DIGITAL FUTURE
OR13 – Piattaforma di Telemedicina: prototipi dimostrativi

Prototipo dimostrativo: Modulo Analytics

PD13.3

AUTORE: EXPRIVIA S.P.A

UNIONE EUROPEA
FONDO EUROPEO DI SVILUPPO REGIONALE

Ministero Sviluppo Economico

REGIONE PUGLIA
AREA POLITICHE PER LO SVILUPPO ECONOMICO
IL LAVORO E L'INNOVAZIONE

Regolamento regionale della Puglia per gli aiuti in esenzione n. 17 del 30/09/2014

Titolo II – Capo 1 "Aiuti ai programmi di investimento delle Grandi Imprese"

POR PUGLIA FESR 2014 – 2020

CONTRATTO DI PROGRAMMA

"Digital Future"

CODICE PROGETTO: CP12PA6

Diritti di autore e riservatezza

Questo documento è proprietà esclusiva della società Exprivia S.p.A. e non può essere riprodotto, anche in forma parziale, senza un'autorizzazione scritta della società stessa.

Indice dei contenuti

INDICE DEI CONTENUTI **4**

INDICE DELLE FIGURE **6**

INDICE DELLE TABELLE **7**

1. INTRODUZIONE **8**

2. ANALYTICS NELL'INTERNET OF MEDICAL THINGS **10**

2.1. ARCHITETTURA GENERALE 10

2.2. DEFINIZIONE DELLO SCENARIO D'USO 14

2.3. CLASSIFICAZIONE DEL RISCHIO CLINICO MEDIANTE APPRENDIMENTO AUTOMATICO 17

2.4. EXPLANATION DEI RISULTATI DI ML MEDIANTE RAPPRESENTAZIONE DELLA CONOSCENZA E RAGIONAMENTO AUTOMATICO TRAMITE ONTOLOGIE 24

2.5. COMUNICAZIONE CON IL MIDDLEWARE DELLA PIATTAFORMA DIGITAL FUTURE 26

3. CLASSIFICAZIONE DEI REFERTI CLINICI **30**

3.1. INTRODUZIONE E ANALISI DEL DOMINIO ..... **ERRORE. IL SEGNALIBRO NON È DEFINITO.**

3.2. TECNICHE PER LA CLASSIFICAZIONE DEI DOCUMENTI .... **ERRORE. IL SEGNALIBRO NON È DEFINITO.**

3.2.1. Le Ontologie ..... ***Errore. Il segnalibro non è definito.***

3.2.2. Information Extraction (IE) ..... ***Errore. Il segnalibro non è definito.***

3.2.2.1. Architettura di riferimento ..... **Errore. Il segnalibro non è definito.**

3.2.2.2. Modulo di pre-processamento di documenti .... **Errore. Il segnalibro non è definito.**

3.2.2.3. Named Entity Recognition ..... **Errore. Il segnalibro non è definito.**

3.3. ANNOTAZIONE SEMANTICA DEI REFERTI MEDICI 30

3.3.1. Approccio integrato per il trattamento dei referti medici in lingua italiana ... *30*

3.3.2. Approccio per l'annotazione dei referti medici in inglese *33*

3.4. OPENSEMANTICSEARCH 34

3.4.1. Componenti principali e architettura *36*

3.4.2. Flusso operativo *44*

3.5. CASI D'USO 45

3.5.1. Caso d'uso 1- Annotazione e classificazione referti Covid19 ..... *45*

3.5.2. Caso d'uso 2- Annotazione e classificazione referti medici in italiano 49

3.5.3. Graph visualization 50

Indice delle figure

Figura 1 - Architettura Edge 10
Figura 2 - Raspberry Pi 3B Model 11
Figura 3 - Configurazione del servizio Bluetooth nel RPi 11
Figura 4 - Creazione del servizio di Advertising Bluetooth 12
Figura 5 - Edge settings.ini 12
Figura 6 - Simulazione dei dispositivi medicali che inviano i dati dei parametri vitali via BlueTooth 14
Figura 7 - Acquisizione dei dispositivi medicali collegati via BlueTooth al nodo Edge .. 14
Figura 8 - Ricezione dei parametri vitali dai dispositivi medicali 17
Figura 9 - Data Science lifecycle 18
Figura 10 – Stralcio del Dataset iniziale 19
Figura 11 – Stralcio del Dataset finale 20
Figura 12 - Visualizzazione dei risultati delle metriche di performance dei modelli di ML 23
Figura 13 - Esempio di avvenuta predizione on Edge del Rischio Clinico 23
Figura 14 - Rappresentazione tassonomica dell'Ontologia di riferimento per lo Scenario d'Uso 25
Figura 15 – Esempio di avvenuta Explanation dei risultati di ML tramite Ontologia 26
Figura 16 - Esempio Tabella Dati registati su DF in Cassandra 28
Figura 17 - Notifica su RPi di invio dati a piattaforma 28
Figura 18 - Notifica di ricezione dati in piattaforma 29
Figura 19 - Intero processo AI on Edge su RPi 29
Figura 20 - (a) Ontologia relativa alle patologie del sistema immunitario; (b) Ontologia relativa ad un flusso ospedaliero ..... **Errore. Il segnalibro non è definito.**

Indice delle tabelle

Tabella 1 - Assegnazione punteggi per il calcolo della classe di rischio secondo l'algoritmo NEWS2 15
Tabella 2 – Selezione del miglior modello di ML 22

1. Introduzione

La piattaforma di Telemedicina sviluppata presenta un sistema che contiene l'insieme di soluzioni e servizi forniti da **Exprivia** per il teleconsulto, il telereporting, la telepresenza e il telemonitoraggio per supportare tutti gli operatori coinvolti nella cura e nel monitoraggio del paziente a casa. La piattaforma rivoluziona l'approccio alla cura del paziente a domicilio in quanto fornisce l'infrastruttura tecnologica Edge Computing e tutti i servizi necessari per la gestione completa e integrata di tutti i processi e servizi di cura, accessibili e utilizzabili da qualsiasi tipo di dispositivo IoT:

- Monitoraggio remoto e rilevamento dei parametri vitali;
- Monitoraggio dell'approvvigionamento di medicinali;
- Sistemi di segnalazione e archiviazione;
- Anamnesi del paziente in cartelle elettroniche.

La piattaforma affronta la nuova sfida di Digital Health che pone il paziente al centro e garantisce servizi continui che migliorano la qualità della vita e, allo stesso tempo, aiutano a limitare i costi delle autorità sanitarie e degli ospedali locali rimuovendo il paziente dall'ospedale. La piattaforma si basa su un'applicazione Web utilizzata per monitorare e gestire tutti i pazienti in tempo reale, fungendo da intermediario operativo tra il paziente e il team medico o lo specialista che segue il paziente.

Focus del presente documento è il **Modulo Analytics della Piattaforma**, che fa fede ad una architettura full-Edge consentendo al team medico di eseguire decisioni supportate dall'Intelligenza Artificiale (IA) per un singolo paziente a casa, in base ai dati della sua cartella clinica su trattamento, parametri vitali, diagnosi, ecc.

Il sistema di Analytics della Piattaforma è incentrato sul conseguimento delle seguenti attività di impiego della piattaforma di telemedicina:

- *Modulo di apprendimento automatico* (Machine Learning, ML): tecnica di IA che consente la definizione di un modello derivante dei fenomeni della realtà. I modelli che ne derivano eseguono task di classificazione, inferenza e

rilevamento delle anomalie. Il sistema in questione è in grado di apprendere e codificare lo stato di salute del paziente.

- *Rappresentazione della conoscenza e ragionamento automatico tramite Ontologie*: definizione di una concettualizzazione tassonomica dei parametri vitali volti a descrivere ed interpretare lo stato di salute di un paziente.
- *Information Retrieval*: è l'insieme delle tecniche utilizzate per gestire la rappresentazione, la memorizzazione, l'organizzazione e l'accesso ad oggetti contenenti informazioni quali documenti, pagine web, cataloghi online e oggetti multimediali. La ricerca semantica utilizzerà i tag definiti per le relazioni ontologiche per ricercare dati non strutturati per sinonimia e specificazione dei termini.
- *Reportistica dinamica*: sugli stati di salute sarà fornito un sistema di reportistica dinamica che permetterà l'analisi puntuale, aggregata e storica degli stati di salute del paziente.

In particolare, in questo documento verranno descritte due specifiche attività del Modulo di Analytics:

1. **Capitolo 2**: La componente di Artificial Intelligence su tecnologia di Edge Computing (**AI on Edge**) della piattaforma raccoglie dati dai dispositivi ed esegue attività di classificazione supervisionata in Machine Learning del rischio clinico del paziente, monitorando i parametri vitali del paziente e dotando i dati raccolti da sensori di piccoli frammenti di annotazioni semantiche dedotte da ontologie mediche sui nodi periferici al fine di dare una spiegazione in linguaggio naturale dei risultati di classificazione ottenuti tramite Machine Learning.
2. **Capitolo 3**: La componente di indicizzazione tramite un motore di ricerca (Solr), metadatazione mediante due ontologie Upper Level ed estrazione terminologica dei referti medici. Tale componente permette di aiutare sia la fase di reperimento di referti afferenti a particolari patologie e sia alla loro leggibilità, infatti la metadatazione tramite ontologie condivise internazionalmente permette di meglio focalizzarsi sui contenuti rilevanti.

2. Analytics nell'Internet of Medical Things

2.1. Architettura Generale

Riportiamo di seguito l'architettura del Modulo di Analytics della Piattaforma di Telemedicina in modo che l'elaborazione possa essere eseguita sui dispositivi (ad es. nodi terminali) o sui gateway (ovvero nodi di bordo rete). Ciò ridurrà il traffico di dati e la latenza di elaborazione non necessari ed è importante per applicazioni come il monitoraggio e l'analisi dei pazienti critici. L'attenzione di seguito si concentra sulla progettazione e comunicazione con i nodi periferici e con nodi terminali associati per l'attività di monitoraggio dei parametri vitali.

