

2008

MIS510 Web
Computing and Mining

Sriram Srinivasan,
Kalpesh Jain, Shankar
Venkataraman

[CRICWEB.COM]

A mash-up of cricket information

1	Introduction	3
2	Business Model	4
2.1	Competitive Analysis.....	4
2.2	Revenue Model	5
3	Architecture	5
4	DATA MINING AND VISUALIZATION	10
4.1	Microsoft SQL Server 2005 Data Mining Platform.....	10
4.2	Player Clustering	10
4.3	Cricweb Predictor.....	11
5	Novelty	12
6	Appendix	14

1 Introduction

Cricket is the most popular sport in countries like India, Pakistan, Sri Lanka, Bangladesh, Australia, South Africa, and England. With billions of followers all around the world, cricket enthusiasts use the internet for variety of reasons to stay updated with the latest happenings in cricket. Cricket fans use to internet for cricket news, live scores, video highlights, player information, cricket related blogging, etc. Currently there is no website that caters to all these needs. Cricweb (Figure 1) is one-stop website for all cricket related information. With a high level of service, and continuously updating cricket information, Cricweb is the ultimate resource for all cricket aficionados.



Figure 1 CricWeb - Home Page

2 Business Model

2.1 Competitive Analysis

CricWeb is a one stop site for all cricket enthusiasts. Cricweb targets all cricket lovers who browse the internet for cricket related information and interact with other people online. It provides rich and varied multimedia content; combination which is not offered by any of the competitors.

For instance:

- Cricinfo.com provides rich information about cricket news; events, player profiles etc, but it does not provide video highlights of cricket matches.

- Howstat.com provides all the information related to cricket statistics but it does not provide any kind of multimedia content
- Yahoo Cricket does not provide any videos or blogs.

2.2 Revenue Model

Cricket is the second most popular game in the world and CricWeb.com can cash in the popularity of the sport to generate revenue.

The following are some of the sources from which revenue could be generated:

1. Google Ad Sense: CricWeb. Com will be posting targeted advertisements from Google.
2. Amazon and EBay: CriWeb can generate revenue through merchandising fees received from EBay and Amazon.
3. Direct Advertisements: CricWeb would be posting direct advertisements on the website. The advertisements will be charged based on location of the ad, size and duration.
4. Collaboration with team and events sponsors: Cricweb can collaborate with team and events sponsors for promoting a particular cricket event or team and generate revenue through these partnerships.

3 Architecture

Following is the architecture of the Cricweb site.

