



Gestione Avanzata dei Dati - Prof. Gennaro Costagliola

Gabriele Lombari | Alessandra Zullo | Gaetano Prisco

Anno Accademico: 2015 - 2016

FantaGAD

è il miglior sito per gli appassionati di Fantacalcio.

Specifica del Problema (1)

L'**obiettivo** che il progetto si pone è riuscire a fornire in un unico sito tutte le informazioni utili ai giocatori di Fantacalcio*.



* *Il Fantacalcio è un popolare gioco di fantasia sul calcio che consiste nell'organizzare e gestire squadre virtuali formate da calciatori reali, scelti tra quelli che giocano il torneo a cui il gioco si riferisce (nel nostro caso alla Serie A).*

Specifica del Problema (2)

I dati forniti dal progetto sono i seguenti:

- **Informazioni** su tutte le squadre di Serie A;
- **Probabili formazioni** di ogni match della prossima giornata di campionato;
- **Calendario** contenente le informazioni di tutti i match del campionato di Serie A;
- **Quote** per le partite della prossima giornata;
- **Live** delle partite in tempo reale;
- **Dettagli tecnici** dei match già conclusi.

DEMO

Fonti (1)

- Lega Serie A: <http://www.legaseriea.it>



- Utilizzata per ottenere il **calendario** del campionato, le **informazioni sulle squadre** e le **informazioni dei match passati**.
- Web Scraping



WIKIPEDIA

- Wikipedia: <https://it.wikipedia.org>

- Utilizzata per ottenere la **storia delle squadre**.
- Web Scraping

- Fastest live scores: <https://docs.crowdscores.com>

- Utilizzata per il **live** di una partita.
- API



Fonti (2)

- Google Maps: <https://maps.google.it>



- Utilizzata per ottenere le **coordinate** della **sede centrale** di ogni squadra (reverse geocoding) e per **visualizzare** la stessa sulla mappa
- API

- Google Custom Search: <https://developers.google.com/web-search/>



- Utilizzata per **ricavare** l'url della pagina **Wikipedia** della squadra di cui ricavare le informazioni.
- API

- Fantagazzetta: <https://www.fantagazzetta.com>



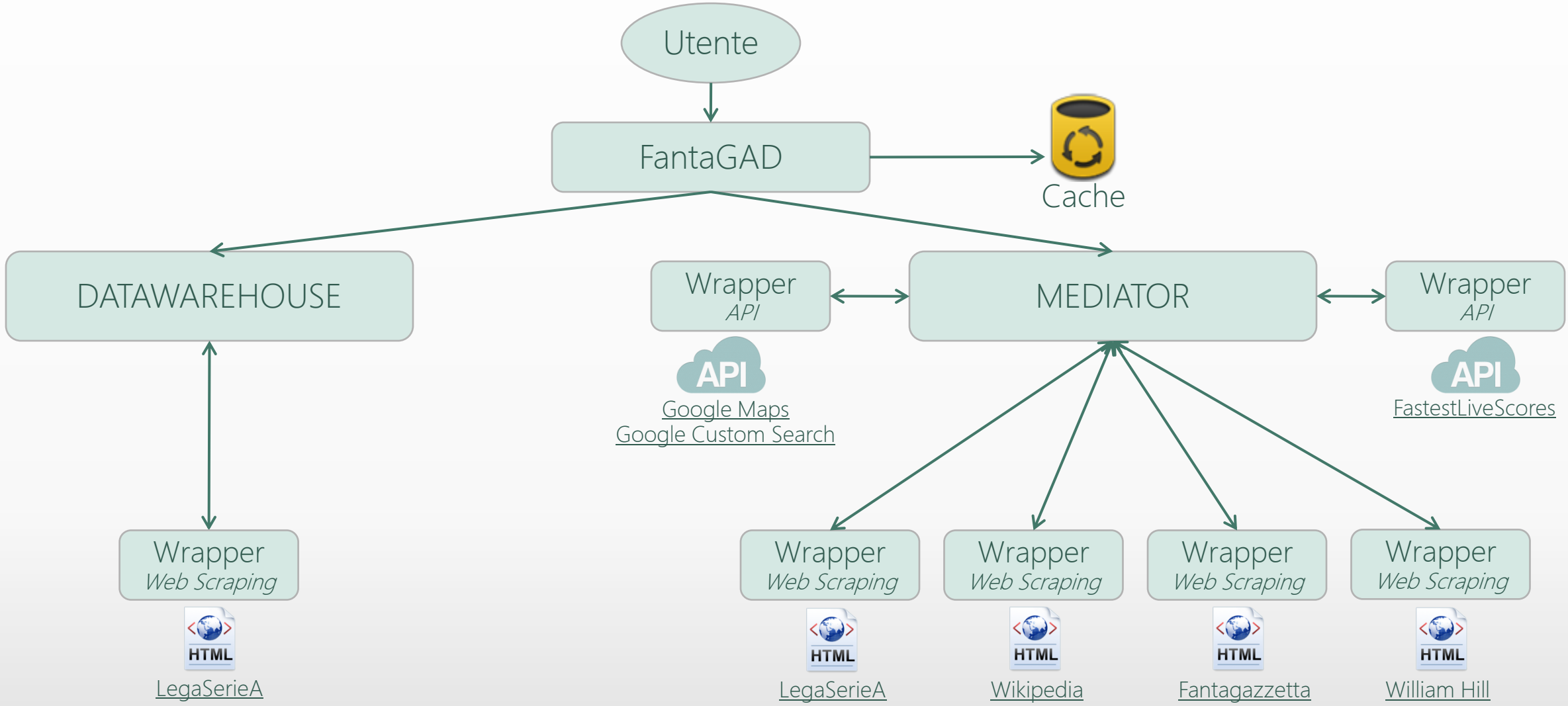
- Utilizzata per ottenere le **probabili formazioni** di ogni match relativo alla prossima giornata di campionato.
- Web Scraping

