MatchJunkies: A vertical search engine for soccer match statistics

Jian Tian

University of Illinois at Urbana-Champaign

jtian4@illinois.edu

August 4, 2014
Outline

- Motivation and project ideas
- Implementation details
- Results
- Future work and summary
Motivation

**What this project is:** building an efficient and accurate vertical search web application for soccer match statistics

**What is it useful:** it can serve as an exploratory tool designed for soccer fans and game commentators for both leisure and professional uses

**What is the outcome:** detailed information of the match history record between two soccer teams

- Existing general search engines usually fail to retrieve such domain-specific information.
- Specialized websites lack the functionality of optimized search operation.
Advantages over existing tools

MatchJunkies include improved functionalities over existing applications:

★ Single and multiple match search
★ Simple query
  ○ *Use only team names for querying*
★ Advanced match information details
  ○ *Longest win/loss streaks, highest scoring game in history, top scorers, etc.*
General structure of the implementation

Web crawler

Crawling-time analysis

"Modified" inverted index

User-input query

Index

Local documents

Lexicon and postings file

Search

Display (web app)

Offline

Online
Web crawling

- **Tool:** PhantomJS embedded in Python script.
- **Data source:** www.11v11.com for match records.
- Crawl both the record pages and match pages, perform analysis on match pages at the same time.

"Record page"

<table>
<thead>
<tr>
<th>Date</th>
<th>Team</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 Apr 2002</td>
<td>Germany v Argentina</td>
<td>L</td>
</tr>
<tr>
<td>09 Feb 2005</td>
<td>Germany v Argentina</td>
<td>D</td>
</tr>
<tr>
<td>21 Jun 2005</td>
<td>Germany v Argentina</td>
<td>D</td>
</tr>
<tr>
<td>30 Jun 2006</td>
<td>Germany v Argentina</td>
<td>W</td>
</tr>
<tr>
<td>03 Mar 2010</td>
<td>Germany v Argentina</td>
<td>L</td>
</tr>
<tr>
<td>03 Jul 2010</td>
<td>Argentina v Germany</td>
<td>W</td>
</tr>
<tr>
<td>15 Aug 2012</td>
<td>Germany v Argentina</td>
<td>L</td>
</tr>
<tr>
<td>13 Jul 2014</td>
<td>Germany v Argentina</td>
<td>W</td>
</tr>
</tbody>
</table>

"Match page"

**Germany:**
- **Manager:** Joachim Low
- **Goals:** Mario Götze (113)

Starting lineup:
- Manuel Neuer (Goalkeeper)
- Benedikt Höwedes (Defender)
- Jerome Boateng (Defender)
- Mats Hummels (Defender)
- Philipp Lahm (captain) (Defender)
- Bastian Schweinsteiger (Midfielder)
- Christoph Kramer (Midfielder)
- Toni Kroos (Midfielder)
- Thomas Müller (Midfielder)
- Mesut Özil (Midfielder)
- Miroslav Klose (Forward)

Substitutions:
- André Schurrle for Christoph Kramer (31)
- Mario Götze for Miroslav Klose (88)
- Per Mertesacker for Mesut Özil (120)

**Argentina:**
- **Manager:** Alejandro Sabella

Starting lineup:
- Sergio Romero (Goalkeeper)
- Marcos Rojo (Defender)
- Martin Demichelis (Defender)
- Pablo Zabaleta (Defender)
- Ezequiel Garay (Defender)
- Javier Mascherano (Midfielder)
- Enzo Pérez (Midfielder)
- Lucas Biglia (Midfielder)
- Lionel Messi (captain) (Midfielder/Forward)
- Gonzalo Higuain (Forward)
- Ezequiel Lavezzi (Forward)

Substitutions:
- Sergio Agüero for Ezequiel Lavezzi (46)
- Rodrigo Palacio for Gonzalo Higuain (78)
- Fernando Gago for Enzo Pérez (86)
Using a “modified” inverted index.

Lexicon contains the list of unique team names.

Postings file contains the “position” of the team in the document, instead of the frequency.

