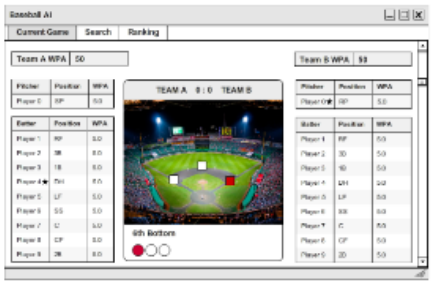

Figure 11. Real-time GUI

- If the base is on with a batter, it filled with color red and if the base is not on, it filled with color white

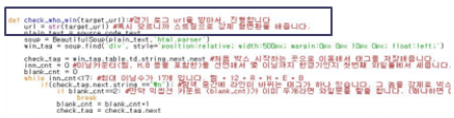

Figure 12. Crawling who won code1

Figure 13. Crawling who won code2

Figure 14. Crawling who won that game source

- When the match ends, real-time crawler gets the information of which team won the game, and ends with setting the value 1 to won team and 0 to lost team.

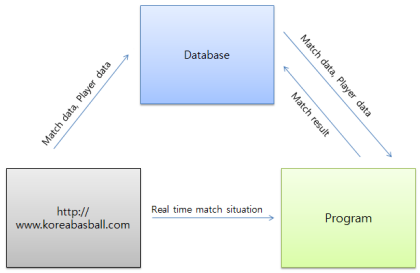

Figure 15. real-time mirroring

4.1.4. Computing.

- We can calculate Winning rate as follows.

$$Winrate = \frac{AllMatches \cap Statement \cap Win}{AllMatches \cap Statement} \quad (1)$$

4.1.5. Data Storage.

- Save every single stats data.
- Divide players into two tables. One table is for players who is in active service, the other for retired.
- Table for players who is in active service needs to be updated constantly, and the other table doesnt.

4.2. Functions

4.2.1. EXCEL Compatibilty.

- User can export data in the program that she/he wants.
- playerorstats.xlsl is an example of excel file daeling with datas of specific player or stats.
- gameres.jpeg is an example of image file that can be exported from a fixed form of MS Excel file of specific game result, which computes changes of KBO algorithm winning rate.

	A	B	C	D
1	시합일	2015.04.10		홈팀
2	구장	고양 대표팀 훈련 구장		어웨이팀
3				
4			1	2
5		넥센	0	0
6		NC	0	1
7				
8	넥센 타자 기록			
9				1
10		합계		
11		二	서건창	2땅
12		좌	고종욱	4구
13		지	이택근	우안
14		우	대니돈	중비
15		三	김민성	삼진
16		一	채태인	
17		포	박동원	
18		유	김하성	
19		중	임병욱	
20		교	홍성갑	
21		교	유재신	
22	NC 타자 기록			
23				1
24		합계		
25		좌	김준완	1땅
26		중	이종욱	3비
27		우	나성범	삼진

Figure 16. excel input file

	A	B	C	D
1	ID	16-001-1		
2	Inning	Attacking Team	Out Count	Statement
3	1	Away	0	1
4	1	Away	1	1
5	1	Away	1	2
6	1	Away	2	3
7	1	Away	2	2
8	1	Home	0	1
9	1	Home	1	1
10	1	Home	2	1

Figure 17. excel output file

- leaguedata.xlsl is an example of fixed form of MS Excel file of specific league(fantasy or amatuer) data to compute WAR stats.

4.2.2. On-Board Posting(X).

- Someone who wants to post any idea or thoughts can share their idea.
- Make another QA board so as to help beginners solve their curiosity.

4.2.3. Board Log-in & Sign-out(X).

- Log-in to or Sign-out from Board.
- User who logged-in the board can upload their post or reply to other post.

4.2.4. Stats Visualization.

- cur_state.myd is an example of mysql data file showing current state of game in a table form.
- By Generating a new pop-up, program can show user current winning average of each team.
- By Generating a new pop-up, program can show user current WPA stats of players.
- By Generating a new pop-up, program can show player's photograph.