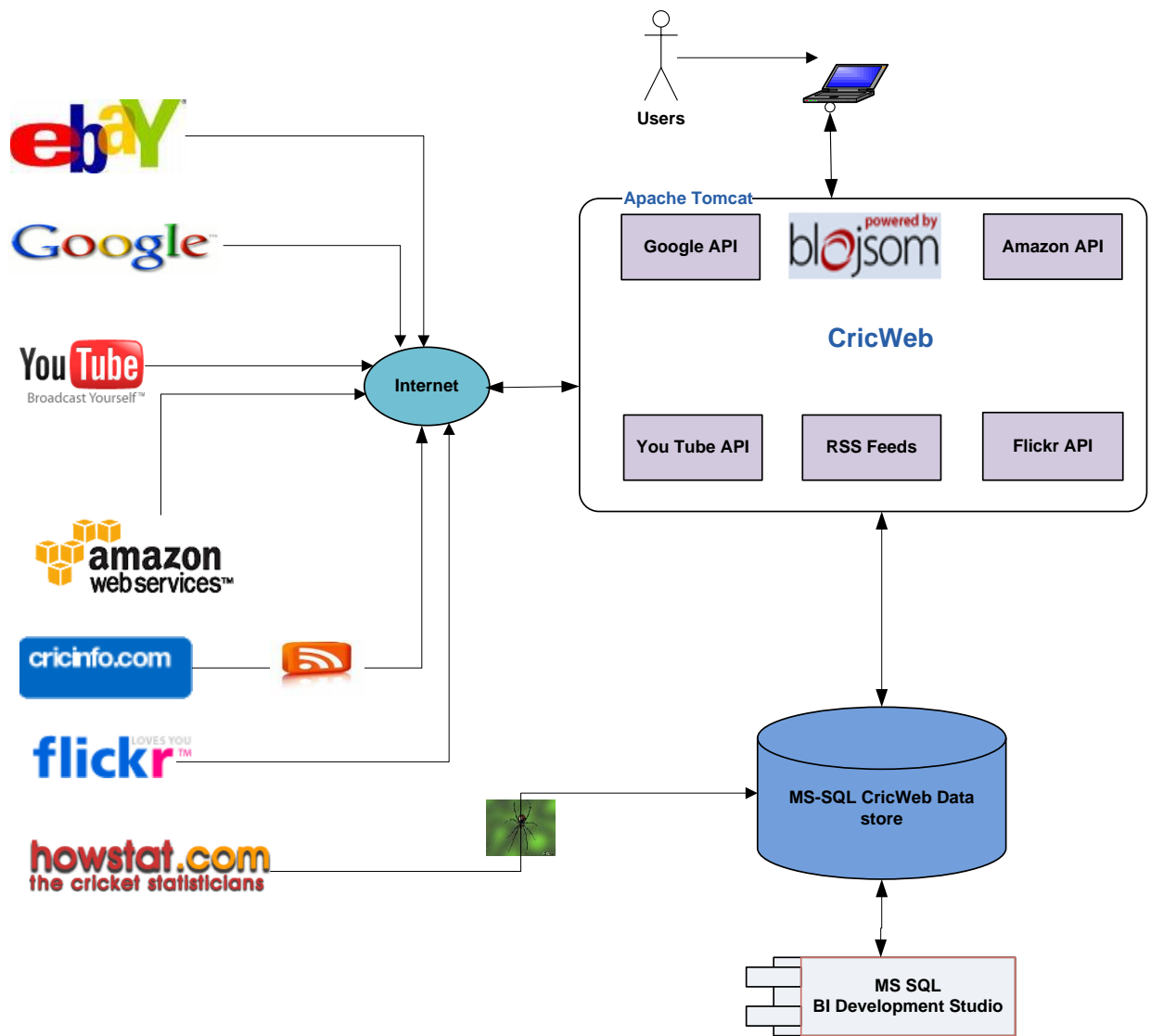


Figure 2 Architecture

Cricweb.com is built on a 3 tier architecture (Figure 2) with Microsoft SQL server 2005 serving as the database and Apache 5.x serving as the web server. The application server is tomcat. For the data mining component, Business Intelligence Studio suite of MS-SQL 2005 is used. The statistics for different players are spidered periodically from howstat.com. Cricweb.com uses the following API's:

Google, You Tube, Amazon, Flickr.com, Cricinfo.com and Cricnext.com RSS feeds. Cricweb.com provides the following functionalities.

- **Cricket News from Major web sites (RSS)** - A user can access the cricket news section from the home page (Refer Figure 1). When the user clicks on the “Read more” button he is redirected to a page where he can view the latest cricket news from cricketnext.com and cricinfo.com. User can also view the latest news about the Indian premium league.
- **Fixtures & Live Scores (RSS)** – A user can view the live scoreboards of all the matches that are happening currently all over the world. The live scores are retrieved from cricinfo.com as rss feeds.
- **Player Profiles** – A user can access the player profiles by accessing the Players & Teams tab in the menu (Refer Figure 1). The user can view the player of the day. His ODI statistics are shown by default. A user can also view the performance of that player against different countries. When the user clicks on the “View Performance Summary” link a graph is shown with number of runs scored against each country and the batting averages against each country (Figure 3).

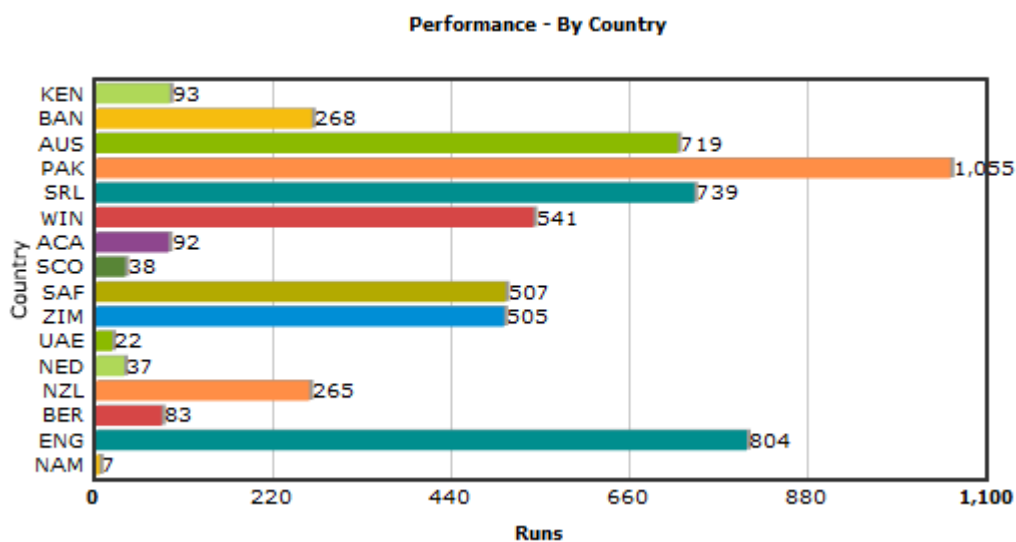


Figure 3 Player Performance Chart

A user also has the option of viewing his recent performances in the last five matches (Figure 4).

