Saberlytics Software Suite

Dylan Hallagan
St. John Fisher College, dylanhallagain@msn.com

Cody Thomas
St. John Fisher College, codythomas2011@hotmail.com

How has open access to Fisher Digital Publications benefited you?

Follow this and additional works at: http://fisherpub.sjfc.edu/cs_undergrad

Part of the Software Engineering Commons

Recommended Citation

Please note that the Recommended Citation provides general citation information and may not be appropriate for your discipline. To receive help in creating a citation based on your discipline, please visit http://libguides.sjfc.edu/citations.

This document is posted at http://fisherpub.sjfc.edu/cs_undergrad/4 and is brought to you for free and open access by Fisher Digital Publications at St. John Fisher College. For more information, please contact fisherpub@sjfc.edu.
Saberlytics Software Suite

**Document Type**
Undergraduate Project

**Department**
Computer Science

**First Supervisor**
Nicholas Bucciarelli

**Subject Categories**
Computer Sciences | Software Engineering

**Comments**
Final presentation for CSCI480: Senior Project.

This undergraduate project is available at Fisher Digital Publications: [http://fisherpub.sjfc.edu/cs_undergrad/4](http://fisherpub.sjfc.edu/cs_undergrad/4)
PROJECT OVERVIEW

- Project Objectives & Description
- System Technical Overview
- Engineering Process Methodology
- Engineering Test Methodology
- Engineering Key Work-Products
  - Requirements Analysis
  - Architecture Development/Models
  - Test Strategy/Test Cases/Expected Results/Corrective Actions
- Project Key Metrics
- Project Schedule/Key Milestones
- System Preview
- System Demonstration
- Academic Key Knowledge Acquired/Applied
- Strategic Value
- Questions
- Lessons Learned

Project Mentor: Nicholas C. Bucciarelli, Ph.D
PROJECT OBJECTIVES & DESCRIPTION

• The primary objective of the Saberlytics Software Suite
  • To provide GMs, coaches, scouts, and data analysts a tool to house separate research in one unified software
  • Readily access the research

Project Mentor: Nicholas C. Bucciarelli, Ph.D
SYSTEM TECHNICAL OVERVIEW

• **Hardware:** Linux based server

• **Soft Technologies:**
  - MySQL Database
  - PHP 5
  - JavaScript
  - HTML
  - Bootstrap CSS 3
SYSTEM TECHNICAL OVERVIEW

- SYSTEM LEVEL DIAGRAM

Browser (Chome and Firefox)

User Interface (Javascript, Ajax, PHP, HTML, CSS)

Application Level (PHP 5.6, Javascript)

Lahman and Saberlytics DB (MySQL 5.5)
ENGINEERING PROCESS METHODOLOGY

• AGILE: Used as a project development methodology
  • Iterative design and build process
    • Allowed revisions and builds to be conducted in parallel
    • Ease of enhancement, refinement, and reduced errors
DEVELOPMENT PHASES - ITERATIVE

• **Phase 1:** Functionality Specifications

• **Phase 2:** Build a prototype for all tables and working data without a login system.

• **Phase 3:** Integrate Prototype 1 with user functionality
  • Login system
  • Project management
    • File uploads
    • Entering new reports
    • Player tear sheets
  • Realigned functionalities based on change in requirements
ENGINEERING TEST METHODOLOGY

• Test Methodology
  • Software Engineering Test – Unit Testing
    • Correct functionalities (Based on requirements)
  • Quality Engineering Test – Systems Acceptance Test
    • Compare SQLs results with industry leading resources
      • BaseballReference.com
      • Fangraphs.com

• Value of Testing
  • Validates the capabilities of the application to work as stated
ENGINEERING KEY WORK-PRODUCTS

• Requirements Document/Traceability
• Test Strategy (plans and executed results)

• Value of Work Products
  • Iterative, Evolving
    • Based upon Final Version of Application
### Requirements Document/Traceability