Fonti (3)



- William Hill: <http://www.williamhill.it>
 - Utilizzata per ricavare le **quote** delle partite della prossima giornata di campionato.
 - Web Scraping

Architettura



Meccanismi di caching

Per alcune query effettuate è salvato in risultato in cache con un tempo di scadenza che dipende dal grado di volatilità della fonte.

Il caching è stato effettuato sia **lato client** che **lato server**.



Cache - Server-side

INFORMAZIONI	SCADENZA
Probabili formazioni: informazioni relative alle probabili formazioni del prossimo match di campionato.	24 h



Cache - Client-side

INFORMAZIONI	SCADENZA
Calendario: informazioni relative ad ogni giornata di campionato.	24 h
Probabili formazioni: informazioni relative alle probabili formazioni del prossimo match di campionato.	24 h



Ruoli e responsabilità del team di progetto

[Gabriele Lombari]

Informazioni Squadre, Probabili
Formazioni

www.gabrielelombari.it



[Alessandra Zullo]

Calendario, Informazioni
Match

www.alessandrazullo.it



[Gaetano Prisco]

Quote, Live



Informazioni squadra



NAPOLI SOCIETA' SPORTIVA CALCIO S.P.A.
Strada Statale Domitiana Km 35,300 81030 Castel Volturno (CE)



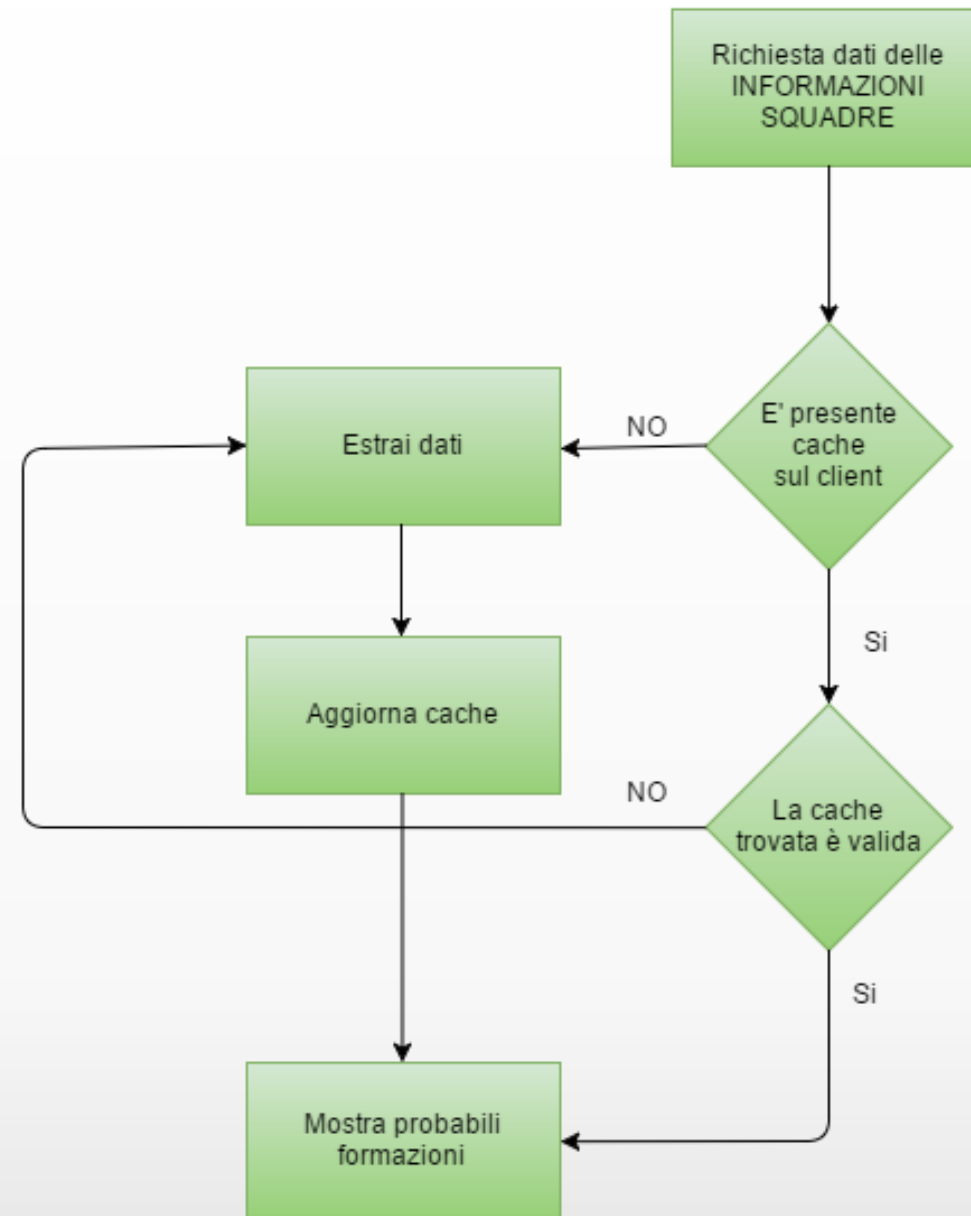
I dati sono stati estratti da wikipedia.it e legaseriea.it mediante WebScraping eseguito con appositi script PHP. Sono state inoltre utilizzate le **Google Custom Search API** e le **Google Maps API**.