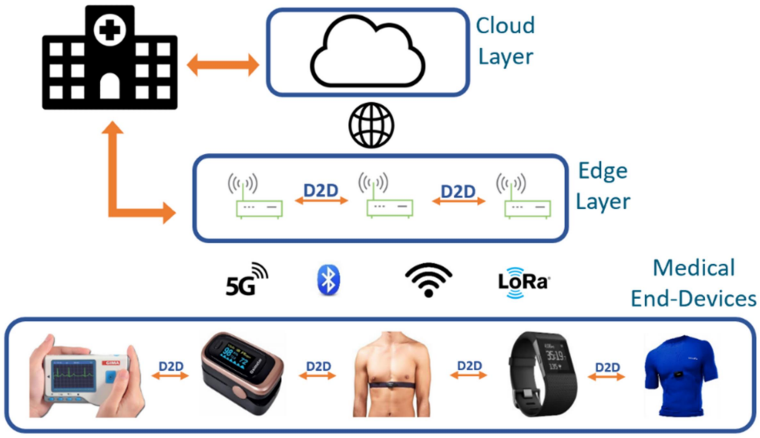


Figura 1 - Architettura Edge

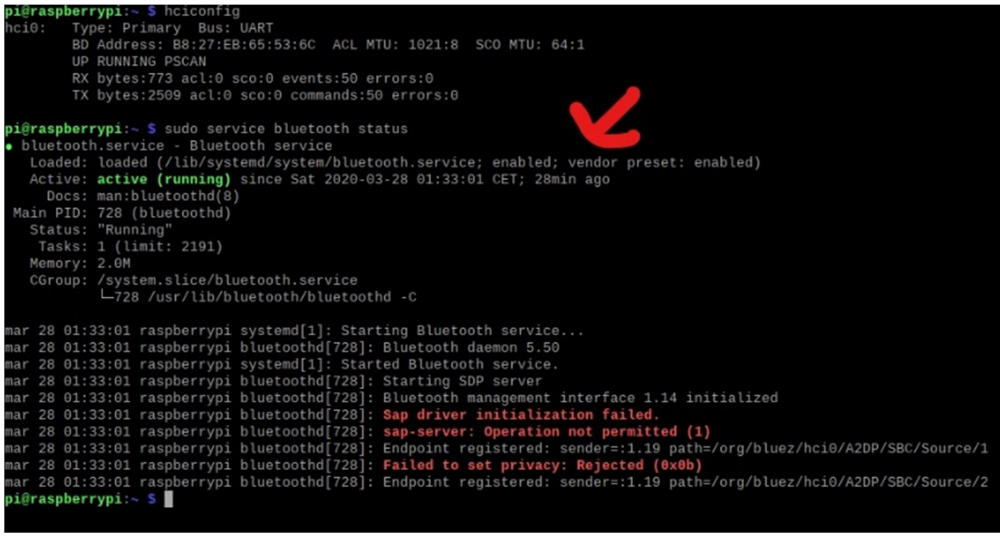
Qui un nodo perimetrale può trovarsi nelle vicinanze di un dispositivo collegabile tramite comunicazione dispositivo-a-dispositivo (D2D), un server collegato a un punto di accesso (ad es. WiFi, router, stazione base), un gateway di rete o persino un centro di micro-dati disponibile per l'uso da dispositivi vicini. Concentrandosi sull'intelligenza artificiale, immaginiamo una rete di Edge Computing in grado di eseguire non solo analisi, classificazione, regressione e / o clustering dei dati tramite formazione e inferenza online Machine Learning / Deep Learning, ma anche una più ampia gamma di tecniche di intelligenza artificiale, aggiungendo consapevolezza del contesto e spiegabilità/affidabilità (cfr. Explanation) dei risultati. In questa visione, i modelli Machine Learning / Deep Learning sono implementati in una modalità ibrida che

combina la modalità di addestramento e/o inferenza decentralizzata con una modalità di revisione o perfezionamento centralizzata.

Il prototipo dimostrativo proposto mostra come i componenti di ogni singolo livello dell'architettura comunichino con i dispositivi dei livelli adiacenti. L'Edge node utilizzato è il single-board computer **Raspberry Pi 3B**, denominato d'ora in poi **RPi**. Grazie alle caratteristiche Hardware e Software ad esse incorporate, il RPi comunica con i dispositivi medicali tramite tecnologia Bluetooth, in modo da ricercare i dispositivi coperti da segnale radio entro un ravvicinato raggio di azione e, garantire lo scambio di informazioni in maniera rapida e sicura.

Figura 2 - Raspberry Pi 3B Model

Prima di poter eseguire il processo di Edge Computing, è necessaria una fase preliminare di configurazione ed abilitazione Bluetooth sul sistema operativo **Raspbian**. Successivamente, si esegue il Pairing con i dispositivi medicali e stabilire con essi, un reciproco livello di Trustworthy.



```
pi@raspberrypi:~ $ hciconfig
hci0:   Type: Primary  Bus: UART
        BD Address: B8:27:EB:65:53:6C  ACL MTU: 1021:8  SCO MTU: 64:1
        UP RUNNING PSCAN
        RX bytes:773 acl:0 sco:0 events:50 errors:0
        TX bytes:2509 acl:0 sco:0 commands:50 errors:0

pi@raspberrypi:~ $ sudo service bluetooth status
● bluetooth.service - Bluetooth service
   Loaded: loaded (/lib/systemd/system/bluetooth.service; enabled; vendor preset: enabled)
   Active: active (running) since Sat 2020-03-28 01:33:01 CET; 28min ago
     Docs: man:bluetoothd(8)
 Main PID: 728 (bluetoothd)
   Status: "Running"
    Tasks: 1 (limit: 2191)
   Memory: 2.0M
   CGroup: /system.slice/bluetooth.service
           └─728 /usr/lib/bluetooth/bluetoothd -C

mar 28 01:33:01 raspberrypi systemd[1]: Starting Bluetooth service...
mar 28 01:33:01 raspberrypi bluetoothd[728]: Bluetooth daemon 5.50
mar 28 01:33:01 raspberrypi systemd[1]: Started Bluetooth service.
mar 28 01:33:01 raspberrypi bluetoothd[728]: Starting SDP server
mar 28 01:33:01 raspberrypi bluetoothd[728]: Bluetooth management interface 1.14 initialized
mar 28 01:33:01 raspberrypi bluetoothd[728]: Sap driver initialization failed.
mar 28 01:33:01 raspberrypi bluetoothd[728]: sap-server: Operation not permitted (1)
mar 28 01:33:01 raspberrypi bluetoothd[728]: Endpoint registered: sender=:1.19 path=/org/bluez/hci0/A2DP/SBC/Source/1
mar 28 01:33:01 raspberrypi bluetoothd[728]: Failed to set privacy: Rejected (0x0b)
mar 28 01:33:01 raspberrypi bluetoothd[728]: Endpoint registered: sender=:1.19 path=/org/bluez/hci0/A2DP/SBC/Source/2
pi@raspberrypi:~ $
```

Figura 3 - Configurazione del servizio Bluetooth nel RPi

```
GNU nano 3.2                                                bluetooth_adv

hciconfig hci0 piscan
sdptool add SP
```

Figura 4 - Creazione del servizio di Advertising Bluetooth

Al fine di garantire la massima esecuzione del caso d'uso, l'applicazione mobile "BlueTooth Serial Controller", installata su smartphone Android, ci permette di simulare il comportamento dei dispositivi medicali. Di seguito, per convenzione, i dispositivi mobile verranno contestualizzati in dispositivi medicali.

La tipologia di comunicazione stabilita tra componenti Edge e dispositivi medicali è di tipo **Client-Server** dove i dispositivi medicali fungono da Client mentre il RPi da Server. I dispositivi medicali dispongono in maniera autonoma di tutte le risorse e le strutture necessarie tipiche di un nodo Client.

Come fase preliminare, è stato creato all'interno del RPi il file di impostazione che consente di memorizzare le modalità di funzionamento. I parametri necessari atti all'esecuzione del modulo Edge Computing sono:

- Numero di dispositivi medicali coinvolti;
- Numero totale di parametri vitali;
- Numero minimo di parametri vitali registrabili in una finestra temporale affinché possa avvenire la predizione;
- Timing di ampiezza della finestra temporale entro il quale si raccolgono il maggior numero di parametri vitali distinti. Il timing dipende dal processo definito dai clinici che governano la reale misurazione dei dati del paziente.

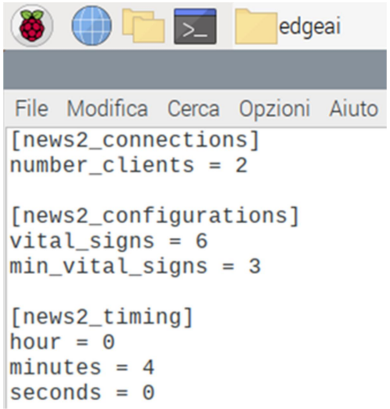


Figura 5 - Edge settings.ini

All'interno del RPi sono stati implementati gli algoritmi che definiscono la soluzione AI on Edge. Il linguaggio di programmazione adottato è stato Python 3.7, un linguaggio flessibile per l'integrazione di una soluzione di Machine Learning in un contesto applicativo come l'IoMT. Le librerie adottate sono le seguenti:

- Configparser, consente di lavorare con file di configurazione in formato .ini;
- PyBluez, che permette la comunicazione tra dispositivi via Bluetooth;
- Threading per definire sottoprocessi in grado di eseguire più operazioni in maniera concorrente, entrando così nell'ambito del Multithreading;
- Time, per la gestione del timing temporale;
- Re, per la creazione di espressioni regolari utili a definire pattern capace di distinguere e disambiguare una serie di simboli ricevuti in input;
- Pickle, modulo che ci consente di salvare e caricare modelli predittivi di Machine Learning;
- Numpy e Pandas per la memorizzazione ed interrogazioni di strutture dati apposite quali DataFrame e Array in grado di gestire enormi mole di dati che soddisfano le caratteristiche tipiche dei Big Data;
- Owlready2, che consente di modellare un'ontologia, caricarla direttamente da file in formato .owl ed interrogarla al fine di estrarre del contenuto informativo semantico;
- Paho.mqtt per definire una comunicazione di tipo MQTT;
- Json per la raccolta e la memorizzazione dei dati in formato JSON.

Il processo Edge Computing parte con il caricamento del file di impostazione e lettura dei parametri in esso contenuti. Si procede, successivamente con la fase di configurazione Bluetooth in cui viene creato, lato Server, il Bluetooth Socket che stabilisce l'endpoint di connessione Bluetooth, per poi eseguire il bind del Bluetooth Socket su una porta referenziata da Bluetooth ed infine per rimanere in attesa di tutti i dispositivi medicali coinvolti nel processo Edge Computing.

Una volta accettata la comunicazione tra i tutti i Client associati e il Server, viene eseguita la fase di condivisione e trasmissione dati, che verrà descritto accuratamente nel caso d'uso predisposto.