- Argentina 1:-1, 1036:-1, 1083:-1, ···, 48:1, 49:1, ···
- France 1005:-1, 1052:-1, 1099:-1, ···, 800:1, 801:1, ···
- Germany 1006:-1, 1053:-1, 1100:-1, ···, 851:1, 852:1, ···
- Italy 1009:-1, 1035:1, 1036:1, ···, 1198:-1, 1245:-1, ···
- Spain 1027:-1, 1074:-1, 1121:-1, ···, 1881:1, 1882:1, ···
- USA 1033:-1, 1080:-1, 1127:-1, ···, 2163:-1, 2164:-1, ···
- ···

“1” for “query team”, “-1” for “opponent team”.
Searching

- Take in user-input queries.
  - Two team names, such as “Germany” and “Argentina”
- Find the inverted index of the two teams.
  - Germany 1006:-1, 1053:-1, 1100:-1, · · · , 851:1, 852:1, · · ·
  - Argentina 1:-1, 1036:-1, 1083:-1, · · · , 48:1, 49:1, · · ·
- Union by the document IDs.
  - Germany 848:1, 65:-1
  - Argentina 65:1, 848:-1
- Return the relevant document, which is docID of 848.
Web interface

Web url: https://web.engr.illinois.edu/~jtian4/matchjunkies/
Germany national team record vs. Argentina

Match summary

Games won: 8
Games drawn: 5
Games lost: 9
### Match history

<table>
<thead>
<tr>
<th>Date</th>
<th>Match</th>
<th>Result</th>
<th>Score</th>
<th>Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>08 Jun 1958</td>
<td>West Germany v Argentina</td>
<td>W</td>
<td>3-1</td>
<td>FIFA World Cup</td>
</tr>
<tr>
<td>16 Jul 1966</td>
<td>Argentina v Germany</td>
<td>D</td>
<td>0-0</td>
<td>FIFA World Cup</td>
</tr>
<tr>
<td>14 Feb 1973</td>
<td>West Germany v Argentina</td>
<td>L</td>
<td>2-3</td>
<td>Internal</td>
</tr>
<tr>
<td>05 Jun 1977</td>
<td>Argentina v West Germany</td>
<td>W</td>
<td>1-3</td>
<td>Internal</td>
</tr>
<tr>
<td>12 Sep 1979</td>
<td>West Germany v Argentina</td>
<td>W</td>
<td>2-1</td>
<td>Internal</td>
</tr>
<tr>
<td>01 Jan 1981</td>
<td>Argentina v West Germany</td>
<td>L</td>
<td>2-1</td>
<td>Gold Cup</td>
</tr>
<tr>
<td>24 Mar 1982</td>
<td>Argentina v West Germany</td>
<td>D</td>
<td>1-1</td>
<td>International</td>
</tr>
<tr>
<td>12 Sep 1984</td>
<td>West Germany v Argentina</td>
<td>L</td>
<td>1-3</td>
<td>International</td>
</tr>
<tr>
<td>29 Jun 1986</td>
<td>Argentina v West Germany</td>
<td>L</td>
<td>3-2</td>
<td>FIFA World Cup</td>
</tr>
<tr>
<td>16 Dec 1987</td>
<td>Argentina v West Germany</td>
<td>L</td>
<td>1-0</td>
<td>International</td>
</tr>
<tr>
<td>02 Apr 1988</td>
<td>West Germany v Argentina</td>
<td>W</td>
<td>1-0</td>
<td>International</td>
</tr>
<tr>
<td>08 Jul 1990</td>
<td>West Germany v Argentina</td>
<td>W</td>
<td>1-0</td>
<td>FIFA World Cup</td>
</tr>
<tr>
<td>15 Dec 1993</td>
<td>Argentina v Germany</td>
<td>L</td>
<td>2-1</td>
<td>International</td>
</tr>
<tr>
<td>17 Apr 2002</td>
<td>Germany v Argentina</td>
<td>L</td>
<td>0-1</td>
<td>International</td>
</tr>
<tr>
<td>09 Feb 2005</td>
<td>Germany v Argentina</td>
<td>D</td>
<td>2-2</td>
<td>International</td>
</tr>
<tr>
<td>21 Jun 2005</td>
<td>Germany v Argentina</td>
<td>D</td>
<td>2-2</td>
<td>FIFA Confederations Cup</td>
</tr>
<tr>
<td>30 Jun 2006</td>
<td>Germany v Argentina</td>
<td>W</td>
<td>1-1</td>
<td>FIFA World Cup</td>
</tr>
<tr>
<td>03 Mar 2010</td>
<td>Germany v Argentina</td>
<td>L</td>
<td>0-1</td>
<td>International</td>
</tr>
<tr>
<td>03 Jul 2010</td>
<td>Argentina v Germany</td>
<td>W</td>
<td>1-0</td>
<td>FIFA World Cup</td>
</tr>
<tr>
<td>15 Aug 2012</td>
<td>Argentina v Germany</td>
<td>W</td>
<td>1-3</td>
<td>International</td>
</tr>
<tr>
<td>13 Jul 2014</td>
<td>Germany v Argentina</td>
<td>W</td>
<td>1-0</td>
<td>FIFA World Cup</td>
</tr>
</tbody>
</table>