Le informazioni delle squadre sono memorizzate in locale sul *client* (**localStorage**) quando la pagina è visitata.





In questa pagina è stata effettuata l'operazione di **JOINT**.

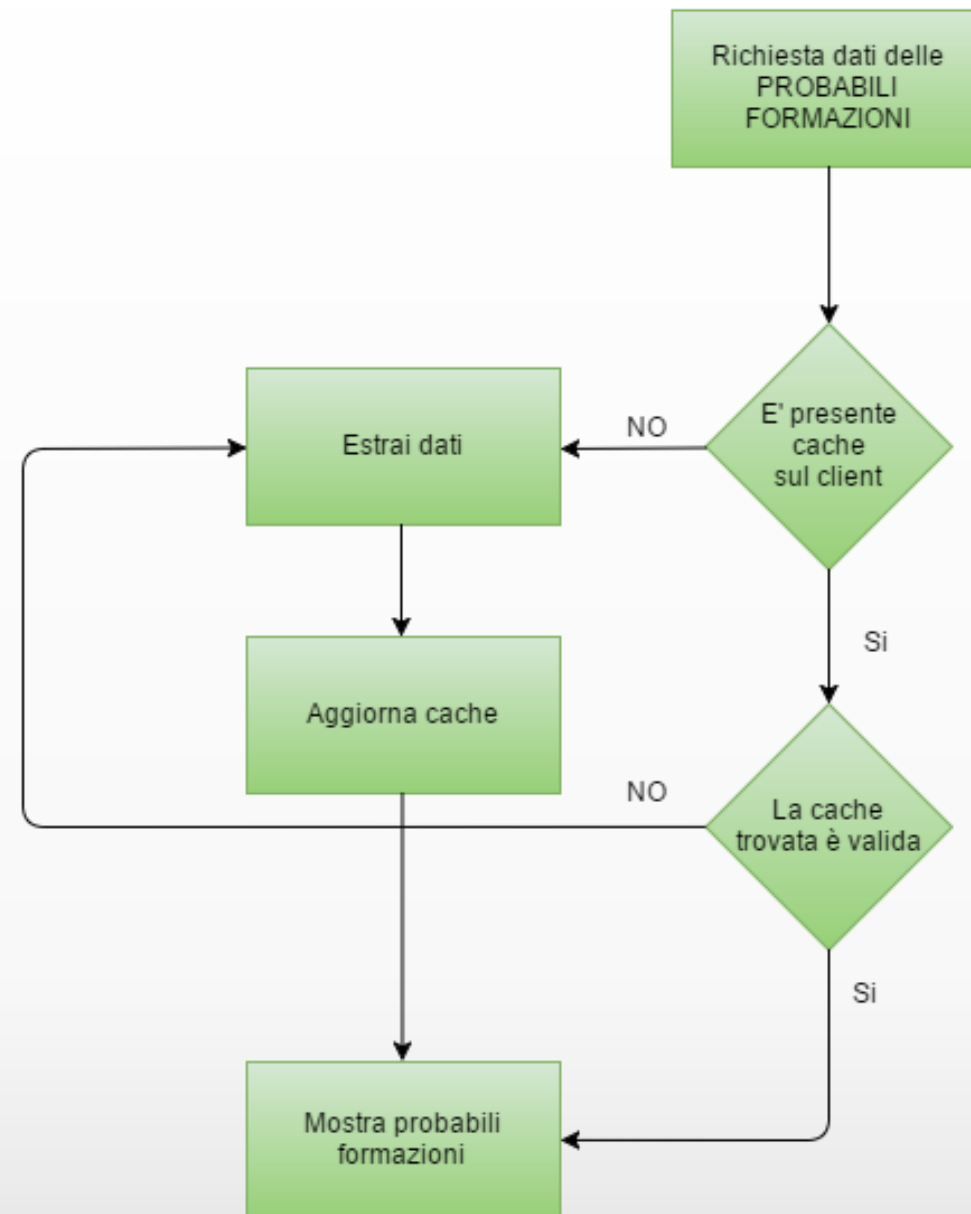


Probabili formazioni

PROBABILI FORMAZIONI



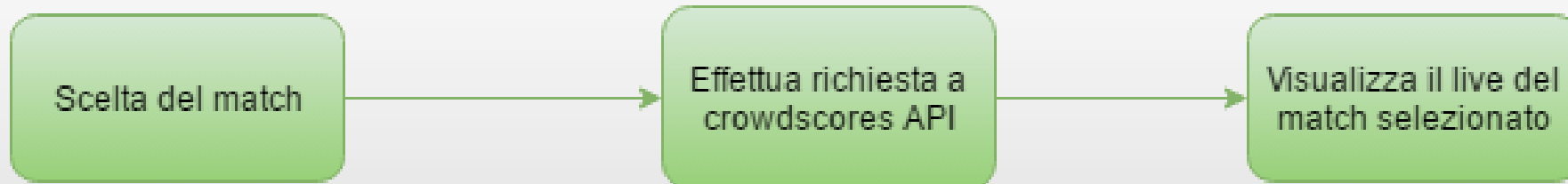
-  I dati sono stati estratti da fantagazzetta.it mediante WebScraping eseguito con appositi script PHP.
-  Le informazioni sulle probabili formazioni sono memorizzate su un **database NOSQL** sul *server* e in locale sul *client* (**localStorage**) quando la pagina è visitata.



Live

Juventus	3-0	Sampdoria
5' 🟡 Patrice Evra		
		13' 🔴 Milan Škriniar
15' 🟡 Paulo Dybala		
20' 🟡 Giorgio Chiellini		
34' 🟡 Hernanes		
37' 🟡 Paulo Dybala		