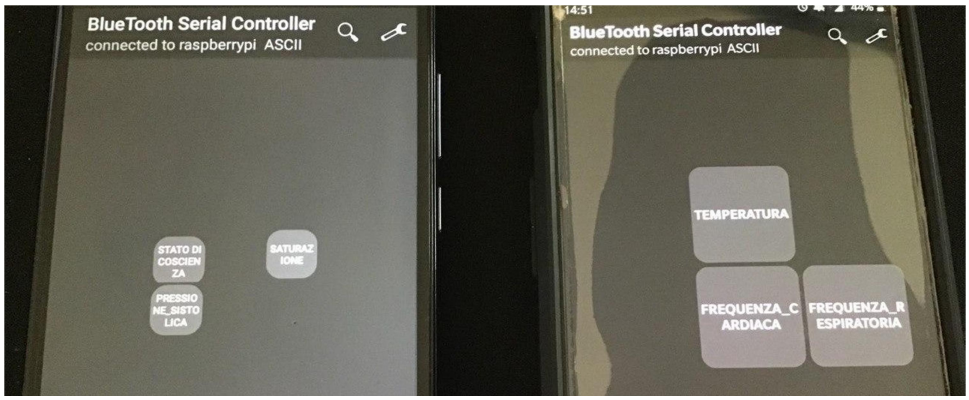


Figura 6 - Simulazione dei dispositivi medicali che inviano i dati dei parametri vitali via BlueTooth

Per ogni dispositivo medicale connesso, è associato ad un apposito Thread in maniera tale che il Server gestisca più dispositivi contemporaneamente e garantisca una condivisione simultanea delle misurazioni dei parametri vitali inviati dai Client. In caso di disconnessione di uno dei dispositivi medicali, il processo verrà eseguito normalmente, purché garantisca la registrazione di misurazioni minime dei parametri vitali consentite, altrimenti, in caso di disconnessione di tutti i dispositivi, il componente Edge rimarrà in ascolto in attesa dei dispositivi medicali.

```
pi@raspberrypi:~ $ cd edgeai
pi@raspberrypi:~/edgeai $ sudo ./bluetooth_adv
Serial Port service registered
pi@raspberrypi:~/edgeai $ sudo python3 blu_simulator_server.py

Avvio Processo AI on Edge for Digital Future

Acquisizione dei dispositivi medicali

In attesa di 2 dispositivi medicali..
Dispositivo medicale Medical Device 2 connesso

In attesa di 1 dispositivo medicale..
Dispositivo medicale Medical Device 3 connesso

Registrazione Parametri vitali del Paziente..
```

Figura 7 - Acquisizione dei dispositivi medicali collegati via BlueTooth al nodo Edge

Una volta definita la configurazione iniziale, è possibile eseguire il processo di Analytics sull'Internet of Medical Things (IoMT) attraverso il modulo di AI on Edge.

2.2. Definizione dello Scenario d'Uso

La possibilità di un monitoraggio continuo delle condizioni di salute rappresenta un aspetto cruciale per il miglioramento delle condizioni di vita, la prevenzione di potenziali

patologie e la pronta risposta in situazioni critiche. In particolare, in terapia intensiva o situazioni di emergenza, la valutazione del grado di malattia di un livello di rischio clinico può essere considerata un compito predittivo in situazioni in cui flussi di dati di segni vitali vengono raccolti da dispositivi medici e sensori di IoMT.

In questo caso d'uso si farà riferimento alla progettazione definita nel documento di Analisi dei Servizi di Piattaforma per l'Healthcare (cfr. TR4.10), sviluppando la componente on Edge del modello di Machine Learning riguardante la predizione del rischio clinico mediante Protocollo NEWS2 (cfr. Cap. 4.1 del TR4.10). Definiamo la tecnica di Machine Learning (ML) più adatta per la previsione delle classi di rischio cliniche di un paziente costantemente monitorato in una particolare condizione in cui è disponibile un numero limitato di parametri vitali. Questa analisi è dunque preparatoria per l'obiettivo finale di progettare un dispositivo periferico collegato a uno o più dispositivi medici indossabili tramite IoMT, che sfrutta in modo adattivo il miglior modello ML per prevedere una classe di rischio NEWS2 affidabile.

Il protocollo NEWS2, ampiamente definito nel Cap. 4.1 del TR4.10, si basa su un sistema di punteggio aggregato in cui un punteggio viene assegnato a misurazioni fisiologiche quando il paziente viene monitorato.

Nel protocollo NEWS2 un punteggio è associato a ciascun valore di parametro vitale rilevato, come mostrato nella Tabella I. La somma di questi valori determina un punteggio finale aggregato, che viene utilizzato per delineare una classe di rischio clinico (basso-medio-alto) a soglie diverse.

NEWS2 Vital Signs	3	2	1	0	1	2	3
Respiratory rate (bpm)	≤ 8		$9-11$	$12-20$		$21-24$	≥ 25
Oxygen saturation (SpO_2)	≤ 91	$92-93$	$94-95$	≥ 96			
Systolic blood pressure (mmHg)	≤ 90	$91-100$	$101-110$	$111-219$			≥ 220
Heart rate (bpm)	≤ 40		$41-50$	$51-90$	$91-110$	$111-130$	≥ 131
Temperature (°C)	≤ 35.0		$35.1-36.0$	$36.1-38.0$	$38.1-39.0$	≥ 39.1	
State of consciousness				A			VPU

Tabella 1 - Assegnazione punteggi per il calcolo della classe di rischio secondo l'algoritmo NEWS2

Il processo Edge Computing viene eseguito entro la finestra temporale, prefissata nel file di configurazione, in cui si raccolgono il maggior numero di parametri fisiologici. Per memorizzare tali dati, il RPi si avvale dell'utilizzo di un dizionario, ovvero una struttura dati che colleziona coppie della forma (chiave, valore). Il RPi riserva per ogni slot del

dizionario, la coppia composta dai parametri fisiologici e le relative misurazioni. Gli slot rimangono per tutta la durata del processo in continuo aggiornamento. Infatti, nel caso in cui vengano registrati più misurazioni dello stesso parametro vitale, il RPi aggiorna gli slot al valore più recente. Qualora vengano registrati tutti i parametri fisiologici prima del termine della finestra temporale, si procede con la fase di predizione del rischio clinico NEWS2 del paziente. In caso contrario, verrà eseguita la predizione sulla base dei parametri fisiologici registrati purché siano congrui a quelli predisposti. Le combinazioni di misurazione accettate dal sistema per le seguenti:

- 6 parametri vitali: frequenza respiratoria, saturazione di ossigeno, temperatura, pressione sistolica, frequenza cardiaca, stato di coscienza;
- 5 parametri vitali: frequenza respiratoria, saturazione di ossigeno, temperatura, pressione sistolica, frequenza cardiaca;
- 4 parametri vitali: frequenza respiratoria, saturazione di ossigeno, temperatura, frequenza cardiaca;
- 3 parametri vitali: frequenza respiratoria, temperatura, frequenza cardiaca.

La scelta di rimozione dei parametri vitali è motivata dalla necessità di poter dare comunque una predizione del rischio clinico nonostante l'assenza di alcuni parametri vitali che non sono stati rilevati o per mancanza degli opportuni dispositivi medicali di più difficile reperimento, come uno Sfigmomanometro, oppure per andare incontro ad esigenze temporali più ristrette di definizione di un early warning. Nel caso in cui non venga registrato il minor numero di misurazioni necessarie, il processo non riuscirà a predirne il rischio clinico.

```
pi@raspberrypi:~/edgeai $ sudo ./bluetooth_adv
Serial Port service registered
pi@raspberrypi:~/edgeai $ sudo python3 blu_simulator_server.py

Avvio Processo AI on Edge for Digital Future

Acquisizione dei dispositivi medicali

In attesa di 2 dispositivi medicali..
Dispositivo medicale Medical Device 2 connesso

In attesa di 1 dispositivo medicale..
Dispositivo medicale Medical Device 3 connesso

Registrazione Parametri vitali del Paziente..

Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'94'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'saturazione'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'150'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'pressione_sistolica'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'36'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'temperatura'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'76'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_cardiaca'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'0'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'stato_di_coscienza'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'15'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_respiratoria'
```

Figura 8 - Ricezione dei parametri vitali dai dispositivi medicali

2.3. Classificazione del Rischio Clinico mediante Apprendimento Automatico

Per ottenere dei modelli accurati, in grado di restituire una predizione affidabile, è stato adottato il tipico ciclo di vita di un'applicazione Data Science.

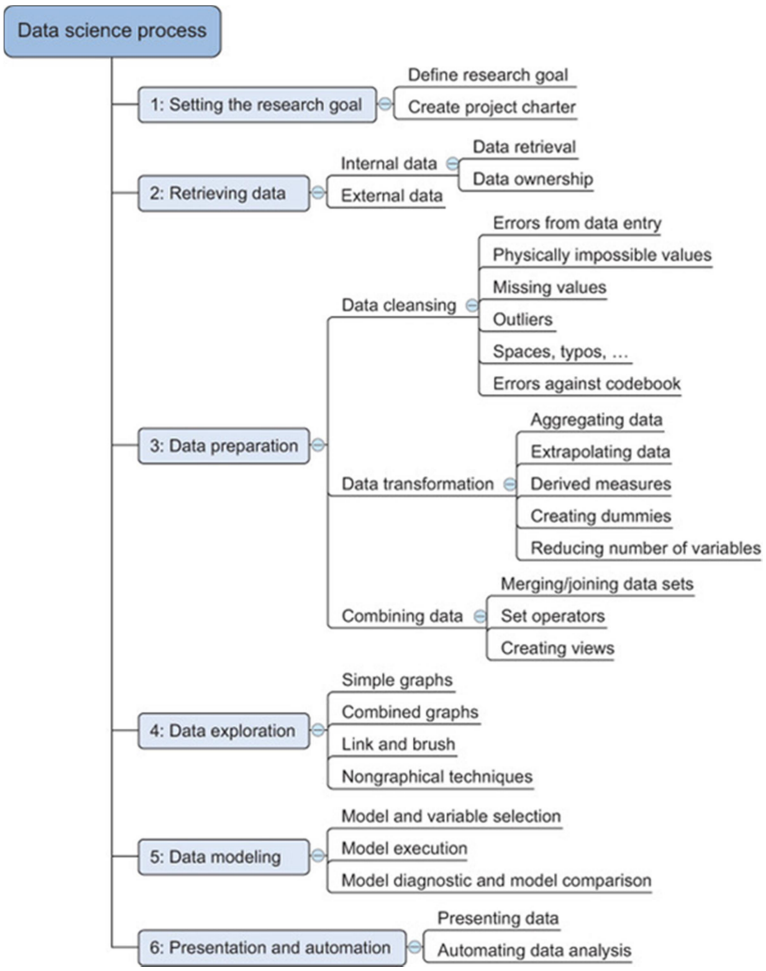


Figura 9 - Data Science lifecycle

Descrivere come nel prototipo dimostrativo sono state affrontate le fasi di Retrieving Data, Data Preparation e Data Exploration, ci permette di definire al meglio e nel dettaglio la fase di Data Modeling, cui obiettivo principale è la predizione del rischio clinico a fronte di un numero distinto p di parametri vitali, anche nel caso in cui non tutti e sei parametri vitali sono disponibili al momento della predizione.

Predire il rischio clinico di un paziente, a fronte dei parametri vitali misurati e relativo calcolo del NEWS2 Score, è un problema di classificazione risolvibile con tecniche di Machine Learning di tipo Supervisionato in cui è possibile predire dati futuri, a partire dal Dataset iniziale aventi osservazioni ed output già noti, che supervisionano la bontà del modello.