<table>
<thead>
<tr>
<th>ID</th>
<th>Req. ID</th>
<th>Business Requirement</th>
<th>Functional Requirement</th>
<th>Technical / System Requirement</th>
<th>System Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>001/002</td>
<td>We need a system that will provide a software suite to analyze baseball player statistics</td>
<td>The system shall take data input from the RetroSheet Database</td>
<td>The system shall allow the user to make selections through an SQL query builder</td>
<td></td>
</tr>
<tr>
<td>002</td>
<td>002/003/004/005</td>
<td>The system suite shall contain reporting services for custom projections based on the statistical analysis and data modeling.</td>
<td>The reporting services shall allow the user to create custom statistical selections to develop a model</td>
<td>The system shall export the query to R for data modeling</td>
<td></td>
</tr>
<tr>
<td>003</td>
<td>006</td>
<td>The system suite shall contain a player profile</td>
<td>The system shall allow for scouts to add input to an athlete’s player profile through a comment interface</td>
<td>The comment section shall allow filters based on scout, date, and tags</td>
<td></td>
</tr>
<tr>
<td>004</td>
<td>007/008</td>
<td></td>
<td>The system shall incorporate a queued comment section in the player profile containing scout name, scout picture, date added, tags, and the comment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>005</td>
<td>009/10</td>
<td>The system shall display the athlete’s picture and basic statistics in the player profile</td>
<td></td>
<td>The comment section shall allow filters based on scout, date, and tags</td>
<td></td>
</tr>
<tr>
<td>006</td>
<td>011</td>
<td>The system shall contain GameDay Reporting functionality</td>
<td>The system shall be able to rapidly produce advanced opponent and player statistics for coaches</td>
<td></td>
<td></td>
</tr>
<tr>
<td>007</td>
<td>012/013</td>
<td></td>
<td>The system shall display recorded Heat Map, Pitch FX, and matchup w/ Pitch FX images in the GameDay report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>014</td>
<td></td>
<td>The GameDay report shall provide an opponent pitch breakdown for analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>009</td>
<td>015</td>
<td></td>
<td>The GameDay report shall queue recorded analyst comments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>010</td>
<td>016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>011</td>
<td>017/018/019</td>
<td>The system shall incorporate a user login database</td>
<td>The system shall provide specific login interfaces based on role</td>
<td>The system shall allow for the roles of admin, coach, scout, and analyst</td>
<td></td>
</tr>
</tbody>
</table>

Project Name: Statistical Baseball Analytic Software
National Center: St. John Fisher College
Project Manager Name: Dylan Hallagan & Cody Thomas

Project Description: Provide a software suite to analyse baseball players using R for custom projections, a player profile for scouts in the organization, and on demand reports for coaches during games.
PROJECT KEY METRICS

• Number Open Tasks/Stories per Sprint
• Number of Changes in Requirements in Total
• Number of Changes to Source Lines of Code (SLOC)
## PROJECT SCHEDULE/KEY MILESTONES

### Standard Senior Project Class Schedule for All Teams

<table>
<thead>
<tr>
<th>ID</th>
<th>% Work Complete</th>
<th>% Complete</th>
<th>Task Mode</th>
<th>Task Name</th>
<th>Duration</th>
<th>Start</th>
<th>Finish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0%</td>
<td>24%</td>
<td></td>
<td>Senior Project 2015</td>
<td>74 days</td>
<td>Mon 1/19/15</td>
<td>Thu 4/30/15</td>
</tr>
<tr>
<td>2</td>
<td>100%</td>
<td>100%</td>
<td></td>
<td>Executive Summary</td>
<td>10 days</td>
<td>Mon 1/19/15</td>
<td>Fri 1/30/15</td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Requirements</td>
<td>6 days</td>
<td>Mon 2/2/15</td>
<td>Mon 2/9/15</td>
</tr>
<tr>
<td>4</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Key Metrics</td>
<td>2 days</td>
<td>Tue 2/10/15</td>
<td>Wed 2/11/15</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Architecture</td>
<td>5 days</td>
<td>Thu 2/12/15</td>
<td>Wed 2/18/15</td>
</tr>
<tr>
<td>6</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Test Strategy</td>
<td>3 days</td>
<td>Thu 2/19/15</td>
<td>Mon 2/23/15</td>
</tr>
<tr>
<td>7</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Prototype Completed</td>
<td>15 days</td>
<td>Tue 2/24/15</td>
<td>Mon 3/16/15</td>
</tr>
<tr>
<td>8</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>System Reviews_1st Iterative</td>
<td>10 days</td>
<td>Tue 3/17/15</td>
<td>Mon 3/30/15</td>
</tr>
<tr>
<td>9</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Test Strategy/Exection</td>
<td>2 days</td>
<td>Tue 3/17/15</td>
<td>Wed 3/18/15</td>
</tr>
<tr>
<td>10</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Corrective Actions</td>
<td>3 days</td>
<td>Thu 3/19/15</td>
<td>Mon 3/23/15</td>
</tr>
<tr>
<td>11</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Regression Test</td>
<td>2 days</td>
<td>Mon 3/23/15</td>
<td>Tue 3/24/15</td>
</tr>
<tr>
<td>12</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>System Reviews_2nd Iterative</td>
<td>5 days</td>
<td>Mon 3/23/15</td>
<td>Fri 3/27/15</td>
</tr>
<tr>
<td>13</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Test Strategy/Exection</td>
<td>1 day</td>
<td>Wed 3/25/15</td>
<td>Wed 3/25/15</td>
</tr>
<tr>
<td>14</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Corrective Actions</td>
<td>1 day</td>
<td>Thu 3/26/15</td>
<td>Thu 3/26/15</td>
</tr>
<tr>
<td>15</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Regression Test</td>
<td>1 day</td>
<td>Fri 3/27/15</td>
<td>Fri 3/27/15</td>
</tr>
<tr>
<td>16</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Review with Department</td>
<td>8 days</td>
<td>Mon 4/6/15</td>
<td>Wed 4/15/15</td>
</tr>
<tr>
<td>17</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>1st Review</td>
<td>1 day</td>
<td>Mon 4/6/15</td>
<td>Mon 4/6/15</td>
</tr>
<tr>
<td>18</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Changes, if needed</td>
<td>2 days</td>
<td>Tue 4/7/15</td>
<td>Wed 4/8/15</td>
</tr>
<tr>
<td>19</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>2nd Final Review</td>
<td>1 day</td>
<td>Mon 4/13/15</td>
<td>Mon 4/13/15</td>
</tr>
<tr>
<td>20</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Baseline Final Presentation</td>
<td>1 day</td>
<td>Wed 4/15/15</td>
<td>Wed 4/15/15</td>
</tr>
<tr>
<td>21</td>
<td>0%</td>
<td>0%</td>
<td></td>
<td>Submit to Department Chairperson</td>
<td>1 day</td>
<td>Wed 4/15/15</td>
<td>Wed 4/15/15</td>
</tr>
</tbody>
</table>