- 📁 I dati sono aggiornati in real-time(ogni 15 secondi).
- 📁 I dati sono recuperati mediante le API di CrowdScores.
- 📁 Per tale pagina non sono previsti meccanismi di caching.



Calendario (1)

CALENDARIO



1ª GIORNATA 23/08/2015							
	Hellas Verona	1	-	1	Roma		
	Lazio	2	-	1	Bologna		
	Juventus	0	-	1	Udinese		
	Empoli	1	-	3	Chievo Verona		
	Fiorentina	2	-	0	Milan		
	Frosinone	1	-	2	Torino		
	Inter	1	-	0	Atalanta		
	Palermo	1	-	0	Genoa		
	Sampdoria	5	-	2	Carpi		
	Sassuolo	2	-	1	Napoli		

Per ogni giornata di campionato è stata creata una tabella:

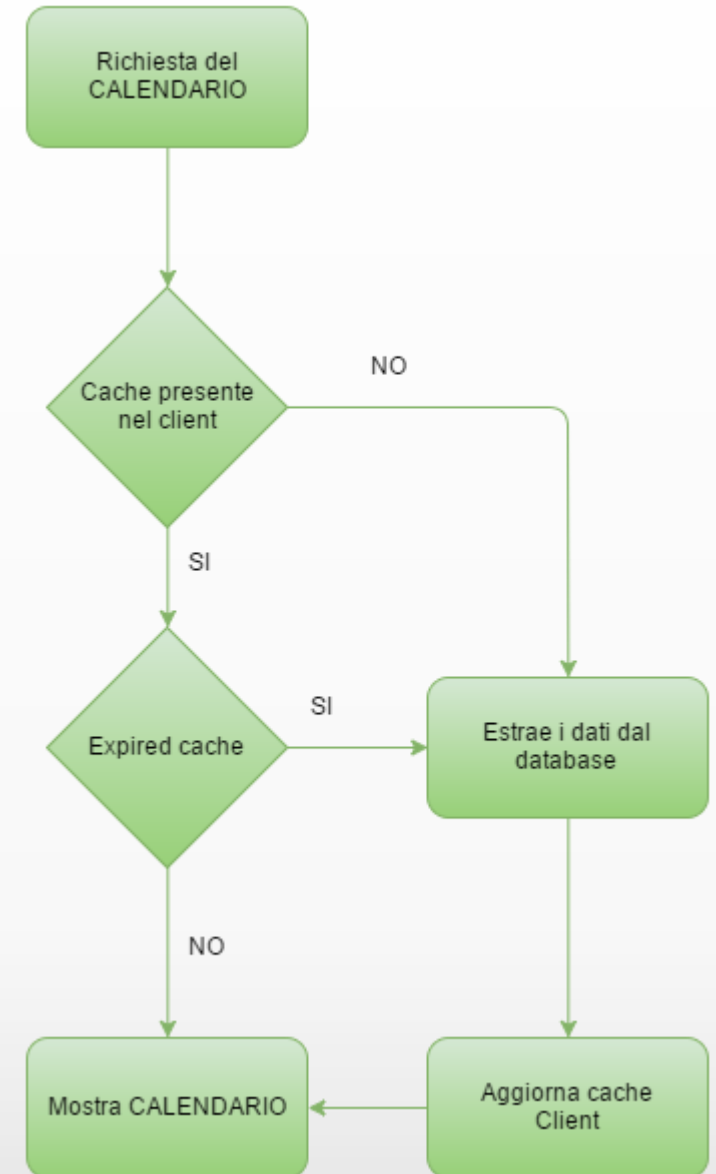
- **Intestazione:** numero giornata + data giornata.
- **Riga:** scudetto team1 + nome team1 + score + nome team2 + scudetto team 2 + pulsante informazioni o pulsante live.