4.3. User Interface

4.3.1. Window Minimization Window Maximization.

- Window default size : 960*540
- User who wants to conceal the program from the screen can do that by clicking window minimization button.
- User who wants to see the program widely can do that by clicking window maximization button.
- User can turn off the program by clicking x button at the top-right corner of the program.

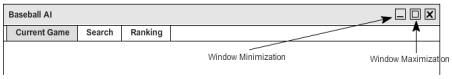

Figure 18. Windows Minim.&Maxim.

4.3.2. Program Turn On & Turn Off.

- User can turn on the program by clicking desktop icon.

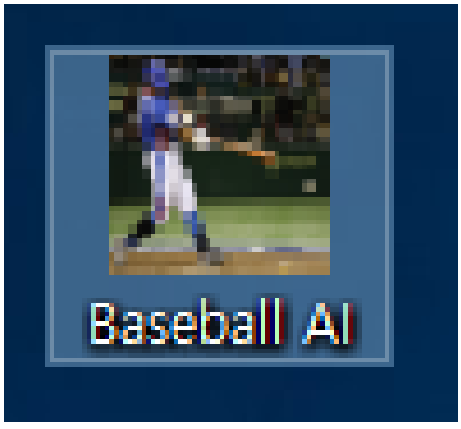

Figure 19. Program Turn On

- If user try to power on the program even if that is already turned on, terminate existing program and launch program again.
- User can turn off the program by clicking x button at the top-right corner of the program.

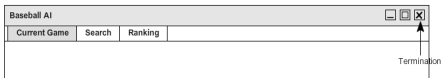

Figure 20. Program Turn Off

4.3.3. Mouse Click Event.

- Provide user with three options [To Home, Window Minimization, Termination] when user right-click any area within program.

4.3.4. Player Stat pop-up.

- When user clicks certain player, program shows his profile by generating a new pop-up.
- If player is a pitcher, pop up list of the first string who has not been on the match yet.
- If player is taking the field, pop up his profile as batter.
- Pitcher pop-up profile stats list : ERA(Earned Run Average) for applicable season, WPA, WAR, WHIP for last 5 matches, (K/BB 9), hyperlink connected to NAVER article about him.
- Batter pop-up profile stats list : BA(Batting Average), WAR, WPA, OPS for last 5 matches, BABIP for applicable season, hyperlink connected to NAVER article about him.
- The number of pop-up cannot be over two.

4.3.5. Current Game.

- Star mark indicates that pitcher(player 4) is on the mound and the batter is at bat.
- Squares on the field indicate each base. If runner is on the base, then it is colored with red.
- Circles at the bottom of the field indicate out count. Red is counted and blank is not.

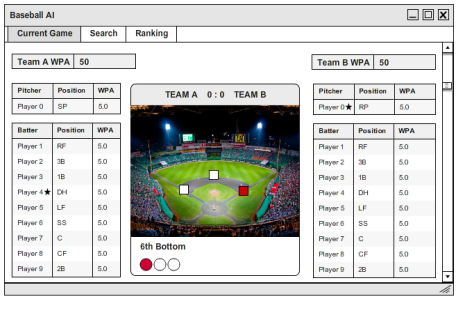

Figure 21. Current Game

4.3.6. Player Ranking.

- Sort players by team, position with WPA stats.

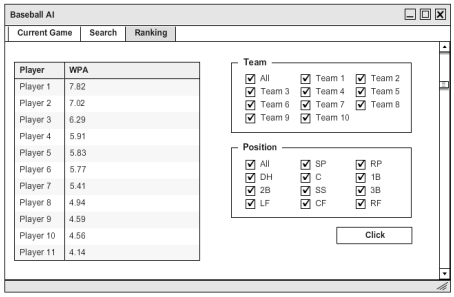

Figure 22. Player Ranking

4.3.7. Data Searching.

- Search option constitutes of Search Match/ Search Player

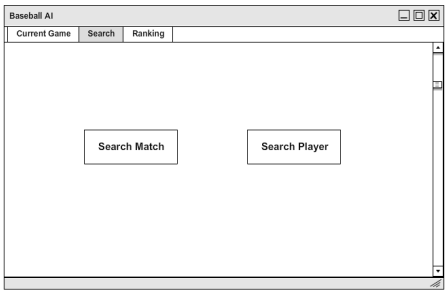

Figure 23. Search Option

- If option Search Match is chosen, program shows match schedule as a calendar.