Project Mentor: Nicholas C. Bucciarelli, Ph.D
SYSTEM PREVIEW

• Dashboard Page
System Preview

- Team Page

New York Yankees

The New York Yankees are an American professional baseball franchise based in the Bronx borough of New York City that competes in Major League Baseball's American League East division.

Total Payroll: $231,978,085.00  Average Age: 31  Manager: Joe Girardi

Batting

<table>
<thead>
<tr>
<th>Season</th>
<th>Team</th>
<th>G</th>
<th>AB</th>
<th>PA</th>
<th>H</th>
<th>HR</th>
<th>1B</th>
<th>2B</th>
<th>3B</th>
<th>HR</th>
<th>R</th>
<th>RB</th>
<th>IBB</th>
<th>SO</th>
<th>HBP</th>
<th>SF</th>
<th>SH</th>
<th>SB</th>
<th>CS</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>NYA</td>
<td>162</td>
<td>5449</td>
<td>6044</td>
<td>1321</td>
<td>594</td>
<td>247</td>
<td>24</td>
<td>24</td>
<td>144</td>
<td>659</td>
<td>614</td>
<td>466</td>
<td>33</td>
<td>1214</td>
<td>57</td>
<td>36</td>
<td>3</td>
<td>115</td>
<td>31</td>
</tr>
<tr>
<td>2012</td>
<td>NYA</td>
<td>162</td>
<td>5524</td>
<td>6231</td>
<td>1462</td>
<td>421</td>
<td>290</td>
<td>13</td>
<td>245</td>
<td>304</td>
<td>774</td>
<td>565</td>
<td>31</td>
<td>1176</td>
<td>62</td>
<td>49</td>
<td>2</td>
<td>93</td>
<td>27</td>
<td>0.2647</td>
</tr>
<tr>
<td>2011</td>
<td>NYA</td>
<td>162</td>
<td>5619</td>
<td>6306</td>
<td>1462</td>
<td>453</td>
<td>267</td>
<td>33</td>
<td>222</td>
<td>967</td>
<td>836</td>
<td>627</td>
<td>34</td>
<td>1136</td>
<td>74</td>
<td>51</td>
<td>8</td>
<td>147</td>
<td>46</td>
<td>0.2631</td>
</tr>
<tr>
<td>2010</td>
<td>NYA</td>
<td>162</td>
<td>5667</td>
<td>6379</td>
<td>1485</td>
<td>543</td>
<td>275</td>
<td>32</td>
<td>201</td>
<td>859</td>
<td>823</td>
<td>662</td>
<td>36</td>
<td>1136</td>
<td>73</td>
<td>44</td>
<td>8</td>
<td>163</td>
<td>30</td>
<td>0.2668</td>
</tr>
<tr>
<td>2009</td>
<td>NYA</td>
<td>162</td>
<td>5669</td>
<td>6447</td>
<td>1684</td>
<td>586</td>
<td>325</td>
<td>21</td>
<td>244</td>
<td>915</td>
<td>881</td>
<td>663</td>
<td>35</td>
<td>1014</td>
<td>54</td>
<td>39</td>
<td>8</td>
<td>111</td>
<td>28</td>
<td>0.2834</td>
</tr>
<tr>
<td>2008</td>
<td>NYA</td>
<td>162</td>
<td>5672</td>
<td>6257</td>
<td>1512</td>
<td>643</td>
<td>289</td>
<td>20</td>
<td>180</td>
<td>789</td>
<td>758</td>
<td>535</td>
<td>36</td>
<td>1016</td>
<td>80</td>
<td>39</td>
<td>11</td>
<td>118</td>
<td>39</td>
<td>0.2714</td>
</tr>
</tbody>
</table>

