

# TECNOLOGIE SEMANTICHE RDFa

Prof. Sabrina Senatore

Rosario Frontino

MAT: 0622701484

Pierluigi Giangiacomi

MAT: 0622701454



## Il Problema: RDFa Lite

- RDFa: Resource Description Framework in attributes;
- RDFa Lite è un sottoinsieme di RDFa, che consiste di soli cinque attributi che possono essere usati per rendere alcuni dati, presenti nel web, processabili dalle macchine;
- Lo scopo del progetto è dunque quello di aggiungere RDFa Lite a un sito web. Il sito utilizzato riguarda una squadra di calcio, i suoi giocatori e la sua storia





## Il Problema: RDFa Lite

I cinque attributi previsti da RDFa Lite sono:

- vocab: definizione di un vocabolario in cui siano previste le keyword utili per descrivere il nostro dato (es: foaf per le persone);
- typeof: utile a specificare il tipo del dato;
- property: per specificare una proprietà particolare (es: nome o email per le persone);
- resource: utile a identificare una risorsa;
- prefix: definizione di vocabolari aggiuntivi.

## Il Problema: RDFa Lite

In questo esempio viene usato il vocabolario schema.org per annotare un dato di tipo Persona, le cui proprietà sono nome, numero di telefono e URL del sito della persona.

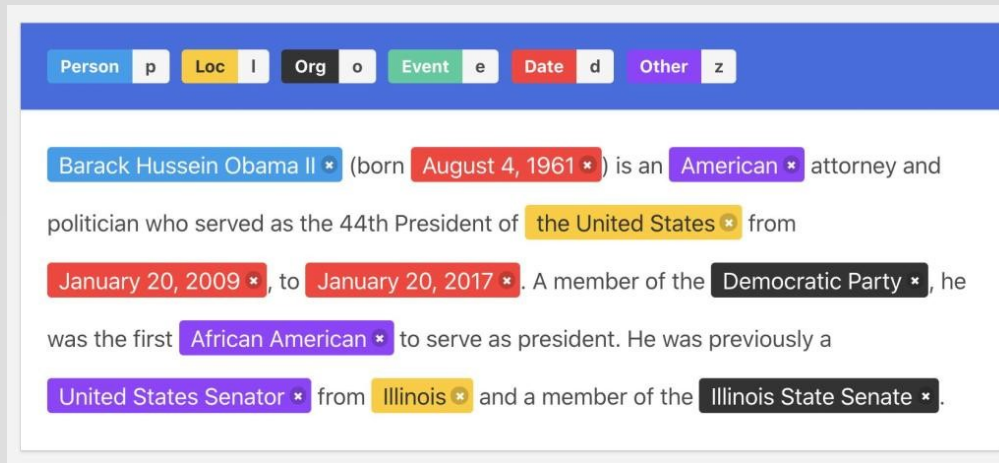
```
<p vocab="http://schema.org/" typeof="Person">
  My name is
  <span property="name">Manu Sporny</span>
  and you can give me a ring via
  <span property="telephone">1-800-555-0199</span>
  or visit
  <a property="url" href="http://manu.sporny.org/">my homepage</a>.
</p>
```

My name is Manu Sporny and you can give me a ring via 1-800-555-0199 or visit [my homepage](http://manu.sporny.org/).

RDFa è invisibile in output



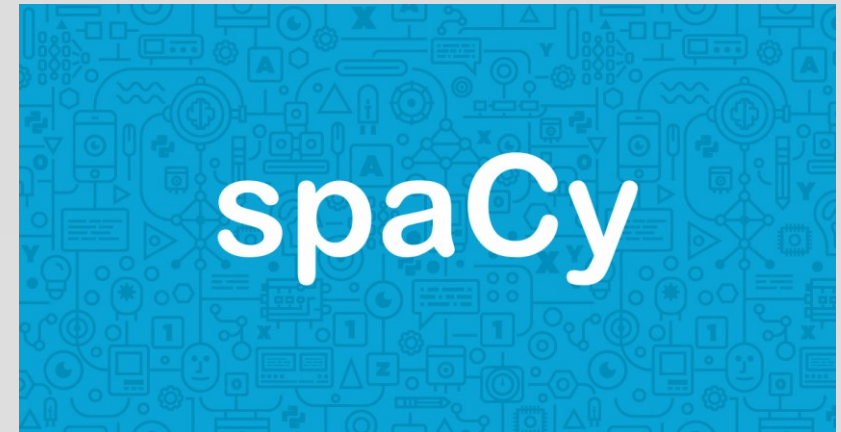
# Descrizione e Architettura: NLP e NER



- Il Natural Language Processing (NLP) è il processo che fornisce alle macchine la possibilità di “comprendere” un testo, in maniera simile a come lo fa l'uomo;
- La Named Entity Recognition (NER) è una sotto-applicazione del NLP che consiste nel fare in modo che, dato un testo, la macchina riconosca e identifichi le entità come persone, organizzazioni o luoghi, presenti nel testo.