Il dataset iniziale contiene, per diversi pazienti, le osservazioni dei sei parametri vitali misurati in una sessione temporale di circa sei ore. Lungo la sessione temporale, le osservazioni per ogni parametro vitale sono state misurate in maniera indipendente ed identificate univocamente da un ID Sensore che verrà successivamente rimosso in quanto non contiene alcun contenuto informativo utile per la predizione dei dati.

id_assistito	id_sensore	cod_tipo_valore	valore	data
80	1	1	28	25/03/2019 08:30
80	4	2	10	25/03/2019 13:35
80	1	1	146	25/03/2019 14:10
80	8	4	38.08	25/03/2019 14:25
80	7	3	1	25/03/2019 08:35
80	1	0	113	25/03/2019 12:14
80	4	2	41	25/03/2019 10:13
80	8	4	38.49	25/03/2019 12:58
80	1	0	76	25/03/2019 13:26
80	1	1	68	25/03/2019 14:04
80	7	3	2	25/03/2019 10:30
80	8	4	34.95	25/03/2019 13:10
80	7	3	3	25/03/2019 13:41

Figura 10 – Stralcio del Dataset iniziale

Nella Data Preparation l'obiettivo principale è stata la ridefinizione del dataset principale in modo tale che per ciascun paziente venisse registrata la tupla contenente i sei parametri vitali, mediante l'utilizzo di tecniche di Time Series Analysis, e calcolare, una volta individuata la tupla completa, i valori NEWS2. Dalle analisi scaturite, l'intervallo di tempo minimo tra due consecutive misure è di 60 secondi, mentre l'intervallo massimo è 53 minuti. Il numero medio di misurazioni a cui un paziente è stato sottoposto è circa 74 in un intervallo tra 66 e 82 misure. Al fine di simulare un processo in tempo reale, le osservazioni sono state ridefinite secondo l'ordine cronologico di rilevazione lungo ogni sessione di misurazione. Di conseguenza, otteniamo per ciascuno paziente una matrice (n × p) record in cui n è il numero di misure simultanee e p è il numero di features. Successivamente, la numerosità delle osservazioni è stato reso omogeneo per ogni paziente presente nel Dataset X. Seguendo tal procedimento, sono stati individuati 201 pazienti distinti in cui per ciascuno di essi, sono state eseguite 43 osservazioni aventi tutte e 6 parametri vitali nell'arco temporale di circa 6 ore. Infine, una volta calcolato lo score NEWS2, è stata definita la relativa classe di appartenenza.

id_paziente	t	pressione_sistolica	frequenza_cardiaca	frequenza_respiratoria	stato_di_coscienza	temperatura	saturazione	news_score	news_class
80	0	117	24	26	0	35.4	103	7	2
80	1	90	122	17	0	35.43	104	6	1
80	2	86	137	6	1	33.15	104	15	2
80	3	125	21	19	2	33.59	89	12	2
80	4	131	82	42	2	44.97	91	11	2
80	5	127	134	5	2	38.05	103	9	2
80	6	84	25	44	1	35.98	97	13	2
80	7	77	83	8	2	38.84	101	10	2
80	8	85	81	34	2	34.02	81	15	2
80	9	87	69	13	3	33.48	101	9	2

Figura 11 – Stralcio del Dataset finale

Una volta generato il Dataset X, conforme strutturalmente a quanto richiesto da un modello predittivo supervisionato, nella fase di Data Exploration sono state create copie del Dataset X in cui per ciascuno di essi, sono stati rimossi uno o più parametri vitali.

Nei Dataset così generati, sono stati rimossi i seguenti set di parametri vitali:

- Dataset X_6 con 6 parametri vitali: presenti tutti i parametri vitali misurabili;
- Dataset X_5 con 5 parametri vitali: eliminato lo Stato di Coscienza;
- Dataset X_4 con 4 parametri vitali: eliminati lo Stato di Coscienza e la Saturazione;
- Dataset X_3 con 3 parametri vitali: eliminati Pressione Sistolica, Stato di Coscienza e Saturazione.

La fase di Data Modeling è un processo iterativo ed è stato applicato, senza alcuna distinzione, su tutti i Dataset X_n creati. Di seguito, verrà menzionato il solo Dataset X_6 e generalizzato come Dataset X. La fase di Data Modeling è suddivisa in 3 step:

1. Campionamento del Training Dataset;
2. Hyper-parameters Tuning;
3. Comparazione dei Modelli

La fase di campionamento del Training Dataset ha il compito di estrarre dal dataset X un numero di sotto dataset e conformi matematicamente per l'addestramento del modello. Infatti, le osservazioni essendo temporalmente dipendenti, non garantiscono una predizione affidabile portando il modello in Overfitting. È necessario che i dati siano linearmente indipendenti. È stato definito un metodo che generi sotto dataset che rispettano i seguenti criteri:

1. Ogni dataset deve contenere m osservazioni indipendenti, associati ad m differenti pazienti. Ciò evita il problema di autocorrelazione, caratteristica tipica delle serie temporali;
2. Le proporzionalità delle classi NEWS2 devono rimanere preservate per ciascun dataset, in questo modo si prendono in considerazione le informazioni contenute nella popolazione dati di riferimento;
3. L'unione dei Dataset deve raggiungere almeno il 75% del Dataset X originale.

Il processo di campionamento ha generato 61 sotto Dataset D = $\{D_1, D_2, \ldots, D_{61}\}$ e 17 Dataset K = $\{K_1, K_2, \ldots, K_{17}\}$ ottenuti dalla differenza tra il Dataset X e l'unione dei sotto dataset che compongono il Dataset D. I dataset D verranno utilizzati nella fase di Training del modello, mentre ogni dataset K è sottoposto ad Hyper-parameter Tuning. Nella fase di Hyper-parameter Tuning, si definiscono i migliori parametri per migliorare le performance del modello predittivo che, non sono ricavabili tramite dati e sono scelti dal Data Analyst. L'approccio utilizzato prevede la combinazione tra 5-Fold Cross Validation e Grid Search al fine di individuare i parametri che massimizzano la capacità predittiva. Per effetto del Grid Search, per ogni classificatore si individua un numero finito di possibili valori che ciascun parametro potrebbe assumere e per ogni dataset appartenente a K poi si effettua il training dell'algoritmo utilizzando tutte le possibili combinazioni dei valori al fine di individuare quelli che massimizzano le metriche di valutazione. Nella 5-Fold Cross Validation, il Dataset viene suddiviso in k = 5 sotto Dataset dove i k-1 vengono utilizzati come Training del modello e il singolo dataset come Validation del modello. Il procedimento viene ripetuto per k volte utilizzando k combinazioni differenti.

I modelli di Machine Learning, che risolvono il problema della Classificazione, utilizzati e processati per Hyper-parameter Tuning sono:

1. Support Vector Machines (SVMs);
2. Random Forest (RF);
3. MultiLayer Perceptron (MLP).

I tre algoritmi sono i più diffusi e risolvono il problema della classificazione con tre approcci differenti. Nello specifico, SVMs riesce a risolvere problemi di classificazione

non lineare proiettando dati non separabili linearmente in uno spazio a più dimensioni, trasformandoli i dati linearmente separabili tramite una Funzione Kernel quali Radial Basis Function (RBF) Kernel e Polynomial Kernel. RF rappresenta un esemble di alberi decisionali, garantendo un modello più robusto i cui risultati non sono influenzati da normalizzazione, outlier e, di conseguenza, da overfitting. Infine, MLP è una tipologia particolare di Artificial Neural Network che utilizza l'algoritmo supervisionato di BackPropagation per l'addestramento e, tecniche di computazione non lineari per gestire features che non sono linearmente separabili.

I parametri utilizzati per ogni modello di Machine Learning sono:

1. SVMs: Radial Basis Function (RBF) Kernel e Polynomial Kernel;
2. RF: Numero tra 2 fino a 50 alberi decisionali;
3. MLP: Numero tra 1 fino a 100 Strati Nascosti e le funzioni di attivazione Logistica, Tangente, Rettificatore (ReLu).

Per ogni dataset e per ogni classificatore, sono stati scelti i migliori classificatori tra i 17 dataset sottoposti nella fase di Hyper-parameter Tuning e denominati come SVM_{best}, RF_{best} and MLP_{best}.

Dataset	5-fold best	5-fold best	5-fold best
K_1	$SVM_b^{(1)}$	$RF_b^{(1)}$	$MLP_b^{(1)}$
K_2	$SVM_b^{(2)}$	$RF_b^{(2)}$	$MLP_b^{(2)}$
...	...	...	...
K_{17}	$SVM_b^{(17)}$	$RF_b^{(17)}$	$MLP_b^{(17)}$
Best Overall	SVM_{best}	RF_{best}	MLP_{best}

Tabella 2 – Selezione del miglior modello di ML

Dai modelli SVM_{best}, RF_{best} and MLP_{best} sono stati estratti i parametri migliori da utilizzare per ogni dataset D durante il Training dei modelli, classificando le classi di rischio NEWS2. Il Dataset D è stato suddiviso come 80% Training e 20% Testing. L'accuratezza media di tutti e tre i classificatori supera il 75%. Attraverso la comparazione tra i modelli si evince come il classificatore più accurato e il più idoneo per la classificazione del rischio clinico è il Random Forest.

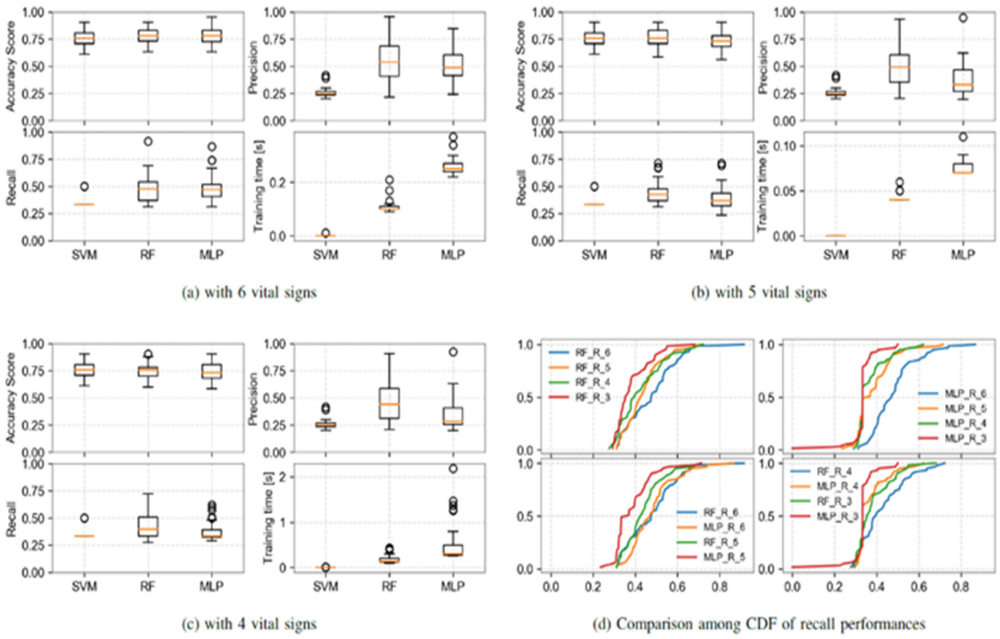


Figura 12 - Visualizzazione dei risultati delle metriche di performance dei modelli di ML

I modelli RF, per ogni numero distinto di parametri vitali misurabili, sono stati salvati in un apposito file Pickle e caricati nel RPi. Durante il processo Edge Computing nella fase di predizione del rischio clinico viene utilizzato il relativo modello in base ai parametri registrati.

```
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'94'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'saturazione'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'150'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'pressione_sistolica'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'36'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'temperatura'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'76'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_cardiaca'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'0'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'stato_di_coscienza'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'15'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_respiratoria'

La predizione del rischio clinico del Paziente avverrà sulla base delle seguenti misurazioni:
    pressione_sistolica  frequenza_cardiaca  frequenza_respiratoria  \
0                150.0               76.0                    15.0

   stato_di_coscienza  temperatura  saturazione
0                 0.0         36.0         94.0

Predizione rischio clinico del Paziente: basso
```

Figura 13 - Esempio di avvenuta predizione on Edge del Rischio Clinico

2.4. Explanation dei Risultati di ML mediante Rappresentazione della Conoscenza e Ragionamento Automatico tramite Ontologie

Un modello di ML supervisionato che risolve un task di classificazione è in grado di indicarci, a partire da un set di ipotesi, in quale classe di appartenenza vengono collocate le osservazioni del dataset. Tuttavia, la classe predetta non fornisce alcun ragionamento intrinseco e per questo è necessario che venga associata un'ontologia in grado di rappresentare la conoscenza e dedurre da essa, un ragionamento semantico. Un'ontologia può essere interpreta come una rete semantica in cui include l'insieme dei concetti (entità, attributi e processi), che ne definiscono il vocabolario di termini, gli assiomi e le relazioni tra concetti.