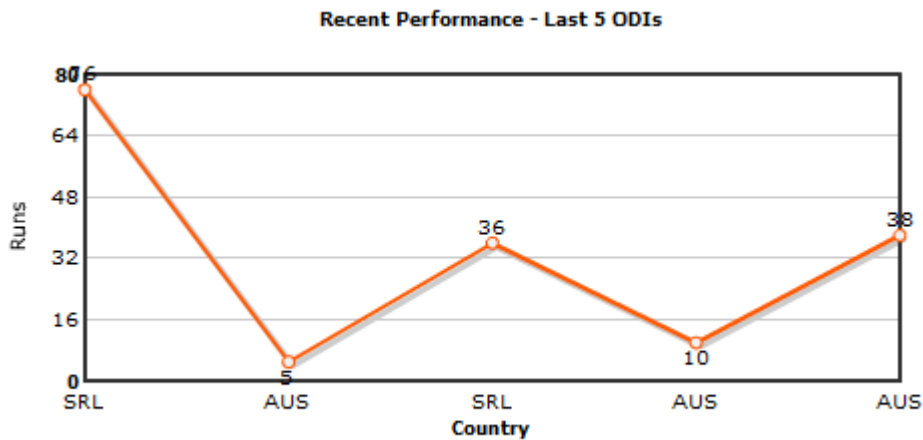


Figure 4 Recent Performance

The graph shows the runs scored by the player in recent matches. The user is also provided with a list of similar players and user has the option of viewing the reasons for similarity (Figure 5).

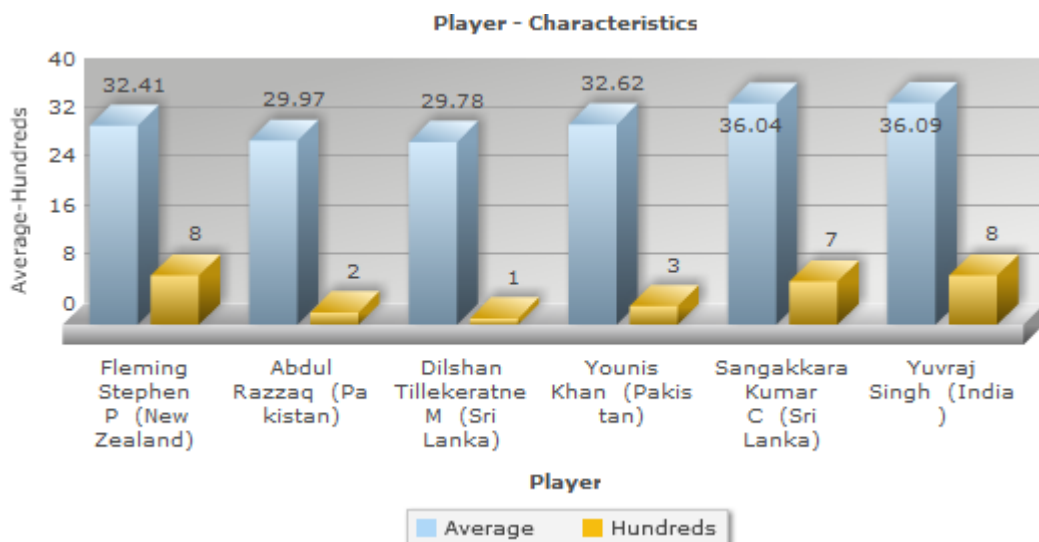


Figure 5 Player Characteristics

Test matches statistics and performance graphs can also be viewed in this section. While considering the recent performance in test matches both first and second innings are taken into consideration (Figure 6).