# Implementazione: SpaCy

- Il NLP/NER utilizzato per la realizzazione del progetto è SpaCy, che in particolare è una libreria Python per il NLP, al cui interno sono presenti metodi per il NER;
- La libreria, inoltre, mette a disposizione un modello già addestrato per riconoscere entità di carattere generale come persone o luoghi, altrimenti andrebbe opportunamente addestrato per uno specifico caso di studio;
- SpaCy può essere impiegato in varie lingue: per il caso in esame è stato scelto l'italiano.



# Implementazione: dbpedia e SPARQL(Wrapper)



Le entità riconosciute da SpaCy vengono usate per ricavare il link della pagina Wikipedia a essi associati, tramite dbpedia e SPARQL

- dbpedia.org: è un progetto open source capace di estrarre, da Wikipedia, dati strutturati in RDF;
- SPARQL: linguaggio per interrogazioni per dati rappresentati in RDF;
- SPARQLWrapper: endpoint SPARQL con interfaccia in Python.



# Implementazione: algoritmo

Il codice per l'aggiunta delle annotazioni RDFa è così strutturato:

- Il file .html in input viene visto come testo, e su di esso si richiama il NLP/NER;
- Su tutto ciò che SpaCy riconosce come persona, luogo e organizzazione sportiva viene eseguita una query SPARQL su dbpedia.org. L'obiettivo di tale query è di recuperare l'URL della pagina Wikipedia associata all'entità corrente.
- ...





# Implementazione: algoritmo

- A questo punto è possibile annotare con RDFa: la parola corrente viene rimpiazzata con la concatenazione dei tag di apertura, dell'entità e dei tag di chiusura di RDFa. Il link di Wikipedia, laddove presente, è annotato con `rdfs:seeAlso` ed è inserito in un elemento vuoto.
- Ogni entità viene processata una e una sola volta, grazie alla lista di supporto “processed”.
- Il testo così ottenuto viene scritto su un nuovo file .html.

# Implementazione: RDFa

Le tre tipologie di entità sono state annotate come segue:

ENTITÀ	vocab	typeof	property
Persone	FOAF	Person	name
Luoghi	schema.org	Place	name
Organizzazioni Sportive	schema.org	SportsOrganization	name

Dove:

- ✓ FOAF (acronimo di "Friend of a friend") è un'ontologia riguardante le persone e permette di descriverle e legarle tra loro, tramite la proprietà foaf:knows;
- ✓ schema.org è un vocabolario semantico di tag quali, oltre ai due già citati, quelli per Eventi e Prodotti.



# Implementazione: query

Sia newWord un'entità rilevata dal NER, la query è così formata:

```
SELECT ?object  
      WHERE {{ dbr:{newWord} foaf:isPrimaryTopicOf ?object.}}
```

- Tutte le risorse presenti su dbpedia sono legate al loro URL di Wikipedia tramite foaf:isPrimaryTopicOf
- Questa proprietà lega una risorsa con un documento che riguarda (principalmente) la risorsa stessa

# Validazione e testing

Tutte le entità rilevate dal NER nelle varie pagine sono state correttamente annotate.

Per ogni entità, la trasformazione nel codice sorgente del sito è di questo tipo:

`<li> Leonardo Moracci </li>`



```
<li> <span
vocab="http://xmlns.com/foaf/0.1/"
typeof="Person"><span
property="name"><span
property="rdfs:seeAlso" =
"http://en.wikipedia.org/wiki/Leonardo_
Moracci"/> Leonardo Moracci
</span></span> </li>
```



# Validazione e testing

Come anticipato, RDFa è invisibile rispetto al sito di partenza, per cui ci aspettiamo che le pagine prima e dopo l'applicazione dell'algoritmo siano uguali.

## Origini del calcio a pagani

Nel 1912 Pasquale Carosella istituisce all'interno della sua società, l'Associazione Giovanile Paganese, con sezione calcistica e ciclistica, la Pagani Football Club, che non parteciperà mai ad alcun campionato ma disputerà incontri di carattere amichevole.