L'approccio utilizzato per la costruzione dell'ontologia medicale è di tipo Top-Down in cui è stato definito il livello Upper Domain Ontology, ovvero i concetti generali comuni per poi specificare i concetti che non possono essere decomposti nel livello Lower Domain Ontology. I sei parametri vitali rappresentano le entità di alto livello e per ciascuna di essi sono state definite più sotto entità, stabilendo una relazione di ereditarietà. La modellazione dell'ontologia medicale riproduce fedelmente il sistema di punteggio assegnabile a ciascun valore di ogni parametro fisiologico, grazie all'utilizzo dell'Annotation Property "Valore Numerico" aventi le sottoproprietà "ValoreMin" e "ValoreMax" che hanno permesso la definizione delle sottoclassi per ciascuno dei parametri vitali. Per ogni Annotation Property è indicata una etichetta alle sottoclassi, affibbiate secondo il ragionamento derivante dal sistema di score del NEWS2:

- +3: "Estremamente Basso (Criticità)";
- +2: "Molto Basso";
- +1: "Basso";
- +0: "Normale";
- +1: "Alto";
- +2: "Molto Alto";
- +3: "Estremamente Alto (Criticità)".

Per lo stato di coscienza diviso è stato imposto la label “Vigile” con sottoproprietà 0, il quale indica la lettera ‘A’, e “Stato non Vigile” con sottoproprietà 3, il quale indica le lettere V (Reazione Vocale), P (Reazione al Dolore), U (Coma).

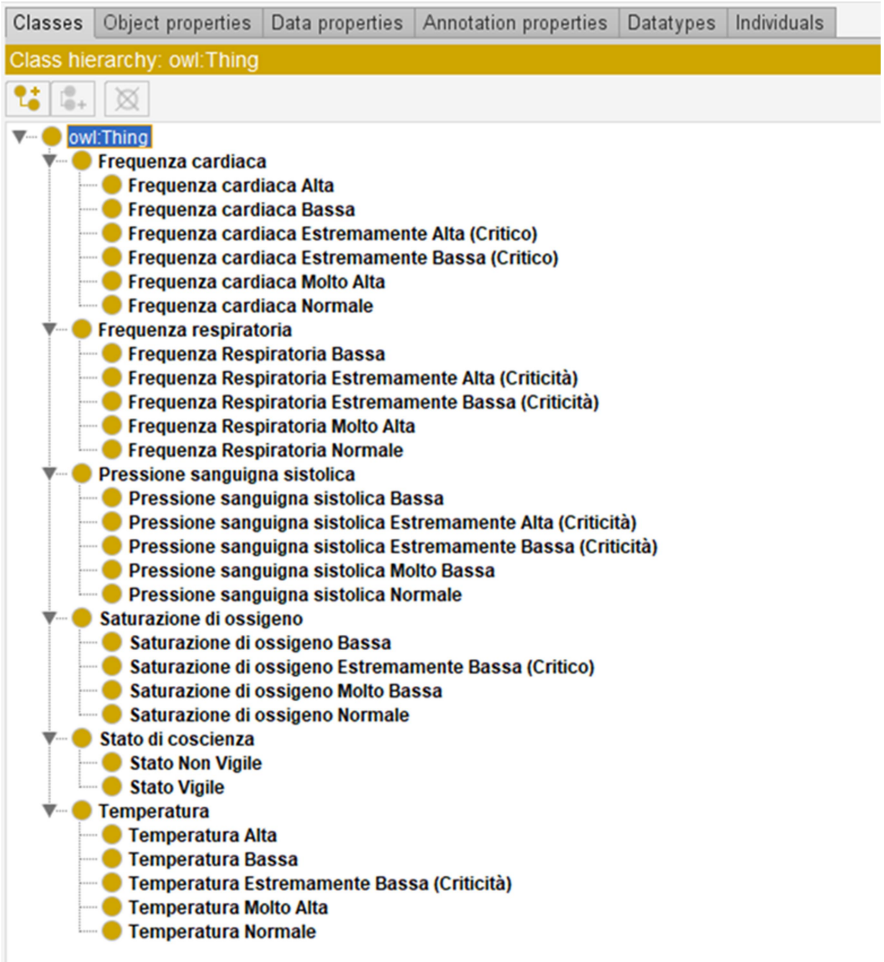


Figura 14 - Rappresentazione tassonomica dell'Ontologia di riferimento per lo Scenario d'Uso

Nel processo di Edge Computing, una volta eseguita la predizione del rischio clinico del paziente, la fase di Explanation della predizione consente nell’ estrarre, attraverso una procedura di query dei dati, le sottoclassi di un'entità per ogni parametro vitale registrato e nel ricercare la sottoclasse di appartenenza confrontando la misurazione corrente del parametro vitale con gli intervalli numerici definite dalle sottoproprietà dell’Annotation Property. Per ogni sottoclasse di appartenenza individuata, si procede con l’estrazioni delle label che ne identificano il ragionamento semantico.

Al termine della computazione, la fase di explanation fornisce un numero di deduzioni pari al numero di misurazioni esaminate, ciascuna dei quali espone il corretto grado di gravità. Combinando il potere informativo estratto dalle fasi di Apprendimento e Ragionamento Semantico, non solo si ottiene la predizione del rischio clinico ma anche la motivazione che spinge l'algoritmo ad aver indicato tale decisione.

```
Registrazione Parametri vitali del Paziente..

Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'94'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'saturazione'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'150'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'pressione_sistolica'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'36'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'temperatura'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'76'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_cardiaca'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'0'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'stato_di_coscienza'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'15'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_respiratoria'

La predizione del rischio clinico del Paziente avverrà sulla base delle seguenti misurazioni:
    pressione_sistolica  frequenza_cardiaca  frequenza_respiratoria  \
0                150.0               76.0                    15.0

   stato_di_coscienza  temperatura  saturazione
0                 0.0         36.0         94.0

Predizione rischio clinico del Paziente: basso

Explanation della predizione tramite Ontologia

Il Paziente presenta rischio clinico basso perché presenta le seguenti misurazioni:
Saturazione di ossigeno Bassa
Pressione sanguigna sistolica Normale
Temperatura Bassa
Frequenza cardiaca Normale
Stato Vigile
Frequenza Respiratoria Normale
Timer: 0:03:55
```

Figura 15 – Esempio di avvenuta Explanation dei risultati di ML tramite Ontologia

2.5. Comunicazione con il Middleware della Piattaforma DIGITAL FUTURE

Per completare il percorso architetturale definito, è necessario che le informazioni vengano gestite da remoto ed inviate nella piattaforma Digital Future sfruttando il Cloud Computing.

È stato definito uno script di configurazione che permette il collegamento tra RPi e piattaforma Digital Future. In questa fase bisogna comunicare alla piattaforma lo schema dei messaggi che saranno inviati dal sensore sul topic Kafka.

Nel processo Edge Computing lo schema dei messaggi sarà rappresentato da una tupla formata da:

- ID Paziente;
- Le misurazioni dei parametri vitali;
- Timestamp di visita del Paziente.

In prima istanza, si indica alla piattaforma lo schema messaggi inviato dal sensore sul topic Kafka. Lo schema messaggi è definito dai seguenti campi:

- device_id - l'id del dispositivo che sta instaurando la comunicazione;
- event_time - timestamp del momento in cui viene inviato il messaggio di configurazione;
- real_time_data - configurazione del sensore in Real Time contenenti le informazioni:
    - payload - informazioni sul tipo e versione del sensore e configurazione delle tabelle in Cassandra:
        - Version - versione del sensore utilizzata nella creazione dei topic MQTT/Kafka;
        - device_type - tipo del dispositivo utilizzato nella creazione dei topic mqtt/kafka;
        - table_info - informazioni della tabella cassandra quali:
            - name - la cui nomenclatura deve essere uguale al device type;
            - table_format - struttura della tabella cassandra:
                - cols - array di JSON contenenti nome e tipo delle colonne di Cassandra;
            - options - opzioni della tabella;
            - primay_keys - colonne che costituiscono la chiave primaria;
            - indexes - eventuali indici;
- frequency - frequenza con cui verranno inviati i dati.

Si tratta di una fase di configurazione ed è un passaggio indipendente dal processo Edge Computing.

Nel processo Edge Computing, invece, il record dati viene inviato su Digital Future in formato JSON sul topic Plug And Play il quale estrae il contenuto informativo del file e lo istanzia sottoforma di record sulla tabella Cassandra.