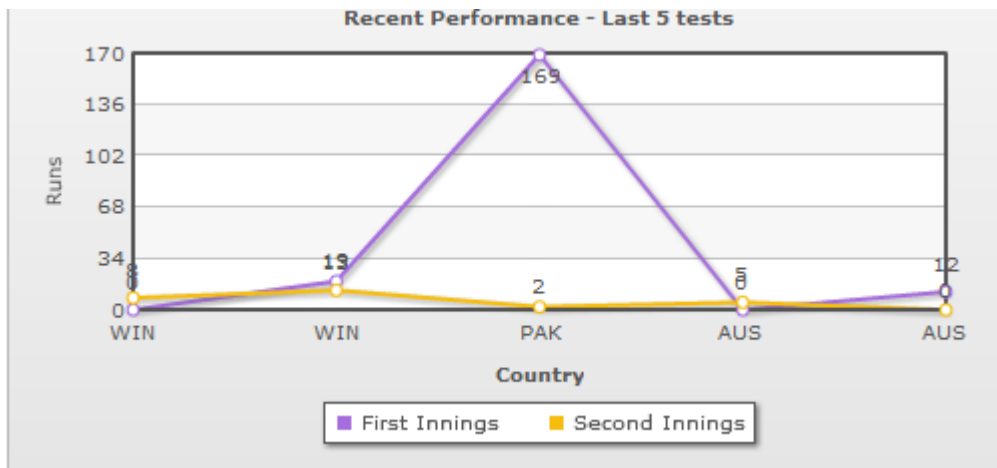


Figure 6 Recent Performance - Test Matches

The photos of the particular player are retrieved from flickr.com. Users can search for different players using the search feature.

- Google Map & Weather Information** – The user can access the fixtures of all the matches which are going to happen in the current week. This section allows the user to view the different stadiums on the Google map and view the weather details at that particular city where the match is going to be played.
- Cricket Highlights, Videos, and Pictures** – Videos & Gallery section which can be accessed from the home page provides user with videos, pictures and latest cricket highlights. A user can access the latest cricket videos available at youtube.com. Search feature available in this section allows the user to search for videos of their choice. The cricket gallery allows the user to view pictures of their favorite players.
- Blogs** – The blog section allows the user to view blogs from different websites at a single glance. A user can also express his opinions on the latest news in cricket through our own blogging site. The blogs section is built on bloksom, an open source blogging application.

- **CricWeb Prediction System** – In this section, an user can predict the performance of team India based on certain conditions such as the opposition against whom the match will be played, result of the toss, venue and whether India bats first or not.
- **Cricket Shopping** – A user can also buy items related to cricket through our website. Our websites displays cricket related items from both Amazon.com and E-Bay. The user also has the option of searching for his favorite items on Amazon and E-Bay through our site.

4 DATA MINING AND VISUALIZATION

4.1 Microsoft SQL Server 2005 Data Mining Platform

The Microsoft SQL Server 2005 Data Mining Platform provides capabilities to predict future results based on input, or attempt to find relationships among data or cluster data in previously unrecognized yet similar groups. Microsoft data mining tools are different from traditional data mining applications in significant ways. First, they support the entire development lifecycle of data in the organization, which Microsoft refers to as Integrate, Analyze, and Report. Microsoft data mining algorithms can be run in real time, allowing for the real-time validation of data against a set of mined data.

4.2 Player Clustering

Player profiles which are spidered from HowStat.com were used to group similar players.

Attributes like runs number of matches, scored, average, and hundreds scored were used. In case of test matches, runs scored from both the innings were taken into consideration. A mining model was constructed using Microsoft K-Means clustering algorithm. Figure 8 shows the mining model. The clusters with stronger links are represented by thick lines and those with weaker links are shown using thin lines. The results of this model are stored into the SQL server

tables. In future, when a new player's profile is fed into this model, it will place him into one of the 10 clusters according to attributes selected.

