Baseball-AI
Simultaneous Winning Rate Computing

Ji Am Chung*, Young Jae Byun†, Seung Myeon Park‡, Jun Jeon§
*Department of Information System
Hanyang University
Goyang, Gyeonggi Province
Email: hummernk@gmail.com
†Department of Information System
Hanyang University
Goyang, Gyeonggi Province
Email: qusdudwowo@gmail.com
‡Department of Information System
Hanyang University
Yongin, Gyeonggi Province
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 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 separately by entire players, team, position, monthly
- 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 Substitute : 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?.

- There are some technical reasons why we chose programming language Go made by google. After a long discussion that lasted several days, group members agreed to implement the program with google Go. Because, different from C or Java, we can easily reflect up-to-date version through cloud. That would help us to keep up with newest trends in software developing, so that we can always maintain the high level stability of the program.

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).

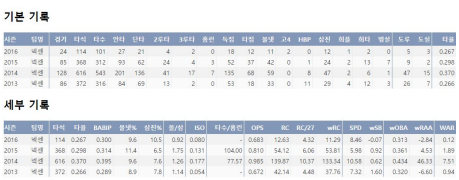
- OS
  Windows 10 pro(10586.218 build)
- Language Set
  Korean
- Computer Mode l
  MSI GT60
- Processor
  Intel(R) Core(TM) i7-3600QM CPU @ 2.4GHz
- Main Memory
  8GB RAM
- Internet Connection
  IPTIME WiFi

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 present concept of WPA.
- Rather, there are some websites providing some information which can help users WPA by themselves.

Statiz


기본 기록

시즌	팀명	경기	타석	타수	안타	단타	2루타	3루타	홈런	득점	타점	볼넷	고4	HBP	삼진	병살	희타	희플	도루	도실	타율
2016	넥센	24	114	101	27	21	4	2	0	18	12	11	2	0	12	1	2	0	5	3	0.267
2015	넥센	85	368	312	93	62	24	4	3	52	37	42	0	1	24	2	13	7	9	2	0.298
2014	넥센	128	616	543	201	136	41	17	7	135	68	59	0	8	47	2	6	1	47	15	0.370
2013	넥센	86	372	316	84	69	13	2	0	53	18	33	0	11	29	4	12	3	26	7	0.266

세부 기록

시즌	팀명	타석	타율	BABIP	장타%	삼진%ㅣ	볼/삼	ISO	타수/홈런	OPS	RC	RC/27	wRC	SPD	wSB	wOBA	wRAA	WAR
2016	넥센	114	0.267	0.300	9.6	10.5	0.92	0.080	-	0.683	12.63	4.32	11.29	8.46	-0.07	0.313	-2.64	0.12
2015	넥센	368	0.298	0.314	11.4	6.5	1.75	0.131	104.00	0.810	54.12	6.06	53.81	5.98	0.92	0.361	4.53	1.89
2014	넥센	616	0.370	0.395	9.6	7.6	1.26	0.177	77.57	0.985	139.87	10.37	133.34	10.58	0.62	0.434	46.33	7.51
2013	넥센	372	0.266	0.289	8.9	7.8	1.14	0.054	-	0.672	42.14	4.48	37.76	7.32	1.60	0.320	-6.60	0.94

Figure 1. Statiz

KBReport

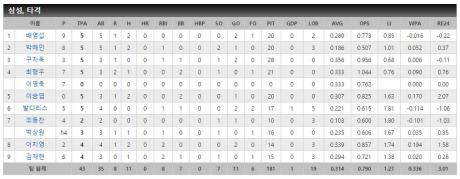


Figure 2. KBReport

TheBook

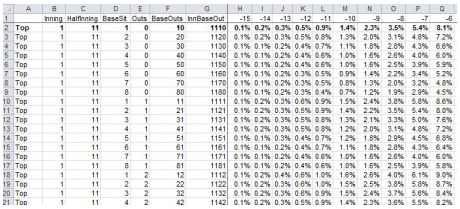


Figure 3. TheBook

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 and etc.
User		
Customer		
Software developer		
Delopment manager		

4. Specification

4.1. Data Handling

Figure 4. crawling

4.1.1. Crawling.

- Get every single data from baseball webpage to construct the database.
- Data means game box score which contains every statement in actual game.
- Crawiling source : http://www.koreabaseball.com/Schedule/GameBoxScore.aspx?leagueId=n&seriesId=0&gameId=yyyymmddt1t2x&gyear=yyyy

n of leagueId is arbitrary 1 digit number, yyyymmddt1t2x of gameId is year, month, day, team1, team2, 1 digit ordinal number, yyyy of gyear is year

- Crawler tries to crawl web data every 30 second.

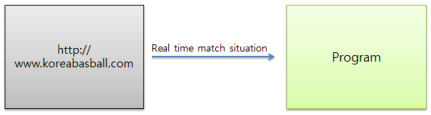


Figure 5. capturing

4.1.2. Capturing.

- Get real-time data when match is underway.