Calendario (2)



CALENDARIO



- ⚽ I dati sono stati estratti da **LegaSerieA** mediante WebScraping eseguito con appositi script PHP.
- ⚽ Il calendario è memorizzato su un **database NOSQL** sul *server* e in locale sul *client* (**localStorage**) quando la pagina è visitata.
- ⚽ I dati sono aggiornati mediante uno script PHP di update invocato ogni 24 ore dal sito cron-job.org.



Informazioni match

-  I dati sono stati estratti da **LegaSerieA** mediante WebScraping eseguito con appositi script PHP.
-  Per tale funzionalità non è previsto nessun meccanismo di caching.



STAGIONE 2015-16 | 27^a GIORNATA



28/02/2016

STADIO: LUIGI FERRARIS - GENOVA

ARBITRO: DANIELE ORSATO

DATI COMPLETI DELLA PARTITA

ALLENATORI

SAMPDORIA : Vincenzo Montella

FROSINONE : Roberto Stellone

GOAL SEGNATI

SAMPDORIA : Fernando 44' Fabio Quagliarella 69'

POSSESSO PALLA

SAMPDORIA : 56%

FROSINONE : 44%

TIRI TOTALI

SAMPDORIA : 13

FROSINONE : 8

TIRI IN PORTA

SAMPDORIA : 7 (di cui 2 GOL)

Quote Serie A

 I dati sono stati estratti da **WilliamHill** mediante WebScraping eseguito con appositi script PHP.

 Per tale funzionalità non è previsto nessun meccanismo di caching.

Match					1	X	2
	Juventus	-	Sampdoria		1.28	5.50	10.00
	Sassuolo	-	Inter		1.70	4.00	4.50
	Milan	-	Roma		3.10	3.40	2.25
	Napoli	-	Frosinone		1.10	8.50	26.00
	Genoa	-	Atalanta		1.95	3.30	4.00
	Chievo Verona	-	Bologna		2.20	3.10	3.50
	Empoli	-	Torino		2.62	3.10	2.80
	Lazio	-	Fiorentina		2.25	3.30	3.20

Richiesta Quote

Effettua scraping su
WilliamHill

Visualizza le
informazioni
estratte

Local As View

-

Global As View

Descrizione schemi locali-1

giornate(numero_giornata, data_giornata, stadio, arbitro, diretta, team1, punteggio1, team2, punteggio2)

matchReport(team1, team2, punteggio1, punteggio2, allenatore1, allenatore2, formazione1, formazione2, panchina1, panchina2, possesso1, possesso2, tiri_in_porta1, tiri_in_porta2, tiri_fuori_porta1, tiri_fuori_porta2, parate1, parate2, pali1, pali2, fuorigioco1, fuorigioco2, ammonizioni1, ammonizioni2, corner1, corner2, occasioni1, occasioni2, assist1, assist2)

Info_club(id_club, nome_società, sede_società, stadio, contatti, sito_web, presidente, vice_presidente, consigliere_delegato, head_of_operation_sales_marketing, direttore_amministrativo, direttore_sportivo, direttore_comunicazioni, direttori_processi_amministrativi, segretario_sportivo, team_manager, addetto_stampa, odv, presidente_collegio_sindacale, sindaci_collegio, sindaci_supplenti, societa_revisione)

club_storia(stagione, competizione, posizione, punti, partite, vittorie, pareggi, sconfitte)

staff(allenatore, tattico, preparatore_atletico, preparatore_portieri)

rosa(giocatore, data_nascita, ruolo, nazionalita, partite_giocate, c_rosso, c_giallo, crg)

Descrizione schemi locali-2

WilliamHill.quotes(team1, team2, quota1, quotaX, quota2)

FastestLiveScores.matchEvents(competitionId, matchId, typeEvent, matchTime, nameRound, homeTeam, awayTeam, season, homePlayer, awayPlayer)

FastestLiveScores.matchList(competitionId, matchId, start, homeGoal, awayGoal, season, homeTeam, awayTeam, venue, outcome, round)

GoogleMaps(formatted_address, address_components, latitudine, longitudine, place_id)

Wikipedia(storia_club, cronistoria_club, colori_simboli, strutture, info_societa, staff, palmares, stats_record, tifoseria)