Il nome della tabella Cassandra è "*raspberry_news2*", mentre la sua struttura tabellare è così composta:

Nome Campo	Elemento Dati	Chiave Primaria
timestampo	timestamp	x
id_paziente	bigint	x
pressione_sistolica	double	
frequenza_cardiaca	double	
frequenza_respiratoria	double	
stato_di_coscienza	bigint	
temperatura	double	
saturazione	double	
rischio_clinico	varchar	

Ai parametri vitali non registrati viene assegnato come valore d'ufficio 0, mentre per lo stato di coscienza, valore –1.

timestamp_	id_paziente	frequenza_cardiaca	frequenza_respiratoria	pressione_sistolica	rischio_clinico	saturazione	stato_di_coscienza	temperatura
2020-04-07 15:25:56	80	76	15	150	basso	94	0	37,6
2020-04-07 15:21:52	80	76	15	150	basso	94	0	37,6
2020-04-09 18:12:46	80	76	15	150	basso	94	0	37,6
2020-04-07 15:34:02	80	76	15	0	medio	0	-1	37,6
2020-04-07 15:12:31	84	80	15	120	medio	94	3	34,7
2020-04-07 15:17:39	80	76	15	150	basso	94	0	37,6
2020-04-10 07:12:36	89	76	15	150	basso	94	0	37,6

Figura 16 - Esempio Tabella Dati registati su DF in Cassandra

Nel processo Edge Computing, dopo la predizione del rischio clinico e relativa explanation, i dati sono trasmessi sulla piattaforma Digital Future e, nello stesso istante, il RPi ci indica lo stato di invio.

```
Il Paziente presenta rischio clinico basso perché presenta le seguenti misurazioni:
Temperatura Normale
Saturazione di ossigeno Bassa
Stato Vigile
Frequenza cardiaca Normale
Frequenza Respiratoria Normale
Pressione sanguigna sistolica Normale
Invio dati alla piattaforma Digital Future
Dati inviati alla piattaforma Digital Future con successo
Timer: 0:03:55
```

Figura 17 - Notifica su RPi di invio dati a piattaforma

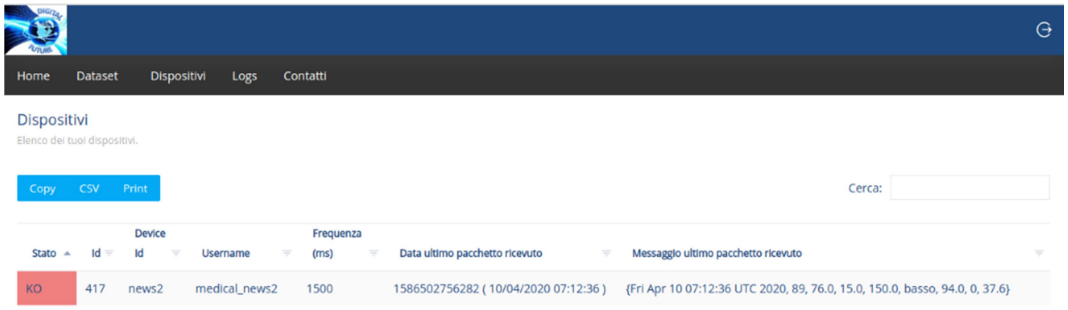


Figura 18 - Notifica di ricezione dati in piattaforma

Il protocollo ISO di comunicazione adottato è MQTT. Riepilogando, l'intero processo Edge Computing viene mostrato nella seguente figura:

```
Avvio Processo AI on Edge for Digital Future

Inserire ID Paziente: 89

Acquisizione dei dispositivi medicali

In attesa di 2 dispositivi medicali..
Dispositivo medicale Medical Device 3 connesso

In attesa di 1 dispositivo medicale..
Dispositivo medicale Medical Device 2 connesso

Registrazione Parametri vitali del Paziente..

Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'37.6'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'temperatura'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'94'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'saturazione'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'0'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'stato_di_coscienza'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'76'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_cardiaca'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'15'
Misurazione ricevuta dal dispositivo medicale Medical Device 2: b'frequenza_respiratoria'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'150'
Misurazione ricevuta dal dispositivo medicale Medical Device 3: b'pressione_sistolica'

La predizione del rischio clinico del Paziente avverrà sulla base delle seguenti misurazioni:
    pressione_sistolica  frequenza_cardiaca  frequenza_respiratoria  \
0                150.0               76.0                    15.0

   stato_di_coscienza  temperatura  saturazione
0                 0.0         37.6         94.0

Predizione rischio clinico del Paziente: basso

Explanation della predizione tramite Ontologia

Il Paziente presenta rischio clinico basso perché presenta le seguenti misurazioni:
Temperatura Normale
Saturazione di ossigeno Bassa
Stato Vigile
Frequenza cardiaca Normale
Frequenza Respiratoria Normale
Pressione sanguigna sistolica Normale
Invio dati alla piattaforma Digital Future
Dati inviati alla piattaforma Digital Future con successo
Timer: 0:03:55
```

Figura 19 - Intero processo AI on Edge su RPi

3. Classificazione dei Referti Clinici

3.1. Annotazione semantica dei referti medici

Nel seguito sono descritte le componenti adottate per l'annotazione semantica dei referti medici in italiano e in inglese e la loro classificazione secondo le due ontologie di riferimento già citate nel paragrafo 4.2 del TR4.10 "Analisi dei Servizi di Piattaforma per l'Healthcare", MeSH per i referti medici in italiano e DOID per i referti medici in inglese.

3.1.1. *Approccio integrato per il trattamento dei referti medici in lingua italiana*

Come espresso nel paragrafo introduttivo, per la presente sperimentazione relativa alla classificazione dei referti in lingua italiana, è stato scelto di adottare un approccio combinato tra tecniche di Information Extraction e approccio ontologico.

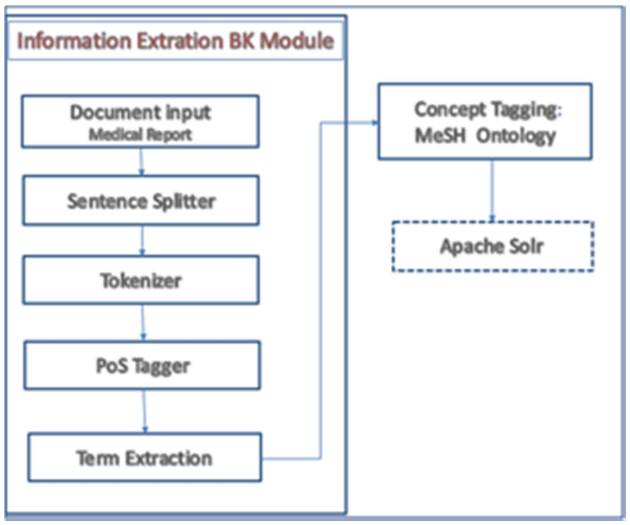
Per quanto concerne nello specifico le tecniche di Information Extraction, è stato scelto di utilizzare le funzionalità già presenti in BigKnowlegde, prodotto proprietario Exprivia dedicato all'acquisizione, organizzazione, condivisione e reperimento della conoscenza sottesa sia alle fonti strutturate (quali database, file CSV, etc..) che alle fonti non strutturate (documenti di testo, email, etc.) che saranno nello specifico oggetto della presente analisi. Le funzionalità rese disponibili da BigKnowledge sono basate su un modello della conoscenza realizzato mediante l'applicazione di una combinazione unica di tecniche di analisi morfosintattica, clustering e analisi semantica latente (LSA) , garantendo una copertura molto elevata dei fenomeni, rispetto a qualsiasi sistema basato sulla definizione manuale di reti semantiche, e non richiede sforzo umano per la definizione di regole per chiarire i sensi delle parole (bassi costi di manutenzione della knowledge base: completezza, correttezza e coerenza del modello non sono influenzate da regole "errate").

Tra le principali funzionalità offerte da BigKnowlegde ricordiamo in particolare:

- **Clustering**: servizio dedicato al reperimento autonomo su rete Internet della documentazione potenzialmente utilizzabile per l'addestramento, al sezionamento della stessa in funzione dell'uniformità dei contenuti, e al raggruppamento delle sezioni così ottenute in cluster semanticamente coesi. In altri termini BigKnowledge è in grado di automatizzare al massimo la fase di addestramento preventivo, lasciando all'esperto di dominio il solo compito di scartare i cluster non di proprio interesse operando direttamente la fase di rafforzamento dell'addestramento.
- **Classificazione**: servizio dedicato alla classificazione di informazioni strutturate e non strutturate secondo un insieme di categorie preesistenti (tassonomia concettuale). La tassonomia concettuale può essere definita dall'utente identificando una serie di argomenti di interesse fornendo descrizioni di BigKnowledge che vengono utilizzate per valutare quali cluster sono maggiormente correlati ad essi. I documenti vengono quindi riorganizzati in base a questa relazione fornendo all'utente un punteggio di categorizzazione.
- **Natural Language Processing**: funzionalità che fornisce funzionalità di recupero delle informazioni sulla base delle query degli utenti espresse in linguaggio naturale, indipendentemente dalla lingua espressa in esse; a seguito di una query, BigKnowledge fornisce solo i documenti relativi ai concetti espressi dall'utente. Inoltre, identifica e classifica automaticamente le sezioni specifiche dei documenti recuperati corrispondenti alla query dell'utente.
- **Information Extraction** e **Named Entity Recognition**: funzionalità dedicata al rilevamento di entità rilevanti (annotazioni) all'interno di testi non strutturati in relazione ad uno specifico dominio al fine di migliorarne la classificazione ed il reperimento di documenti di interesse a fronte di una specifica query espressa dall'utente;

Nella sperimentazione in oggetto è stato scelto in particolare di ampliare le funzionalità di Information Extraction già presenti internamente a BigKnowledge (BK) per l'annotazione di documenti testuali, al fine di identificare nello specifico le informazioni testuali riferibili alle voci dell'ontologia MeSH.

Nella seguente figura viene riportato il flusso elaborativo utilizzato.



Le componenti riportate nella figura precedente sono state già descritte nel più generale paragrafo "Modulo di pre-processamento di documenti" relativo alle tecniche di Information Extraction. Una volta terminata la fase di annotazione del testo, finalizzata a rilevare all'interno del testo i termini riferibili nello specifico alle classi dell'ontologia MeSH, i risultati dell'annotazione vengono inviati a Apache Solr per la fase di indicizzazione.

Apache Solr, è una piattaforma di ricerca open-source basata su Apache Lucene che offre funzionalità come la ricerca full-text, la ricerca per facet, l'indicizzazione in near-real time, il clustering dinamico, l'integrazione di un database, la gestione di documenti in vari formati quali word o PDF, e ricerca geospaziale. Solr, è altamente affidabile, scalabile e fault tolerant in quanto fornisce funzionalità di indicizzazione distribuita, replicazione e interrogazioni load-balanced, failover e ripristino automatizzati e configurazione automatizzata.

Come espresso nei paragrafi precedenti la componente di Information Extraction qui descritta è stata integrata dentro una piattaforma open source denominata OpensemanticSearch (paragrafo 4) opportunamente adattata alle finalità della sperimentazione qui descritta.

In questo caso specifico, come vedremo in dettaglio nel paragrafo "OpenSemanticSearch", la componente di BigKnowledge deputata alle funzionalità di annotazione del testo sarà richiamata come **plugin** esterno nella fase di **enrichment** dei documenti già indicizzati dentro a Solr al fine di aggiungere, come in questo caso, annotazioni ai documenti e agganciarli a risorse esterne quali le ontologie di dominio.

Entrando nello specifico della sperimentazione in oggetto, il flusso definito nella figura precedente è stato applicato a referti in italiano del tipo rappresentato nella figura successiva. Il risultato dell'infomation Extractiion applicata al referto ha dato come risultato quello riportato in formato JSON.

RM cerebrale

Ventricoli cerebrali in sede, con cavità laterali più ampie.

Quarto ventricolo in sede, di normale ampiezza. PiU' ampia la scissura interemisferica. Non anomalie del tentorio cerebellare. Normali anche le strutture cerebellari, in sede, come pure il tronco cerebrale. Loro lieve anomalia morfologica della lamina quadrigemina, con aspetto a becco e e si trova usualmente nelle anomalie associate a Chiari 2. lo studio dell'asse spinale mostra canale rachideo e midollo regolari sino al tratto dorsale inferiore. Il midollo appare ancorato dalla spina bifida in L4 - L5, dove il canale rachideo risulta decisamente più ampio, con modesto meningocele.