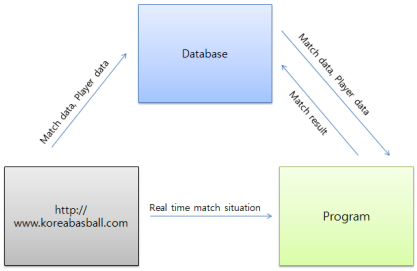


Figure 6. real-time mirroring

4.1.3. Real-Time Mirroring.

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

4.1.4. Computing.

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

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.

	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 7. 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 8. excel output file

- gameres.xlsl or gameres.jpeg are examples of 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.
- 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

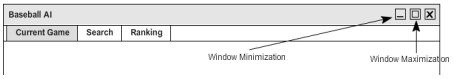


Figure 9. Windows Minim.&Maxim.

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.

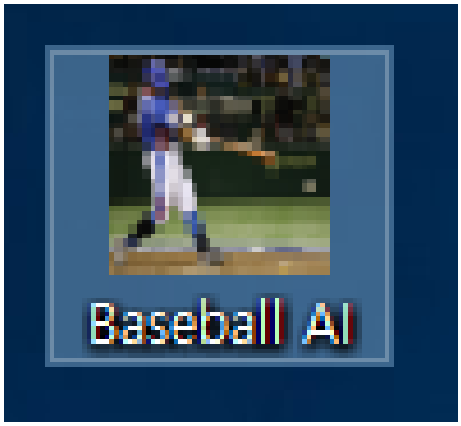


Figure 10. Program Turn On

4.3.2. 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 program again.

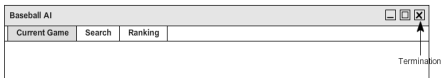


Figure 11. Program Turn Off

- User can turn off the program by clicking x button at the top-right corner of the program.

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 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.

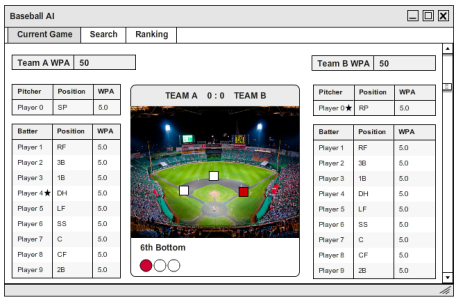


Figure 12. Current Game

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.

4.3.6. Player Ranking.

- Sort players by team, position with WPA stats.

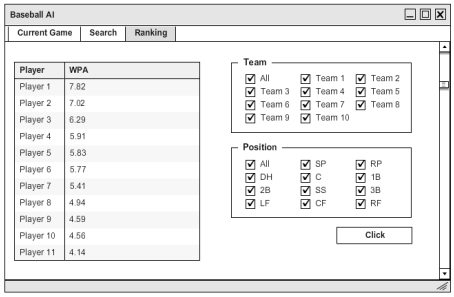


Figure 13. Player Ranking

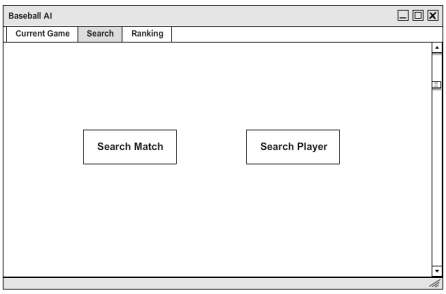


Figure 14. Search Option

4.3.7. Data Searching.

- Search option constitutes of Search Match/ Search Player

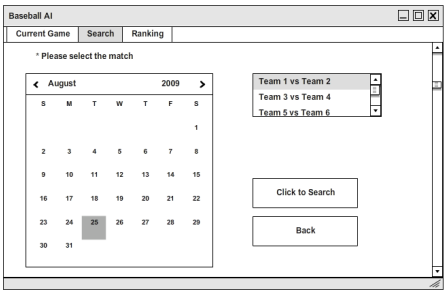


Figure 15. Search-match option

- If option Search Match is chosen, program shows match schedule as a calendar.
- The result is shown above.
- If option Search Player is chosen, input box will appear. It requires user to enter a players name.
- The result is shown as above.

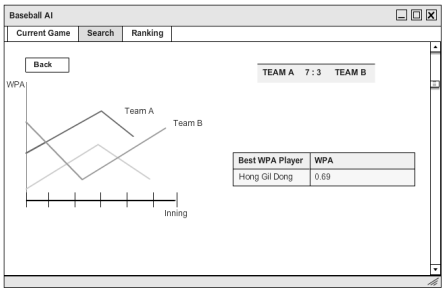


Figure 16. Search match result

Figure 17. Search Player

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 Substitute : 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).

4.3.10. Database.

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

Figure 18. Search Player Result

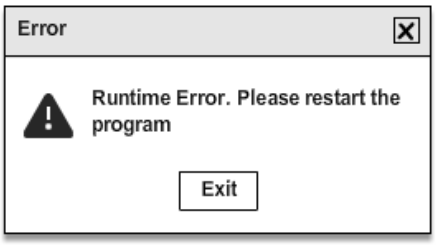


Figure 19. Error Notification

5. design

5.1. design detail

	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 20. 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 21. 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 22. 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 23. 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 24. player table