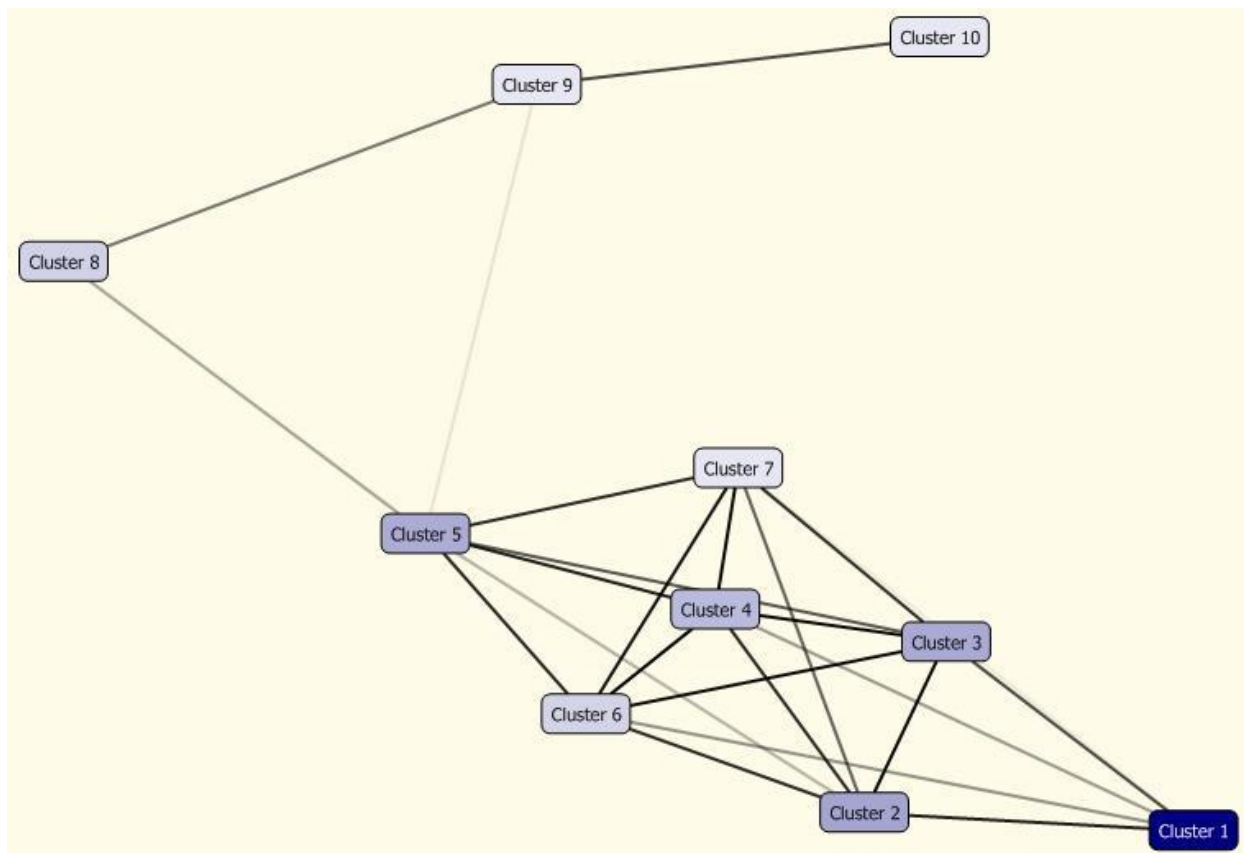


Figure 7 Players Clustering

4.3 Cricweb Predictor

The CricWeb Predictor can predict India's performance against an opposition given the toss, location of the match. Microsoft Neural Networks model was built based on match results for past 20 years. Figure 9 shows the neural network model. The model learns to predict the result based on the results from matches held in last 20 years. The learning is continuous since as new matches are played, their results will be fed into the model. The results of the model are stored into SQL server table and displayed in the browser according to user query.

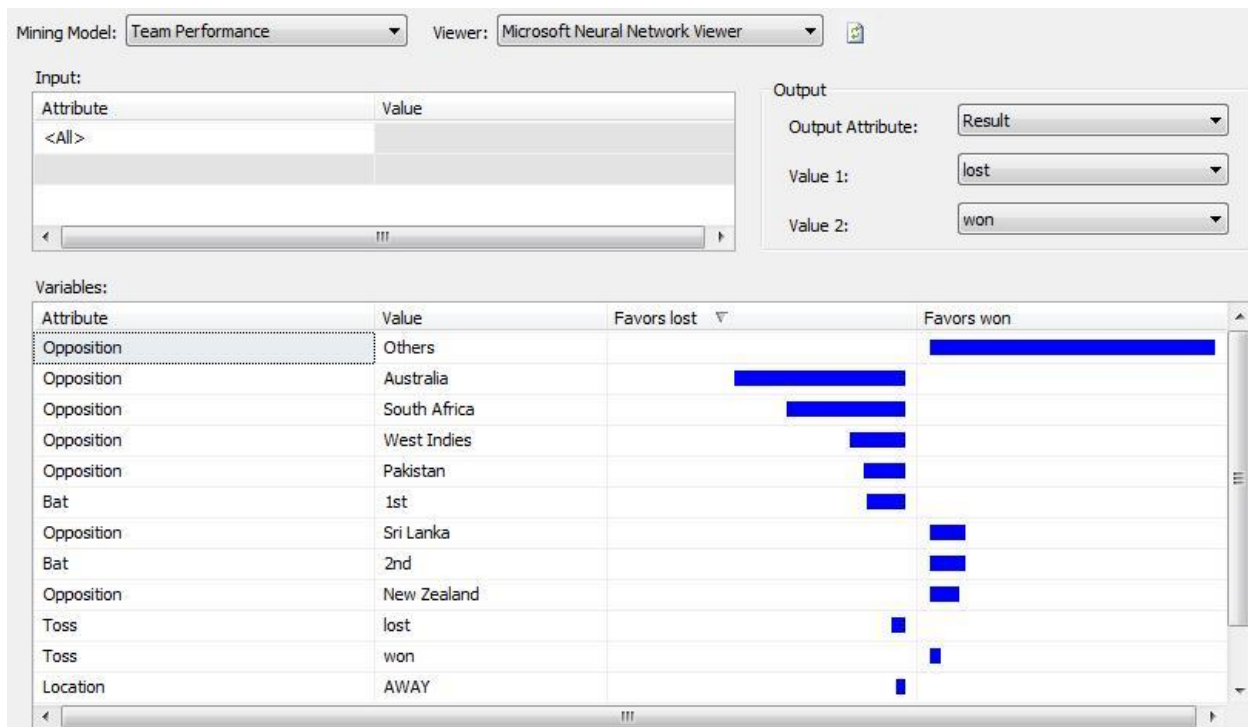


Figure 8 Neural Network Model

5 Novelty

The following are the ways in Cricweb is unique:

1. **Idea:** We are using Web 2.0 technology to create a mash up of all cricket related information. We are using API's from Google, Amazon, eBay, Flickr, YouTube, RSS feeds from Cricinfo. By combining features from these sources and data which we have crawled from Howstat we provide a rich multimedia content.
2. **Player Clustering:** We have data mined players records and using K-Means Algorithm we have divided them to different clusters. When a user searches for a particular player he can also find information about players having similar profiles. Such a feature is unique to Cricweb.com.

3. **Cricweb Predictor:** We have data mined information about India's matches for the past 20 years. Using Neural Network Algorithm we can predict the outcome of India's next match given certain set of conditions. There is no other resource existing on the internet providing such functionality.
4. **Visualization:** To help users better understand data and get a rich user experience, we used several visualization techniques supported by dynamic flash-based charts (Fusion Charts).

6 Appendix

Project Timelines and Resource Allocation

Project timeline

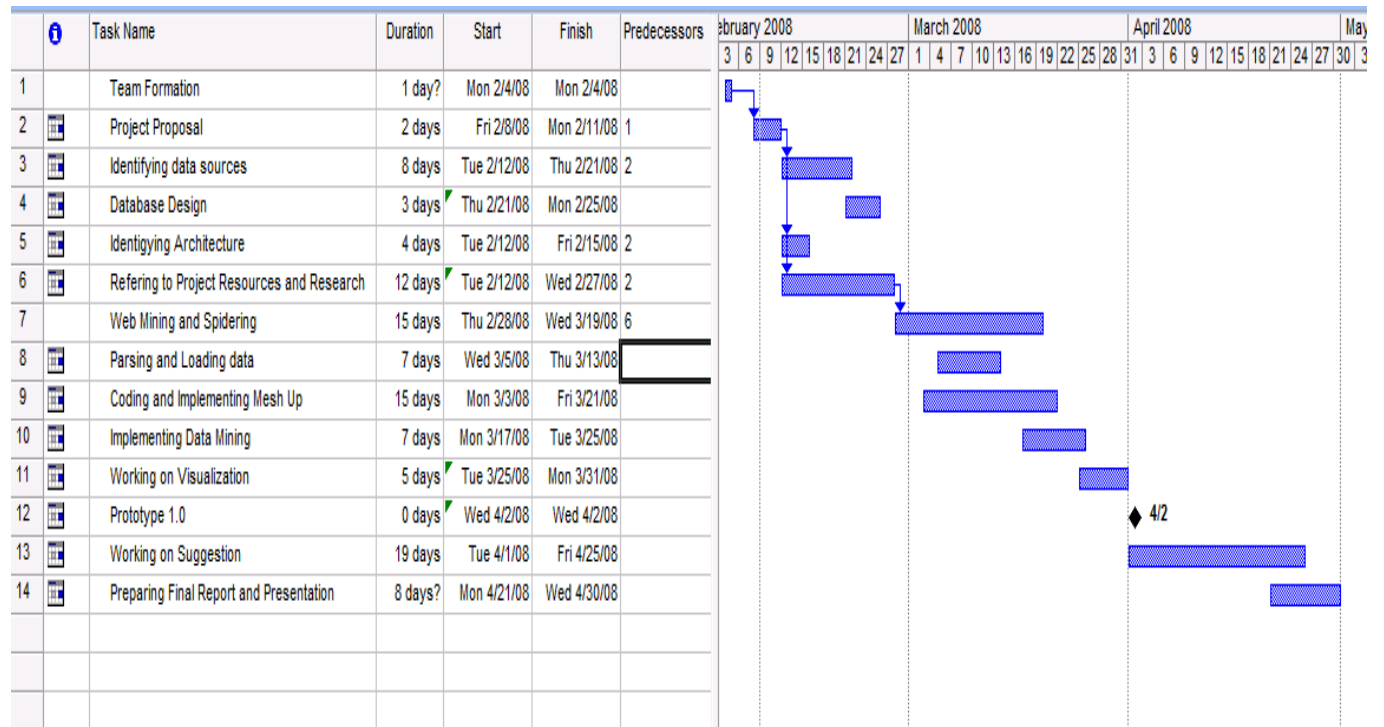


Figure 9 Project Timeline

Figure A: Project Timeline

Player of the Day

Search Player: [Search](#)