**Match: Germany v Argentina**

**Manager:** Joachim Low

**Goals:**
- Mario Götze (113)

**Starting lineup:**
- Manuel Neuer (Goalkeeper)
- Benedikt Howedes (Defender)
- Jerome Boateng (Defender)
- Mats Hummels (Defender)
- Philipp Lahm (captain) (Defender)
- Bastian Schweinsteiger (Midfielder)
- Christoph Kramer (Midfielder)
- Toni Kroos (Midfielder)
- Thomas Müller (Midfielder)
- Mesut Özil (Midfielder)
- Miroslav Klose (Forward)

**Substitutions:**
- Andre Schürrle for Christoph Kramer (71)
- Mario Götze for Miroslav Klose (88)
- Per Mertesacker for Mesut Özil (120)

**Manager:** Alejandro Sabella

**Goals:**

**Starting lineup:**
- Sergio Romero (Goalkeeper)
- Marcos Rojo (Defender)
- Martin Demichelis (Defender)
- Pablo Zabaleta (Defender)
- Ezequiel Garay (Defender)
- Javier Mascherano (Defender)
- Enzo Pérez (Midfielder)
- Lucas Biglia (Midfielder)
- Lionel Messi (captain) (Midfielder/Forward)
- Gonzalo Higuaín (Forward)
- Ezequiel Lavezzi (Forward)

**Substitutions:**
- Sergio Aguero for Ezequiel Lavezzi (65)
- Rodrigo Palacio for Gonzalo Higuaín (78)
- Fernando Gago for Enzo Pérez (86)
Longest winning streak: 2

for the matches on:
- 05 Jun 1977, 12 Sep 1979
- 02 Apr 1988, 08 Jul 1990

Longest losing streak: 3

for the matches on:
- 12 Sep 1984, 29 Jun 1986, 16 Dec 1987

Highest scoring game:

<table>
<thead>
<tr>
<th>Date</th>
<th>Match</th>
<th>Result</th>
<th>Score</th>
<th>Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 Feb 1973</td>
<td>West Germany v Argentina</td>
<td>L</td>
<td>2-3</td>
<td>International</td>
</tr>
<tr>
<td>29 Jun 1986</td>
<td>Argentina v West Germany</td>
<td>L</td>
<td>3-2</td>
<td>FIFA World Cup</td>
</tr>
</tbody>
</table>
Germany top scorer against Argentina

Miroslav Klose

Goals: 7

Argentina top scorer against Germany

Jorge Luis Borruchaga

Goals: 6
Future work

- Auto-correction for user-input queries
- Extend database to include club teams
  - Same approach, run web crawler in parallel for acceleration
- More efficiently retrieve the player’s images
  - Correctly locate the image from google results
We developed a vertical search engine “MatchJunkies” with the purpose of providing users with match history records between two national soccer teams.

MatchJunkies was implemented using a web application which took simple users’ queries and return a comprehensive set of information of the match record in history.

Our system will benefit both the soccer fans and commentators for leisure and professional uses.