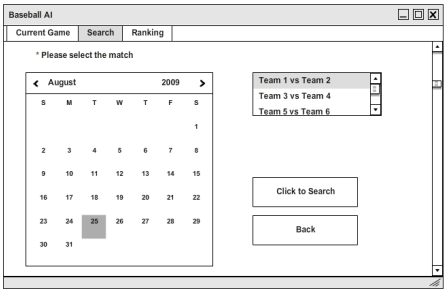

Figure 24. Search-match option

- The result is shown as follows.

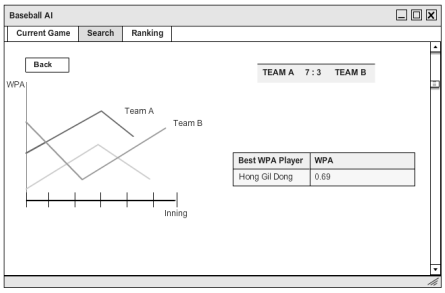

Figure 25. Search match result

- If option Search Player is chosen, input box will appear. It requires user to enter a players name.

Figure 26. Search Player

- The result is shown as follows.

Figure 27. Search Player Result

4.3.8. Get Information Real-Time.

- User can choose the way one gets some information(pop-up or push window).
- Pop-up is a kind of window, so when user have it on the screen, one cannot click main program.
- Information could be as follows.
- Agreed Decision : User can get information about agreed decision and its details.
- Cancellation in case of rain : User can get information when the match is cancelled in case of rain by getting a pop-up or push window.
- Player Substitutition : In case of substituting player, User can get information why the player was substituted with other, and information about that other player.
- When option is pop-up, user can have additional function which is multi-view. By doing so, user can watch several matches simultaneously.
- If there has multiple pop-ups, eliminate pop-up windows sequentially after checking them.
- Refresh in every 30 seconds.

4.3.9. Error.

- If program does not receive signal for 2 minutes, error message pops up with alert sound(windows alert sound).

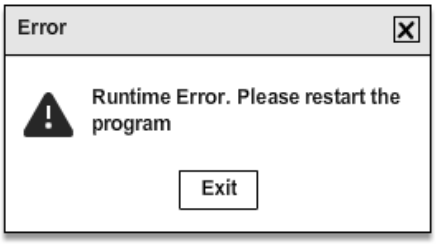

Figure 28. Error Notification

4.3.10. Database.

- We need four data tables, and each every table refers to each other.
- We also need one Season table.

	A	B	C	D
1	ID	16-001-1		
2	Inning	Attacking Team	Out Count	Statement
3	1	Away	0	1
4	1	Away	1	1
5	1	Away	1	2
6	1	Away	2	3
7	1	Away	2	2
8	1	Home	0	1
9	1	Home	1	1
10	1	Home	2	1

Figure 29. the first half match table

E	F	G	H	I
			Statement	
Score Gap	Win Rate		1	Bases Empty
0	50		2	Runner on First
0	52.2		3	Runner on Second
0	49.7		4	Runners on First and Second
0	51.5		5	Runner on Third
1-	42.5		6	Runners on First and Third
1-	44.5		7	Runners on Second and Third
1-	42.1		8	Bases Loaded
1-	40.4			

Figure 30. the second half match table

	A	B	C	D	E
1	15Season	Inning	Attacking Team	Out Count	Statement
2		1	Away	0	1
3		1	Away	0	2
4		1	Away	0	3
5		1	Away	0	4
6		1	Away	0	5
7		1	Away	0	6
8		1	Away	0	7
9		1	Away	0	8
10		1	Away	1	1
11		1	Away	1	2
12		1	Away	1	3
13		1	Away	1	4