© Dylan and Cody 2015

Page rendered in 0.0002 seconds - Saberlytics Software
SYSTEM PREVIEW

- Player Page

---

**Player**

**Derek Jeter**

Positions: First Baseman, Third Baseman and Leftfielder
Bats: R Throws: R
Height: 75 Weight: 195lb.

**Comparable Players**

<table>
<thead>
<tr>
<th>Player</th>
<th>War</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon Stewart</td>
<td>823.5</td>
</tr>
<tr>
<td>Paul Molitor</td>
<td>803.8</td>
</tr>
<tr>
<td>Robin Yount</td>
<td>802.9</td>
</tr>
<tr>
<td>Pete Rose</td>
<td>764.9</td>
</tr>
<tr>
<td>Mark Kotsay</td>
<td>760.3</td>
</tr>
<tr>
<td>Dan Money</td>
<td>741.4</td>
</tr>
<tr>
<td>Rusty Staub</td>
<td>737.4</td>
</tr>
<tr>
<td>Dan Ford</td>
<td>733.1</td>
</tr>
<tr>
<td>Thurman Munson</td>
<td>728.9</td>
</tr>
<tr>
<td>Bob Watson</td>
<td>714.7</td>
</tr>
</tbody>
</table>

**Aging Curve**

War Projection for Derek Jeter

Source: Fangraphs 2015

Equation: $y = -0.04x^2 + 2.11x - 24.98$
# SYSTEM PREVIEW

- **Project Management**

![Project Management Dashboard](image)

## Projects and Assignments

<table>
<thead>
<tr>
<th>No</th>
<th>ID</th>
<th>Player</th>
<th>Assignment</th>
<th>Status</th>
<th>Start Date</th>
<th>Deadline</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0015</td>
<td>Gian Carlo Stanton</td>
<td>Performance Analysis</td>
<td>Planned</td>
<td>April 17, 2015</td>
<td>Deadline Reached</td>
<td><img src="edit" alt="Edit" /></td>
</tr>
<tr>
<td>2</td>
<td>0015</td>
<td>Derek Jeter</td>
<td>Prospect Analysis</td>
<td>Planned</td>
<td></td>
<td>Deadline Reached</td>
<td><img src="edit" alt="Edit" /></td>
</tr>
<tr>
<td>3</td>
<td>0015</td>
<td>Manny Machado</td>
<td>Performance Analysis</td>
<td>Planned</td>
<td></td>
<td>Deadline Reached</td>
<td><img src="edit" alt="Edit" /></td>
</tr>
</tbody>
</table>

© Dylan and Cody 2015

Page rendered in 0.0002 seconds - Saberlytics Software

---

Project Mentor: Nicholas C. Bucciarelli, Ph.D
SYSTEM DEMONSTRATION

Saberlytics.com
ACADEMIC KEY KNOWLEDGE ACQUIRED/APPLIED

• CSCI 231: Data Structure and Algorithms
• CSCI 353: Database Structures
  • Database Architecture
  • Advanced SQL querying
• CSCI 375: Programming the WWW
  • PHP
    • Custom framework developed in PHP
  • Javascript/JQuery
STRATEGIC VALUE

• Value Proposition
  • Automated Valuation
    • Internal metrics
  • All franchise studies and work in one spot
  • Gameday Report Generation
LESSONS LEARNED

• SQL Database Structuring
  • With large databases time to render is a problem
    • Techniques for optimization
      • Crack a database into like queries/datasets
        • Example: Computed Player Statistics

• Advanced PHP
  • Login/Authorization
  • Session Handling
  • Background Fetchs
  • Dynamic pages rendering using AJAX Javascript/JQuery