## I primi campionati

Nel 1926 nasce l'Unione Sportiva Paganese con presidente Giuseppe Baldanza. Una volta costituiti i quadri societari (1927), e con l'avvenuta affiliata alla FIGC (1928), alla neonata squadra se viene concesso il terreno per il campo da gioco. Prima della nascita della società, infatti, i calciatori di Pagani giocavano in una vasca di raccolta di acque piovane, e solo nel marzo 1928 fu concesso il terreno per il campo da gioco. Il primo campo sportivo nacque al centro della città, in corso Padovano, ove oggi si trova la Villa Comunale: fu il "Campo del Pino", così chiamato perché sorgeva al centro del paese, dove si ergeva un gigantesco pino (simbolo della città) piantato trecento anni prima nel giardino dell'allora convento delle Suore Carmelitane. Il primo allenatore fu Venturini, già difensore della Salernitana, che a Pagani ebbe il doppio ruolo di guida tecnica e giocatore. La squadra si piazzò al 40° posto nel campionato di Terza Divisione 1928-1929, e fu promossa in Seconda Divisione. La stagione 1929-30 deve essere annoverata tra le più importanti nella storia del calcio paganese. La formazione allenata dal tecnico Gherlani con presidente Costantino Astarita fu la grande sorpresa del campionato di Seconda Divisione 1929-1930 Girone A. Al termine di 18 combattutissime partite la Paganese si aggiudicò il girone superando squadre del calibro di Reggina, Catania e Cosenza. Grazie al primo posto nel girone Paganese, insieme alla Reggina, approdarono alle finali interregionali che avrebbero decretato la squadra promossa in Prima Divisione. La Paganese vi arrivò esausta, non disponendo di una rosa di calciatori molto ampia a disposizione e così dovette arrendersi in casa sia alla Reggina, che pure aveva battuto nella regular season per 2-0, sia al Molfetta. Il risultato portò lo scoramento tra i tifosi azzurri che con un duro comunicato stigmatizzarono il comportamento degli atleti azzurri rei a loro dire di scarso attaccamento ai colori sociali. Per ragioni economiche la Paganese quindi retrocedette in Terza Divisione Campania. Nella stagione 1932-1933 la Paganese sfiorò per un soffio la promozione in seconda divisione ma al termine di un'annata caratterizzata da alti e bassi, venne ripescata in Terza Divisione Campania.

storia.html

## Origini del calcio a pagani

Nel 1912 Pasquale Carosella istituisce all'interno della sua società, l'Associazione Giovanile Paganese, con sezione calcistica e ciclistica, la Pagani Football Club, che non parteciperà mai ad alcun campionato ma disputerà incontri di carattere amichevole.



## I primi campionati

Nel 1926 nasce l'Unione Sportiva Paganese con presidente Giuseppe Baldanza. Una volta costituiti i quadri societari (1927), e con l'avvenuta affiliata alla FIGC (1928), alla neonata squadra se viene concesso il terreno per il campo da gioco. Prima della nascita della società, infatti, i calciatori di Pagani giocavano in una vasca di raccolta di acque piovane, e solo nel marzo 1928 fu concesso il terreno per il campo da gioco. Il primo campo sportivo nacque al centro della città, in corso Padovano, ove oggi si trova la Villa Comunale: fu il "Campo del Pino", così chiamato perché sorgeva al centro del paese, dove si ergeva un gigantesco pino (simbolo della città) piantato trecento anni prima nel giardino dell'allora convento delle Suore Carmelitane. Il primo allenatore fu Venturini, già difensore della Salernitana, che a Pagani ebbe il doppio ruolo di guida tecnica e giocatore. La squadra si piazzò al 40° posto nel campionato di Terza Divisione 1928-1929, e fu promossa in Seconda Divisione. La stagione 1929-30 deve essere annoverata tra le più importanti nella storia del calcio paganese. La formazione allenata dal tecnico Gherlani con presidente Costantino Astarita fu la grande sorpresa del campionato di Seconda Divisione 1929-1930 Girone A. Al termine di 18 combattutissime partite la Paganese si aggiudicò il girone superando squadre del calibro di Reggina, Catania e Cosenza. Grazie al primo posto nel girone Paganese, insieme alla Reggina, approdarono alle finali interregionali che avrebbero decretato la squadra promossa in Prima Divisione. La Paganese vi arrivò esausta, non disponendo di una rosa di calciatori molto ampia a disposizione e così dovette arrendersi in casa sia alla Reggina, che pure aveva battuto nella regular season per 2-0, sia al Molfetta. Il risultato portò lo scoramento tra i tifosi azzurri che con un duro comunicato stigmatizzarono il comportamento degli atleti azzurri rei a loro dire di scarso attaccamento ai colori sociali. Per ragioni economiche la Paganese quindi retrocedette in Terza Divisione Campania. Nella stagione 1932-1933 la Paganese sfiorò per un soffio la promozione in seconda divisione ma al termine di un'annata caratterizzata da alti e bassi, venne ripescata in Terza Divisione Campania.

storiaOutput.html

GRAZIE PER L'ATTENZIONE