Figure 31. the first half season statement table

F	G	H	I	J	K	L	M	N	O	P
5++	4+	3+	2+	1+	0	1-	2-	3-	4-	5--
					50					

Figure 32. the second half season statement table

	A	B	C	D	E
1	PID	Name	Birth date		
2	11025	서건창	1989.08.22	16Season	1.01
3			1989.08.22	15Season	2.06
4		유희관	1986.06.01	14Season	1.05

Figure 33. player table

5. Architecture Design & Implementation(partial)

5.1. Overall architecture

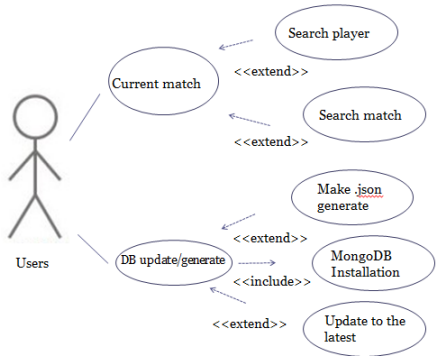

Figure 34. Easy Architecture

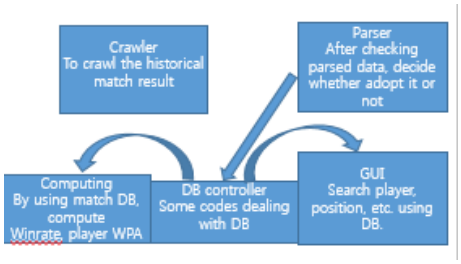

Figure 35. ModuleRelation

- Crawler
  crawl the historical match result
  Class1 : pastSpider

- Parser
  After checking parsed data, decide whether adopt it or not.
  Class1 : game_checker
  Class2 : html_parse

- Computing
  By using match DB, compute Winrate, player WPA.
  Class1 : make_winrate
  Class2 : make_player_wpa
  Class3 : current_game

- DB controller
  Some codes dealing with DB
  Embedded in crawling code. no separate class made.

- GUI Search player, position, etc. using DB.
  Class1 : gui
  Class2 : tabbed
  Class3 : search_p

Class4 : resultplayerpanel
Class5 : search_m
Class6 : dbgen

5.2. Directory organization

Directory	File names	Module names in use	Etc.
./db/	match_log	computing	
./db/	player_wpa	GUI	
./db/	winrate	GUI	
./db/	current_match	GUI	
./	baseball_AI	main module	

5.3. module 1

Crawling :
Crawl raw box data to make past match log.
class1 pastSpider : crawl every single KBO website till when DB is made initially.
And make match_log in db updateSpider : crawl KBO website only for needs to be updated.
(if last db was ver. 20160528, and today is 20160531, update for 29,30 for automatically)

5.4. module 2

Parsing :
Parse the crawled data to make some useful information.
class1 game_checker : parse crawled data to check it is valid website or not.
class2 html_parse : after game_checker adopt specific website, leaving only meaningful values to make DB ”match_log”

5.5. module 3

Computing :
Compute winrate and WPA from match log, by changing box data to statement data.
class1 make_winrate : read box data in ”match_log” and get statement data and save into ”winrate”
class2 make_player_wpa : read ”match_log” and compute the plus and minus value of winrate and save wpa values into ”player_wpa”
class3 current_game : using box data crawled by current_spider make statement data to save into ”current_game”

5.6. module 4

DB manipulator :
not decided yet clearly.

5.7. module 5

GUI :
class1 gui : Root frame that all of the contents are integrated.in
class2 Tabbed : basic format of panel which is used to declare each tab.in frame.
class3 search_p : action which occurs when Search Player button is clicked.
class4 resultplayerpanel : panel which pops up to show the result of ’Search Player’
class5 search_m : action which occurs when Search Match button is clicked.
class6 dbgen : generate or update database when such button is clicked.