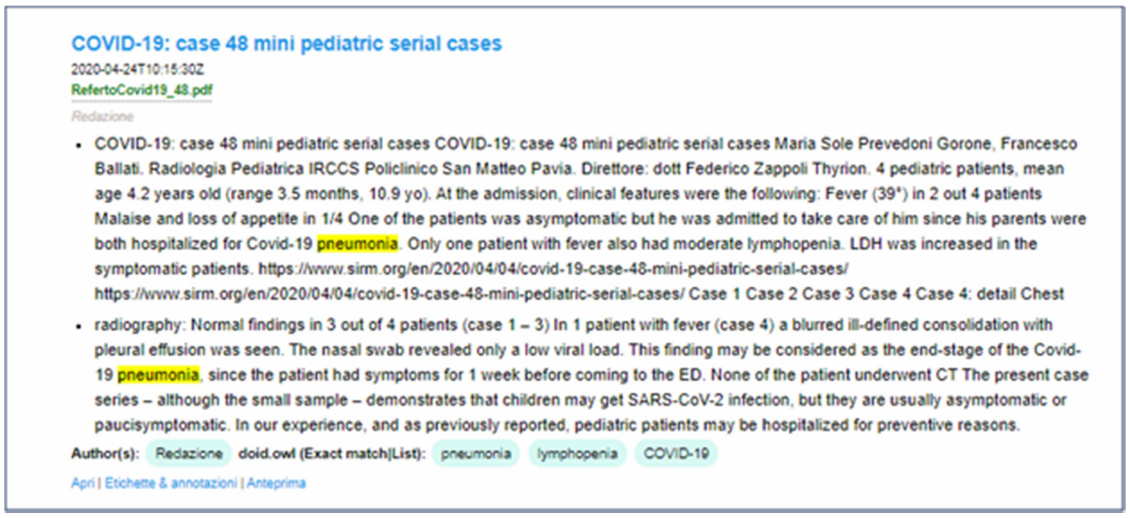
{"msg":{"annotations":[{"surf":"Ventricoli cerebrali","endOffset":34,"startOffset":14,"pos":"MESH","keys":"A08.186.211.276","areas":"A"},{"surf":"cavità laterali","endOffset":63,"startOffset":48,"pos":"TERM"},{"surf":"Quarto ventricolo","endOffset":93,"startOffset":76,"pos":"MESH","keys":"A08.186.211.276.500","areas":"A"},{"surf":"normale ampiezza","endOffset":122,"startOffset":106,"pos":"TERM"},{"surf":"scissura interemisferica","endOffset":162,"startOffset":138,"pos":"TERM"},{"surf":"tentorio cerebellare","endOffset":203,"startOffset":183,"pos":"TERM"},{"surf":"strutture cerebellari","endOffset":243,"startOffset":222,"pos":"TERM"},{"surf":"tronco cerebrale","endOffset":283,"startOffset":267,"pos":"MESH","keys":"A08.186.211.132","areas":"A"},{"surf":"lieve anomalia","endOffset":304,"startOffset":290,"pos":"TERM"},{"surf":"anomalie associate","endOffset":412,"startOffset":394,"pos":"TERM"},{"surf":"asse spinale","endOffset":452,"startOffset":440,"pos":"TERM"},{"surf":"canale rachideo","endOffset":475,"startOffset":460,"pos":"TERM"},{"surf":"midollo regolari","endOffset":494,"startOffset":478,"pos":"TERM"},{"surf":"spina bifida","endOffset":574,"startOffset":562,"pos":"TERM"},{"surf":"canale rachideo","endOffset":610,"startOffset":595,"pos":"TERM"},{"surf":"modesto meningocele","endOffset":665,"startOffset":646,"pos":"TERM"}]},"docKey":"1","error":false}

3.1.2. *Approccio per l'annotazione dei referti medici in inglese*

I referti medici in inglese, come espresso precedentemente, sono stati reperiti dal sito dalla SIRM (Società italiana di Radiologia Medica e Intenventistica) e sono relativi a pazienti su cui è stato riscontrato il COVID19. I suddetti documenti sono stati annotati e classificati utilizzando l'ontologia Doid già citata precedentemente utilizzando il componente **Solr Ontology Tagger** già interno, come vedremo più in dettaglio nel paragrafo "OpenSemanticSearch", all'applicazione OpensemanticSearch.

Tale scelta è stata per lo più dettata dall'evidenza, durante la sperimentazione, dei buoni risultati di classificazione ottenuti da questo componente per i referti in inglese rispetto all'italiano anche senza l'ausilio di plugin esterni quali BigKnowledge.

Nella figura che segue viene mostrato un esempio di referto medico in inglese con evidenza delle classi dell'ontologia Doid rilevate dal componente Solr Ontology Tagger.


COVID-19: case 48 mini pediatric serial cases
2020-04-24T10:15:30Z
RefertoCovid19_48.pdf
Redazione

- COVID-19: case 48 mini pediatric serial cases COVID-19: case 48 mini pediatric serial cases Maria Sole Prevedoni Gorone, Francesco Ballati. Radiologia Pediatrica IRCCS Policlinico San Matteo Pavia. Direttore: dott Federico Zappoli Thyrion. 4 pediatric patients, mean age 4.2 years old (range 3.5 months, 10.9 yo). At the admission, clinical features were the following: Fever (39°) in 2 out 4 patients Malaise and loss of appetite in 1/4 One of the patients was asymptomatic but he was admitted to take care of him since his parents were both hospitalized for Covid-19 pneumonia. Only one patient with fever also had moderate lymphopenia. LDH was increased in the symptomatic patients. https://www.sirm.org/en/2020/04/04/covid-19-case-48-mini-pediatric-serial-cases/ https://www.sirm.org/en/2020/04/04/covid-19-case-48-mini-pediatric-serial-cases/ Case 1 Case 2 Case 3 Case 4 Case 4: detail Chest
- radiography: Normal findings in 3 out of 4 patients (case 1 – 3) In 1 patient with fever (case 4) a blurred ill-defined consolidation with pleural effusion was seen. The nasal swab revealed only a low viral load. This finding may be considered as the end-stage of the Covid-19 pneumonia, since the patient had symptoms for 1 week before coming to the ED. None of the patient underwent CT The present case series – although the small sample – demonstrates that children may get SARS-CoV-2 infection, but they are usually asymptomatic or paucisymptomatic. In our experience, and as previously reported, pediatric patients may be hospitalized for preventive reasons.

Author(s): Redazione doid.owl (Exact match|List): pneumonia lymphopenia COVID-19
Apri | Etichette & annotazioni | Anteprima

3.2. OpenSemanticSearch

Come espresso più volte nei precedenti paragrafi l'intero progetto qui descritto è stato sviluppato all'interno di una piattaforma integrata open-source OpensemanticSearch[1], che fornisce funzionalità di esplorazione e ricerca su grosse collezioni di dati, Text Analytics, Media Monitoring, Document Analysis e Text Mining. Basata sulle piattaforme di ricerca open-source Apache Solr o Elastic Search e sugli Open Standard per Linked Data, Semantic Web e Linked Open Data Integration.

Più in dettaglio le principali funzionalità di OpenSemanticSearch possono essere così riassunte:

[1] https://opensemanticsearch.org/

- **Search Engine** (**Ricerca Full Text**) in molteplici datasource e su molteplici formati di file. La piattaforma offre infatti funzionalità di ricerca mediante l'immissione di una query (che può includere potenti operatori di ricerca) e di navigazione dei risultati.
- **Semantic Search** (attraverso l'integrazione di Thesauri e Ontologie): grazie all'integrazione dei vocabolari il motore di ricerca semantica multilingue è in grado di reperire anche sinonimi, iponimi e alias. Usando l'euristica per le regole grammaticali come la determinazione, è in grado di reperire anche altre forme di parole.
- **Filtri interattivi** (**Faceted Search**): è possibile navigare attraverso i risultati di ricerca con filtri interattivi (ricerca per facet) in grado di fornire risultati aggregati in base alla selezione di specifiche entità estratte dai testi come autori, organizzazioni, persone, luoghi, date, prodotti, tag o tipologie di documenti.
- **Exploration, browsing e preview** (**Exploratory search**): è possibile esplorare i dati o i risultati di ricerca visualizzando i documenti aggregati in base ai valori delle facet definite da specifiche named entity (ad es. path, tag, persone, posizioni, organizzazioni o prodotti), visualizzare le anteprime dei documenti reperiti (ovvero PDF, testo estratto, righe di tabella o immagini), analizzare o rivedere i set di documenti o gli elenchi di parole, estratte mediante le funzionalità di text mining, che afferiscono ai documenti prodotti dalle funzionalità di text mining.
- **Collaborative annotation e tagging** (**Social search e collaborative filtering**): è possibile taggare manualmente i documenti con specifiche keywords, categorie, nomi o note testuali che non sono incluse nel contenuto originale al fine di facilitarne il reperimento (document management e knowledge management) in altri contesti di ricerca o semplicemente poter filtrare i documenti annotati o taggati mediante l'utilizzo di filtri con le apposite funzionalità di ricerca per facet).
- **Datavisualization** (**Dataviz**): è possibile visualizzare utilizzare le date associate documenti per visualizzare dei trend charts o visionare i risultati delle analisi tesuali, mediante la visualizzazione di words cloud, visualizzare le relazioni tra le

entità estratte dai testi mediante l'accesso alle funzionalità di visual graph o visualizzazione i risultati contenti informazioni di tipo geografico mediante mappe interattive.

- **Supporto per molteplici formati di file**: file di testo, file Word e altri documenti di Microsoft Office o OpenOffice, Excel, LibreOffice Calc tables, PDF, E-Mail, CSV, doc, images, photos, pictures, JPG, TIFF, videos e molti altri formati.
- **Supporto per molteplici origini dati**: attraverso le funzionalità di ETL il sistema è in grado di connettersi a molteplici origini dati, come file e cartelle, database, siti web, feed RSS, e raccogliere tutti i dati insieme. I connettori del framework ETL (Extract Transform Load) per l'integrazione dei dati collegano e combinano più fonti di dati rendendoli fruibili al framework integrato di analisi e arricchimento dei dati, migliorando così i risultati delle ricerche.
- **Automatic text recognition**: attraverso l'integrazione dell'OCR (Optical Character recognition), il sistema è in grado di riconoscere il testo contenuto nelle immagini o in formati grafici quali documenti scannerizzati, screenshots, o foto in formato immagini o incluse all'interno di documenti PDF;

Per un maggior dettaglio su quelle appena espresse o su altre funzionalità incluse nella piattaforma si rimanda al sito https://www.opensemanticsearch.org/.