View Similar Players >> [Fleming Stephen P](#) [Dilshan Tillekeratne M](#) [Sehwag Virender](#) [Boucher Mark V](#) [Porterfield William T S](#)

[How are they similar?](#)

Name: **Yuvraj Singh (India)** 

Matches: 204 (2000-)

Batting

Innings:186

Notouts:26

Aggregate:5775

Average:36.09

Highest Score:139

50s:35

100s:8

Ducks:9

Opened Batting:2

Scoring Rate:86.56

Born: 12/12/1981

[View Performance Summary](#)

Bowling

Overs:438.4

Balls:2632

Maidens:13

Runs:2247

Wickets:55

Average:40.85

4 Wicket Innings:1

Best:4/6

Economy Rate:5.12

Strike Rate:47.85

[View Test Profile](#)

[View Recent Performance](#)

Fielding

Catches:64

Most Catches:3

Wicket keeping

Catches:0

Stumpings:0

Most Catches:0

Most Dismissals:0

Captaincy

Matches/Won/Lost

Tosses Won:



Figure 10 Player of the day

Welcome to Cricweb!

Cricweb is one-stop website for all cricket related information.

Cricket enthusiast can use Cricweb for variety of reasons to stay updated with the latest happenings in cricket. Cricket fans can use our website for cricket news, live scores, match fixtures, map information, video highlights, player information, image gallery, online cricket shopping, cricket related blogging and player performance prediction. Currently there is no website that caters to all these needs. Cricweb is an one-stop website for all cricket related information. With a high level of service, and continuously updating cricket information, Cricweb provides the ultimate resource for all cricket aficionados!



**STUART SURRIDGE
OLD CRICKET BAT,
LONDON**



Starting Bid: **USD \$0.93**

Item location: Melbourne, VIC
Ships to: Worldwide

Seller: **licaptainbirdseye**
Feedback: 4197

Bid Now right now on **ebay**



**CRICKET BAT
STICKERS
MANY MORE
DESIGNS**

Figure 11 Home Page



Cricket Videos

Watch latest cricket highlights from Youtube. Also, search for cricket videos.

[Read more](#)

Cricket Gallery

View images of recent matches and events.

[Read more](#)



Search You Tube

[Search](#)