6. Use Cases

6.1. Current Match

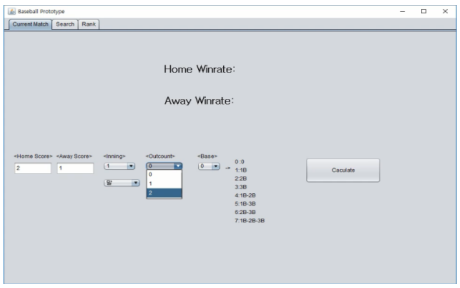

Figure 36. Winrate calculator based on statement

- You can choose the current situation of the match. Ex) score, inning, outcount, base
- If you finish putting the value and push on the calculate tab, the home winrate and away winrate is calculated and showed on the program.
- It is based on the latest database you updated or the existing one.

6.2. Player Search

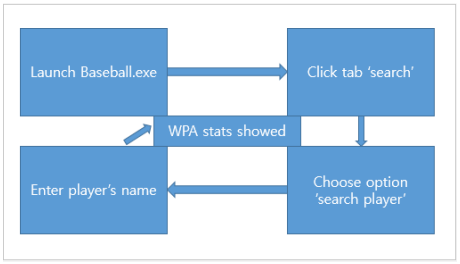

Figure 37. Player Search-flow chart

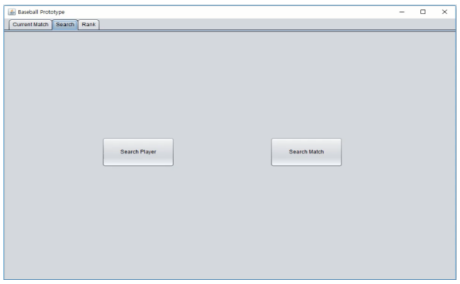

Figure 38. Search button

- If Users choose the search tap, User can choose between Search Player and Search Match.

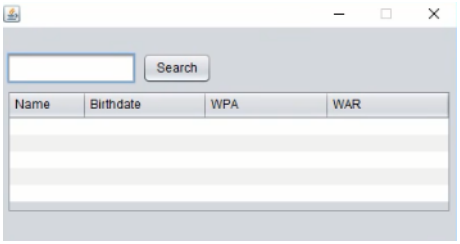

Figure 39. Player search pop-up

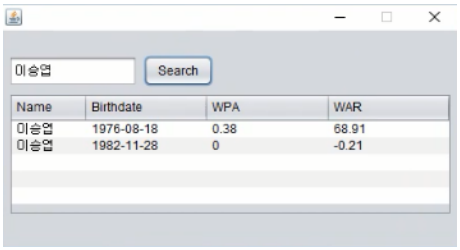

Figure 40. Player search result

- If the user clicks on search player tab, another tab is on which users can put the name of the player.
- If there is many names of the player that have the same name, the all players are shown on the table.
- The existing information is name, birthdate, WPA, WAR of the player.

6.3. Match Search

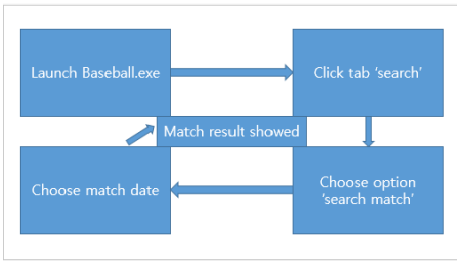

Figure 41. Match Search-flow chart

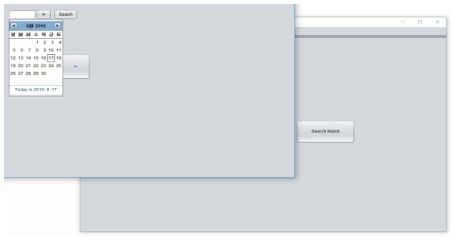

Figure 42. Match Search_date choose

- if users click on the search match tap, there shows calendar on which you can put the date.

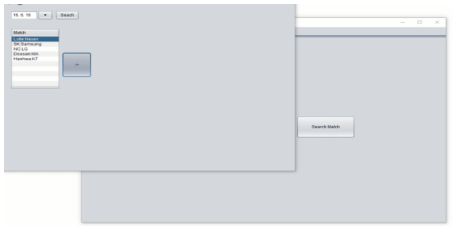

Figure 43. Match Search_match choose

- After users pick the date, you can get all the list of the matches on that date.