ProbabiliFormazioni(squalificati, in_dubbio, indisponibili, titolari_squadra_in_casa, ruoli_titolari_in_casa, percentuali_titolari_in_casa, panchinari_in_casa, ruoli_panchinari_in_casa, percentuali_panchinari_in_casa, titolari_squadra_fuori_casa, ruoli_titolari_fuori_casa, percentuali_titolari_fuori_casa, panchinari_fuori_casa, ruoli_panchinari_fuori_casa, percentuali_panchinari_fuori_casa, modulo_casa, modulo_fuori_casa, arbitro, data, stadio, ballottaggi, diffidati)

Schema globale

calendario (numero_giornata, data_giornata, stadio, arbitro, diretta, team1, punteggio1, team2, punteggio2)

informazioniMatch (team1, team2, punteggio1, punteggio2, allenatore1, allenatore2, possesso1, possesso2, tiri_in_porta1, tiri_in_porta2, tiri_fuori_porta1, tiri_fuori_porta2, occasioni1, occasioni2, assist1, assist2)

quote(team1, team2, quota1, quotaX, quota2)

InformazioniSquadre(nome_squadra, storia, giocatore, data_nascita, ruolo, nazionalita, partite_giocate, c_rosso, c_giallo, crg, indirizzo_sede, palmares, allenatore, presidente, latitudine, longitudine)

ProbabiliFormazioni(squalificati, in_dubbio, indisponibili, titolari_squadra_in_casa, ruoli_titolari_in_casa, percentuali_titolari_in_casa, panchinari_in_casa, ruoli_panchinari_in_casa, percentuali_panchinari_in_casa, titolari_squadra_fuori_casa, ruoli_titolari_fuori_casa, percentuali_titolari_fuori_casa, panchinari_fuori_casa, ruoli_panchinari_fuori_casa, percentuali_panchinari_fuori_casa)

Live(id, tipo_evento, squadra_coinvolta, giocatore_coinvolto, minuto)

Wrapper-1

Calendario: http://www.legaseriea.it/it/serie-a-tim/calendario-e-risultati/2015-16/UNICO/UNI/<numero_giornata>

- **numero_giornata:** /html/body/main/div[1]/section[1]/section/h3
- **data_giornata:** /html/body/main/div[1]/section[1]/section/h3
- **team1:** /html/body/main/div[1]/section[1]/section/div[".<numeroTeam1>."]/div[2]/h4
- **punteggio1:** /html/body/main/div[1]/section[1]/section/div[".<numeroTeam1>."]/div[2]/span
- **team2:** /html/body/main/div[1]/section[1]/section/div[".<numeroTeam2>."]/div[3]/h4
- **punteggio2:** /html/body/main/div[1]/section[1]/section/div[".<numeroTeam2>."]/div[3]/span

Wrapper-2

InfoMatch : http://www.legaseriea.it/it/serie-a-tim/match-report/2015-16/UNICO/UNI/<numero_giornata>/<team1> <team2>

- team1: /html/body/main/div[1]/section/div[1]/h3[1]/span
- team2: /html/body/main/div[1]/section/div[1]/h3[2]/span
- punteggio1: /html/body/main/div[1]/section/div[1]/div[3]
- punteggio2: /html/body/main/div[1]/section/div[1]/div[5]
- allenatore1: /html/body/main/div[1]/section/section[3]/div[1]/table/tbody/tr/td[2]
- allenatore2: /html/body/main/div[1]/section/section[3]/div[2]/table/tbody/tr/td[2]

Quote: http://sports.williamhill.it/bet_ita/it/betting/t/321/Serie+A.html

- match: //table/tbody/tr/td/a/span
- quotazione: //table/tbody/tr/td/div/div/text()

Wrapper-3

Informazioni squadra: [GoogleCustomSearch API](#), [GoogleMaps API](#), [SerieA](#), [Wikipedia](#).

- **palmares:** `//*[@id="mw-content-text"]/table`
- **testo:** `//*[@id="mw-content-text"]/div[@id="toc"]/preceding-sibling::p`
- **rosa:** `//*[@id="rosa-completa"]/table[1]/tr`
- **allenatore:** `/html/body/main/div[1]/section[1]/section/table/tr[1]/td[2]`
- **scudetto:** `/html/body/main/div[1]/section[1]/section/aside/section[1]/img/@src`
- **indirizzo:** `/html/body/main/div[1]/section[1]/section/aside/section[2]/article[1]/p[1]/text()`
- **presidente:** `/html/body/main/div[1]/section[1]/section/table/tr[1]/td[2]/text()`
- **GoogleSearchAPI:** `https://www.googleapis.com/customsearch/v1?key= <Google API Key> &cx= <cx Code> &q= <squadra> +calcio+wiki`
- **GoogleMapsAPI:** `http://maps.google.com/maps/api/geocode/json?address= <indirizzo>`

Wrapper-4

Probabili formazioni: <http://www.fantagazzetta.com/probabili-formazioni-serie-a>