IPL Cricket Highlights - Sehwag Cleaned Up By M.Gony!
Duration: 41s



Live IPL Cricket Action @ Cric-Tube! - Gambhir Steyn Sehwag!
Duration: 52s



IPL Highlights - Dhoni Breaks Bat Off Glen McGrath!
Duration: 21s

Figure 12 YouTube Videos

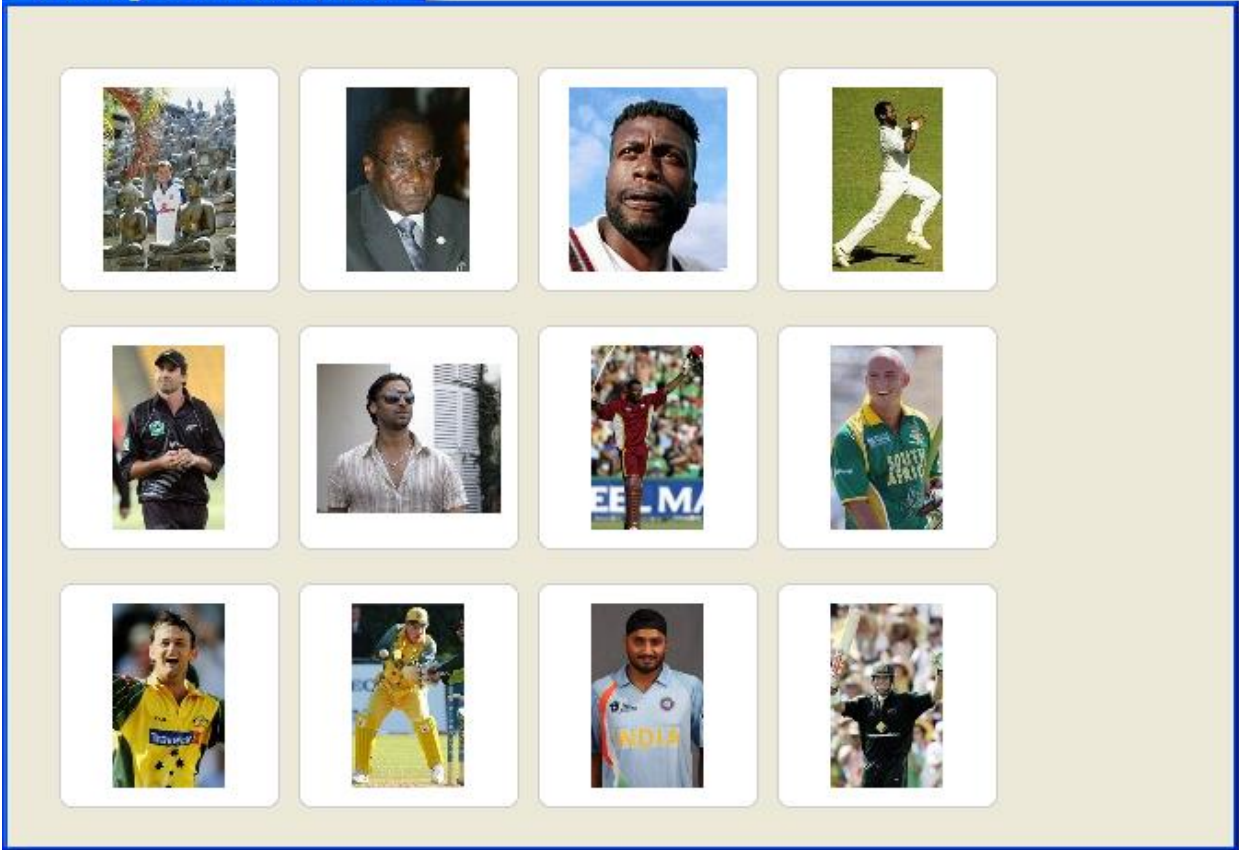


Figure 13 Image Gallery

Home | Players & Teams | Videos & Gallery | Blogs

CRICWEB.COM

Cricket Blogs

Collection of Blogs from major cricketing websites.

Express your ideas on issues related to current tours, players, matches, statistics, etc

[Read more](#)

CricWeb Blog

Welcome to CricWeb Blogs!

Join our Forum. Discuss issues related to cricket. Express yourself

[Read more](#)

IPL BUZZ

Catch up with the latest happenings of Indian Premier League. Express your ideas about various players auctions, team compositions, etc

[Read more](#)

Cricket Blogs!

Tour Diaries

- [Electrifying Eden](#)
- [Indian Foreign Legion?](#)
- [Slapgate - the IPL's first controversy](#)

Whats new on Cricinfo

- [Slogout! - new cricket game](#)
- [World XI Fantasy Cricket League now open for entries](#)

Figure 14 RSS feeds of Blogs

Tuesday, 6 May 2008

Shoaib Akhtar banned for five years

In a dramatic decision, the Pakistan Cricket Board announced that Shoaib Akhtar had been banned for 5 years for violating the players' code of conduct, a document that I'm unable to locate on the PCB website.

As per the ICC code of conduct, a ban of 2-5 years is imposed on a player in the following situations:

Betting on a game/series where the player/his team was a participant.

Tried to sweet-talk someone else into betting.

Gambled on a game/event.

Tried to sweet-talk someone else into gambling.

Received money/benefit/rewards for providing weather, team and pitch information.

Received money/benefit/rewards and thus brought the game into disrepute.

Provided money/benefit/rewards and thus brought the game into disrepute.

Was approached to bet/gamble/provide information but didn't disclose the event to designated staff.

Is aware of another player doing so but didn't report to the designated staff.

Was Shoaib's offense more severe than the above? Now, he cannot play for Pakistan or in

« May »						
Sun	Mon	Tue	Wed	Thu	Fri	Sat
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

About:

Default Owner

Categories:

Uncategorized

Recently:

Shoaib Akhtar
banned for five years

what Graeme Smith
should have said

Let's embrace IPL,
but with a little more
maturity

Figure 15 CricWeb Blog



Rankings

Check out the latest team rankings. Find the best teams, batsmen and bowlers for both Test and ODI matches

[Read more](#)

Top Searched Players

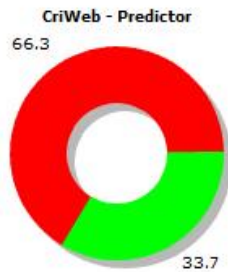
Tendulkar Sachin R (India)
Yuvraj Singh (India)
Ponting Ricky T. (Australia)
Abdul Razzaq (Pakistan)
Sehwag Virender (India)

Predict Team Performance

Eager to know how India will perform in the next match? Find out the chances of India winning the next match

[Read more](#)

[Predict Again](#)



The Chances of India Losing the Game is 66%

*The results are based on performance of the team for the past 20 yrs..

Figure 16 CricWeb Predictor