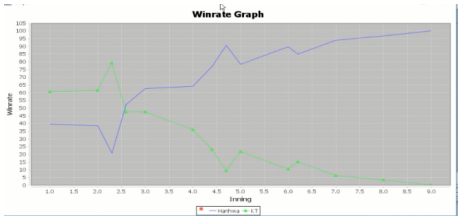

Figure 44. Match Search_Winrate Graph

- The winrate for all innings of each team are shows by linear graph.
- The values of two graphs are added up to 100% for every innings.

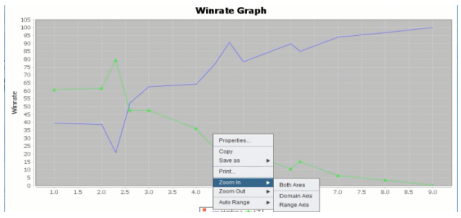

Figure 45. Graph option_zoom in, zoom out

- Users can view the graph in bigger form by clicking on zoom in button, and in smaller form by clicking on zoom out button.

Figure 46. Graph option_print graph

- Users can print the output by clicking on print tab.
- If users click on it, users can set more details such as size of the sheet, the direction of the print, the margins on the sides,top and bottom.

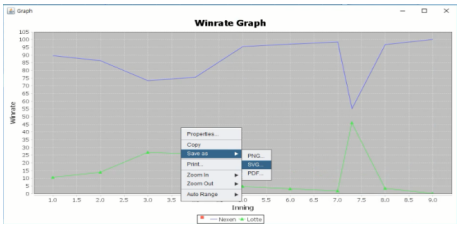

Figure 47. Graph option_save graph

- If users click on the save as tab, users can save the graph output by the file form they want.
- This program supports file form PDF, SVG, and PNG.

6.4. DB generation

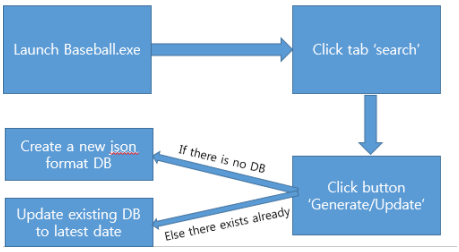

Figure 48. DB generation-flow chart

- This program uses MongoDB as data flatforms. By using MongoDB, much of the program functionality can be accessed thorugh JAVA and python.

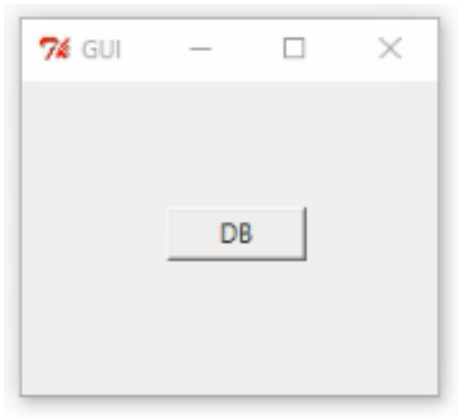

Figure 49. DB generating program

- If users click out the GUI.exe, users can generate/update DB

Figure 50. Before_client DB generation

- Users need to load cmd to generate another DB.
- Users initially generate/update db my putting the line ḿongo, úse playeránd ḍb.player.find()śequentially.

Figure 51. After_client DB generation

- The whole DB is generated or updated with the clicking on the ḎBbutton in ǴUIṕrogram afer you put the line on the cmd.