- info match: `//*[contains(@class,"tab-pane fade") and contains(@class,"in no-gutter")]`
- squalificati – in_dubbio – indisponibili : `//*[div[@class="pgroup"]`
- titolari_squadra_in_casa: `//*[div/div[2]/div[3]/div[@class="col-lg-6 col-md-6 col-sm-6 col-xs-6 probbar p10"]/div[@class="pgroup lf"]/div[2]/div[1]/a`
- ruoli_titolari_in_casa: `//*[div/div[2]/div[3]/div[@class="col-lg-6 col-md-6 col-sm-6 col-xs-6 probbar p10"]/div[@class="pgroup lf"]//*[span[contains(@class,"role")]`
- percentuali_titolari_in_casa: `//*[div/div[2]/div[3]/div[@class="col-lg-6 col-md-6 col-sm-6 col-xs-6 probbar p10"]/div[@class="pgroup lf"]/div[2]/div[@class="tdrole"]//*[span[@class="perc"]`
- panchinari_in_casa: `//*[div[@class="col-lg-12 col-md-12 col-sm-12 col-xs-12"]//*[div[@class="pgroup lf"]/div[2]/div[2]/a`
- ruoli_panchinari_in_casa: `//*[div[@class="col-lg-12 col-md-12 col-sm-12 col-xs-12"]//*[div[@class="pgroup lf"]//*[span[contains(@class,"role")]`
- percentuali_panchinari_in_casa: `//*[div[@class="col-lg-12 col-md-12 col-sm-12 col-xs-12"]//*[div[@class="pgroup lf"]//*[div[contains(@class,"is")]`

Le xPath delle squadre fuori casa sono identiche eccezion fatta per la classe "lf" che deve essere sostituita con "rt"

Wrapper-5

Live partite: <https://docs.crowdscores.com/>

- **Match:** [http://api.crowdscores.com/api/v1/matches?round_ids= <ID round> &competition_id= <ID competizione> &api_key= <API key>](http://api.crowdscores.com/api/v1/matches?round_ids=<ID round>&competition_id=<ID competizione>&api_key=<API key>)
- **Match list:** http://api.crowdscores.com/api/v1/matches/<ID match>?api_key= < API key>

Mapping GAV-1

```
calendario(numero_giornata, data_giornata, team1, punteggio1, team2, punteggio2) :-  
giornate(numero_giornata, data_giornata, _, _, _, team1, punteggio1, team2, punteggio2)
```

```
informazioniMatch (team1, team2, punteggio1, punteggio2, allenatore1, allenatore2, possesso1,  
possesso2, tiri_in_porta1, tiri_in_porta2, tiri_fuori_porta1, tiri_fuori_porta2, occasioni1, occasioni2,  
assist1, assist2) :- matchReport (team1, team2, punteggio1, punteggio2, allenatore1, allenatore2, _, _, _,  
, _, possesso1, possesso2, tiri_in_porta1, tiri_in_porta2, tiri_fuori_porta1, tiri_fuori_porta2, _, _, _, _, _,  
, _, _, _, _, occasioni1, occasioni2, assist1, assist2)
```

Mapping GAV-2

`quote(team1, team2, quota1, quotaX, quota2) :- WilliamHill.quotes(team1, team2, quota1, quotaX, quota2)`

`live(id, tipo_evento, squadra_coinvolta, giocatore_coinvolto, minuto) :- matchEvents(competitionId, matchId, typeEvent, matchTime, _, homeTeam, awayTeam, _, homePlayer, awayPlayer) ^ matchList(competitionId, matchId, _, _, _, _, _, _, _, _) ^ competitionId='idSerieA`

`ProbabiliFormazioni(squalificati, in_dubbio, indisponibili, titolari_squadra_in_casa, ruoli_titolari_in_casa, percentuali_titolari_in_casa, panchinari_in_casa, ruoli_panchinari_in_casa, percentuali_panchinari_in_casa, titolari_squadra_fuori_casa, ruoli_titolari_fuori_casa, percentuali_titolari_fuori_casa, panchinari_fuori_casa, ruoli_panchinari_fuori_casa, percentuali_panchinari_fuori_casa) :- ProbabiliFormazioni(squalificati, in_dubbio, indisponibili, titolari_squadra_in_casa, ruoli_titolari_in_casa, percentuali_titolari_in_casa, panchinari_in_casa, ruoli_panchinari_in_casa, percentuali_panchinari_in_casa, titolari_squadra_fuori_casa, ruoli_titolari_fuori_casa, percentuali_titolari_fuori_casa, panchinari_fuori_casa, ruoli_panchinari_fuori_casa, percentuali_panchinari_fuori_casa, _, _, _, _, _, _)`

Mapping GAV-3

InformazioniSquadre(nome_società, storia, giocatore, data_nascita, ruolo, nazionalita, partite_giocate, c_rosso, c_giallo, crg, indirizzo_sede, palmares, allenatore, presidente, longitudine, latitudine) :-
info_club (_, nome_società, indirizzo_sede, _, _, _, presidente, _, _, _, _, _, _, _, _, _, _, _, _) ^
GoogleMaps(_, _, latitudine, longitudine, _) ^ **rosa** (giocatore, data_nascita, ruolo, nazionalita, partite_giocate, c_rosso, c_giallo, crg) ^ **staff** (allenatore, _, _, _) ^ **Wikipedia**(storia, _, _, _, _, palmares, _, _)