7. Software Installation Guide

7.1. MongoDB

Figure 52. MongoDB Logo

In order to execute the baseball.ai file, we first need MongoDB. Many exe file dont need any language program or other platforms because they all aimed to just execute. However, this baseball.ai file has to be installed to get to the database created by file. MongoDB is a free and open-source cross-platform document-oriented database. Its classified as a NoSQL database, and it avoids the traditional table-based relational database structure in favor of JSON-like documents with dynamic schemas. With the use of MongoDB, much of the functionality can be accessed through JAVA and python. MongoDB can be used as a file system, taking advantage of load balancing and data replication features over multiple machines for storing files. Files can be distributed and copied multiple times between machines transparently, thus effectively creating a load-balanced and fault-tolerant system.
Installation MongoDB on windows.

7.1.1. Step1. Visit downloads page and find your version of Windows. We are going to place all these files directly inside a directory C:/mongodo/. So once the download finishes, extract the zip and open the folders until you find /bin/ with a few other files. Select all these and cut/paste into our new C:/mongodb/directory.

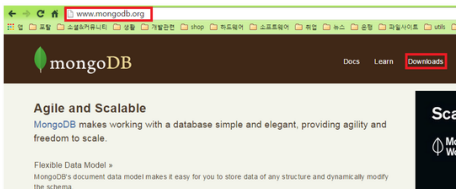

Figure 53. www.mongodb.org/

http://www.mongodb.org/
http://www.mongodb.org/downloads

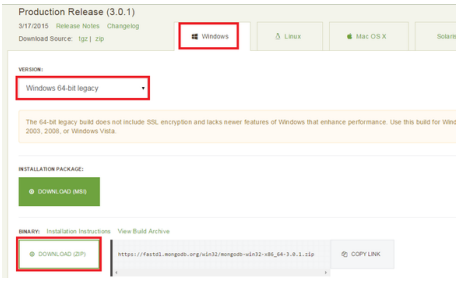

Figure 54. version/os check

Check your version and OS then downloads.

7.1.2. Step2. Add route. Add new route named MONGODB_HOME and put the address on which you uncompressed.
Revise Path as showed below.

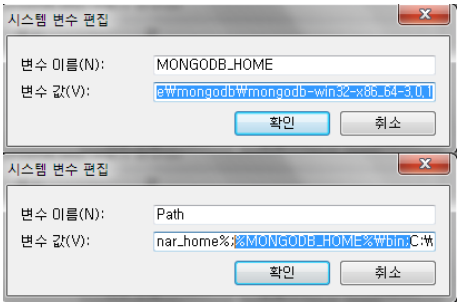

Figure 55. system variable editting

7.1.3. Step3. Now, inside this folder alongside /bin create a new folder named ”log” which will store all the MongoDB system logs. We also need to create two external directories for data storage, C:/data/ and C:/data/db.

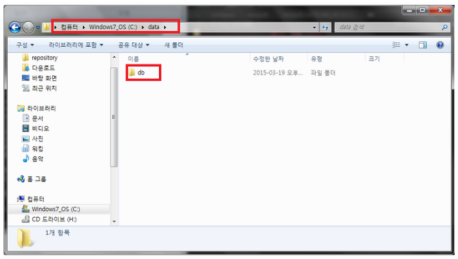

Figure 56. data storage directory

7.1.4. Step4. After creating new directory, and put mongod on the DOS command. Then we can see the server is on like below.

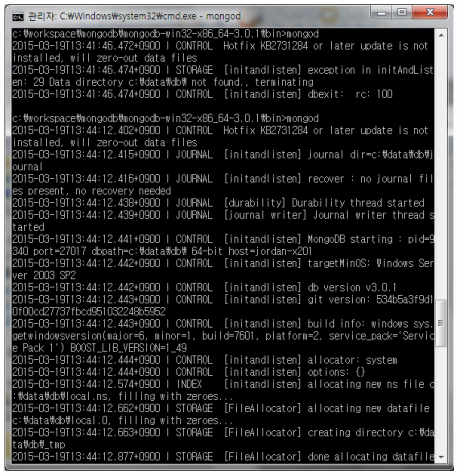

Figure 57. db creation using CMD

Then, lets put new command and put word mongo. If you are succeed to follow up to here, you can see the databased to test , which means that its successfully installed.