Mapping LAV-1

giornate(numero_giornata, data_giornata, stadio, arbitro, diretta, team1, punteggio1, team2, punteggio2) :- **calendario**(numero_giornata, data_giornata, team1, punteggio1, team2, punteggio2)

matchReport(team1, team2, punteggio1, punteggio2, allenatore1, allenatore2, formazione1, formazione2, panchina1, panchina2, possesso1, possesso2, tiri_in_porta1, tiri_in_porta2, tiri_fuori_porta1, tiri_fuori_porta2, parate1, parate2, pali1, pali2, fuorigioco1, fuorigioco2, ammonizioni1, ammonizioni2, corner1, corner2, occassioni1, occasioni2, assist1, assist2) :- **informazioniMatch** (team1, team2, punteggio1, punteggio2, allenatore1, allenatore2, possesso1, possesso2, tiri_in_porta1, tiri_in_porta2, tiri_fuori_porta1, tiri_fuori_porta2, occassioni1, occasioni2, assist1, assist2)

WilliamHill.quotes(team1, team2, quota1, quotaX, quota2) :- **quote**(team1, team2, quota1, quotaX, quota2)

Mapping LAV -3

GoogleMaps(formatted_address, address_components, latitudine, longitudine, place_id) :-
InformazioniSquadre(_, _, _, _, _, _, _, _, _, _, _, _, _, _, latitudine, longitudine)

staff(allenatore, tattico, preparatore_atletico, preparatore_portieri) :- **InformazioniSquadre**(_, _, _, _, _,
_, _, _, _, _, allenatore, _, _, _)

Wikipedia(storia_club, cronistoria_club, colori_simboli, strutture, info_societa, staff, palmares,
stats_record, tifoseria) :- **InformazioniSquadre**(_, storia_club, _, _, _, _, _, _, _, palmares, _, _, _, _)

Query

1) Mostra la storia di una determinata squadra.

Q1(storia)

2) Dato un determinato match in corso mostrare i minuti dei goal effettuati da una squadra.

Q2(minuto)

3) Date due squadre mostrare le quotazioni del match tra esse.

Q3(quota1,quotaX, quota2)

Query in datalog e SQL-2

3) Date due squadre mostrare le quotazioni del match tra esse.

$Q3(\text{quota1}, \text{quotaX}, \text{quota2}) :- \text{quote}(\text{team1}, \text{team2}, \text{quota1}, \text{quotaX}, \text{quota2}) \wedge \text{team1} = \text{"napoli"} \wedge \text{team2} = \text{"milan"}$

`SELECT quota1, quota2, quotaX FROM quote WHERE squadra1='napoli' AND squadra2='milan'`

Riformulazione Query in GAV

Q3(quota1,quotaX, quota2) :- quote(team1, team2, quota1, quotaX, quota2) ^
team1='napoli' ^ team2="milan"

Applicando l'unfolding

Q3(quota1,quotaX, quota2) :- WilliamHill.quotes(team1, team2, quota1, quotaX, quota2) ^
team1=napoli' ^ team2='milan'

Riformulazione Query in LAV

Abbiamo scelto il **bucket algorithm**.

$Q3(\text{quota1}, \text{quotaX}, \text{quota2}) :- \text{quote}(\text{team1}, \text{team2}, \text{quota1}, \text{quotaX}, \text{quota2}) \wedge \text{team1} = \text{"napoli"} \wedge \text{team2} = \text{"milan"}$

1) Costruzione Bucket per ogni atomo:

Bucket_quote [WilliamHill.quotes(team1, team2, quota1, quotaX, quota2)]

2) Query candidata:

$Q3.1(\text{quota1}, \text{quotaX}, \text{quota2}) :- \text{WilliamHill.quotes}(\text{team1}, \text{team2}, \text{quota1}, \text{quotaX}, \text{quota2}) \wedge \text{team1} = \text{"napoli"} \wedge \text{team2} = \text{"milan"}$

3) Checking containment:

$Q3(\text{quota1}, \text{quotaX}, \text{quota2}) :- \text{quote}(\text{team1}, \text{team2}, \text{quota1}, \text{quotaX}, \text{quota2}) \wedge \text{team1} = \text{"napoli"} \wedge \text{team2} = \text{"milan"}$

LA QUERY OTTENUTA E' IDENTICA ALLA QUERY INIZIALE.

Tecnologie Utilizzate

❑ HTML5



❑ JSON



❑ PHP



❑ MONGODB



❑ JAVASCRIPT



❑ GOOGLE CUSTOM SEARCH



❑ XPATH



❑ FASTEST LIVE SCORES API



❑ CSS



❑ GOOGLE MAPS



Grazie per
l'attenzione.