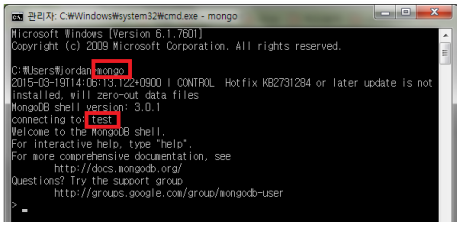

Figure 58. connect on database 'test'

7.1.5. Step5. If we open up the command prompt and run cd C:/mongodb/bin, we are looking to start the mongod.exe in shell, but after running this youll notice the operation will freeze when listening for connections. Well its not actually frozen, we are running Mongo directly through the terminal.
So, to start comman Mongo automatically as a Windows Service, first create a log file and configuration for the service. The code below executes creating the log file.

```
> echo logpath=C:\mongodb\log\mongo.log > C:\mongodb\mongod.cfg
```

Figure 59. creating log file

Now run the next two lines in terminal to create the service and get it started.

```
> C:\mongodb\bin\mongod.exe --config C:\mongodb\mongod.cfg --install
> net start MongoDB
```

Figure 60. creating service and starts

Following line shows all current databases on the server:

```
> show dbs
```

Figure 61. show database

7.2. BaseballAI.exe

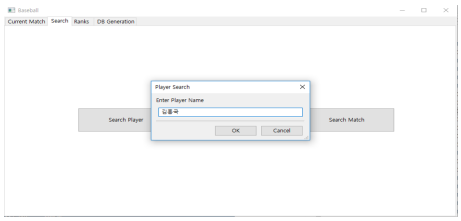

Figure 62. Search Player

7.2.1. Search Player. You can enter the name of player you want. By doing so, you can get his WPA.

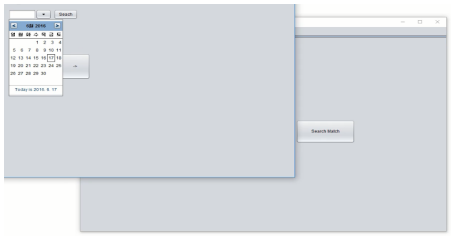

Figure 63. Search Match

7.2.2. Search Match. You can select the date of match, then you can get all the list of the matches on that date.

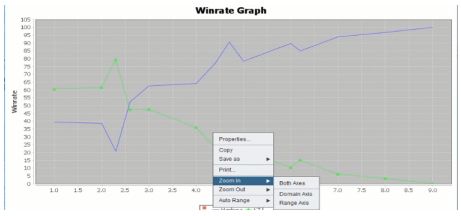

Figure 64. Winrate Graph

7.2.3. Winrate Graph. The win-rate for all innings of each teams are showed in linear graph.

8. Discussion

We decided to use python because python is a handy programming language to deal with. Our real purpose was to crawl data about korea baseball league,

which includes the name(korean). Because python 2.7 doesn't support unicode, at first we didn't know how to handle korean encoding. But at last, we transfer to python 3.5 and java who is free from unicode problem. During that transition, we had to amend what we had done partly or entirely, and it took a non trivial time to accomplish that. That was our first challenge. Second, about parsing. Because our program uses crawled data from HTML source, it is impossible to avoid parsing problem. The basic principle of parsing is quite simple. But like a many cases, the applicated, nested ones made us feel terrible. Sometimes, it tooks 3 to 4 days to parse one web site. Because both python and parsing is something we had never used before, that may be the reason why we had such a difficulties. And, about non-technical difficulties. this might be related to technical problem.. Because we all members are technically not proficient, so there was some obstacle when we take about out project, or program. For example, our members was distributed their task one by one. But because computer software is inter-related, there is some time one should know other software's basic things. But in terms of that, we were insufficient. And that makes inter-communication more difficult, slows down the talking. Through entire curriculum of our department, this class, firstly and mostly, provide students with practical application of what we learn by doing a project. So except for experienced students who did many project before this class, other class including us must be embrassed. But the more milestone of project we finish, the more we could get relief and we were suprised with ourselves that we could do such a things. Especially we really appreciate you and this class that We improved our practical skill by doing this project.