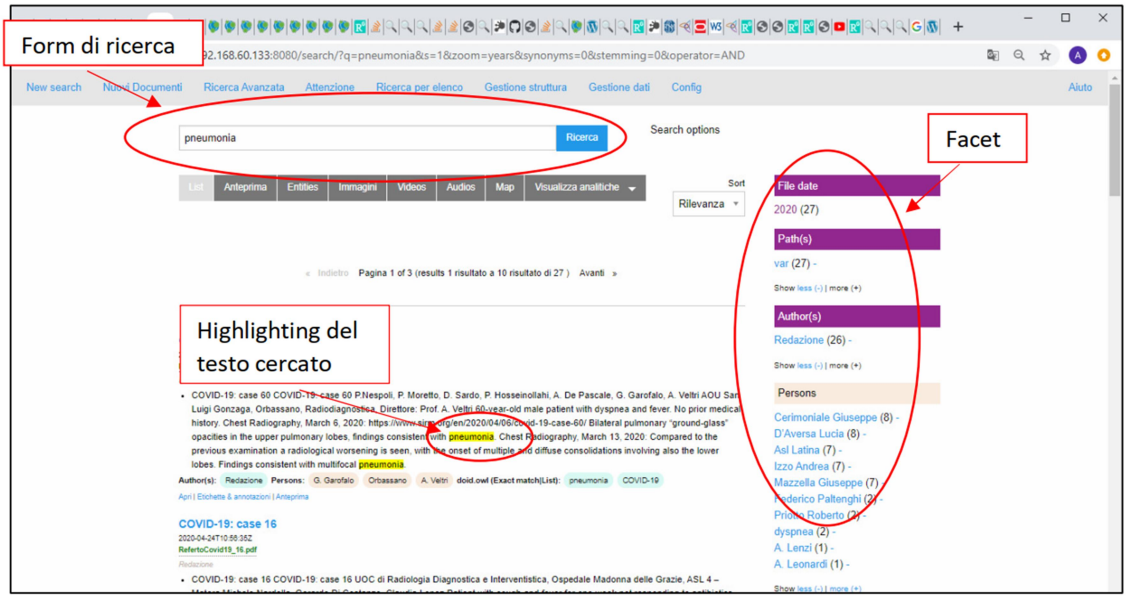
3.2.1. *Componenti principali e architettura*

Nel seguito vogliamo invece soffermarci sulle principali componenti che caratterizzano l'architettura software della piattaforma e che rendono possibile la fruizione delle funzionalità appena descritte, soffermandoci in particolare sulle componenti che sono state maggiormente utilizzate nel corso della presente sperimentazione e che hanno determinato il flusso operativo per alimentare il sistema e renderne fruibili le informazioni agli utenti finali:

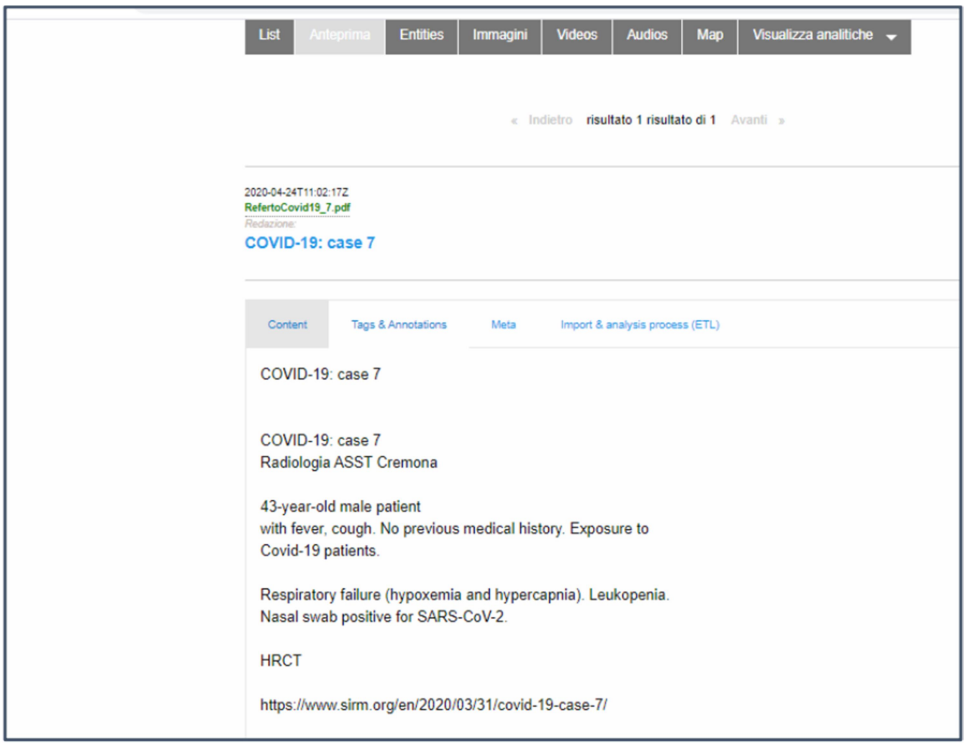
- La componente di **User Interface** (figura seguente), denominata Solr-php-ui, è definita dalle seguenti principali componenti:
    - **Search query form**, componente attraverso cui l'utente può inserire query di ricerca; come è possibile notare dalla figura sottostante

l'immissione della query di ricerca (che può essere anche una query in linguaggio naturale e non una semplice keyword), comporta l'highlighting della/e parola/e cercate all'interno dei documenti reperiti.

- **Explorer e navigator**: per le ricerche full text e la navigazione (exploratory search) dell'indice e dei risultati di ricerca mediante filtri interattivi (ricerca per facet);

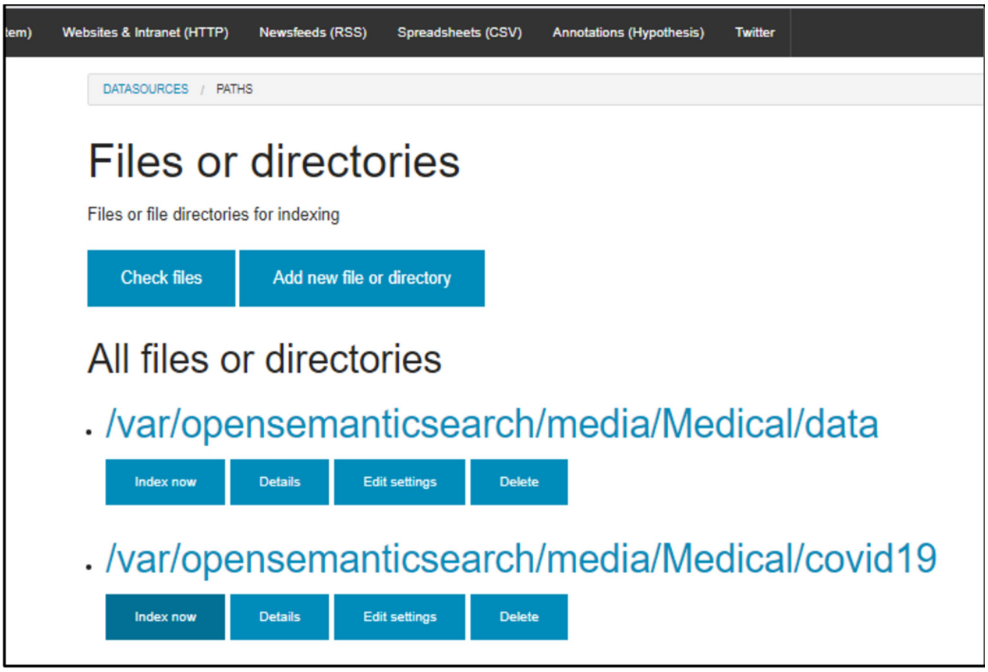


- **Viewers**: componente dedicato all'analisi dei dati (come words cloud o trend charts) e anteprime (vedere figura) per specifici formati (documenti, immagini, e-mail, etc.).



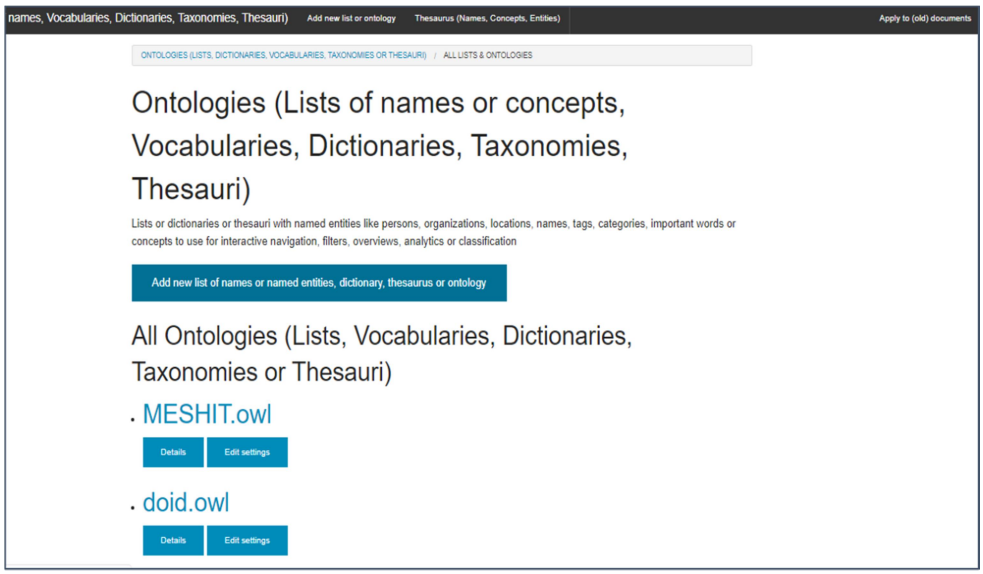
- **Annotators**: web application deputata all'annotazione dei documenti o CMS con forme e campi per la gestione di metadati come tag o annotazioni.
- **Search apps**: applicazioni e interfacce utente per effettuare ricerche in linguaggio naturale o mediante l'utilizzo di named entity estratte dai documenti di testo che saranno disponibili come filtri interattivi (ricerca per facet).
- **Add new annotation**: è possibile attraverso l'utilizzo di un'apposita funzionalità aggiungere manualmente delle annotazioni ai documenti al fine di arricchirne il contenuto e renderli più facilmente reperibile attraverso le apposite funzionalità di ricerca.

- **Datasources**: componente per la selezione delle sorgenti dati per l'indicizzazione dei contenuti. Attraverso questo componente è possibile non solo selezionare le sorgenti dati (ad esempio directory contenenti i documenti di interesse) ma anche indicizzare la sorgente di interesse selezionando l'apposito pulsante Index now.



- **Ontologies e thesauri**: questa componente consente l'aggiunta e la gestione di ontologie o di vocabolari esterni alla piattaforma. Questa

componente, come è possibile notare dalla figura 16, è stata utilizzata per l'aggiunta nel sistema delle 2 ontologie MeSH (per i referti in italiano) e Doid (per i referti in inglese).



- **Index e search server** (**Solr or Elastic Search**): server di ricerca che gestisce l'indice (indicizzatore) ed esegue query di ricerca (gestore query).
- **Open Semantic ETL**: framework per data integration, data analysis, data enrichment e ETL (extract, transform, load):
    - Connectors, importers, ingestors e crawlers: Importa dati da un'origine dati (ad es. File system, directory di file, condivisione file, sito Web o feed di notizie).
    - Parsers: viene utilizzato Apache Tika[2] per l'estrazione del testo e dei metadati da diversi formati di file e formati di documenti (PPT, XLS e PDF): Tutti questi tipi di file possono essere analizzati attraverso un'unica interfaccia, rendendo Tika particolarmente adatto per

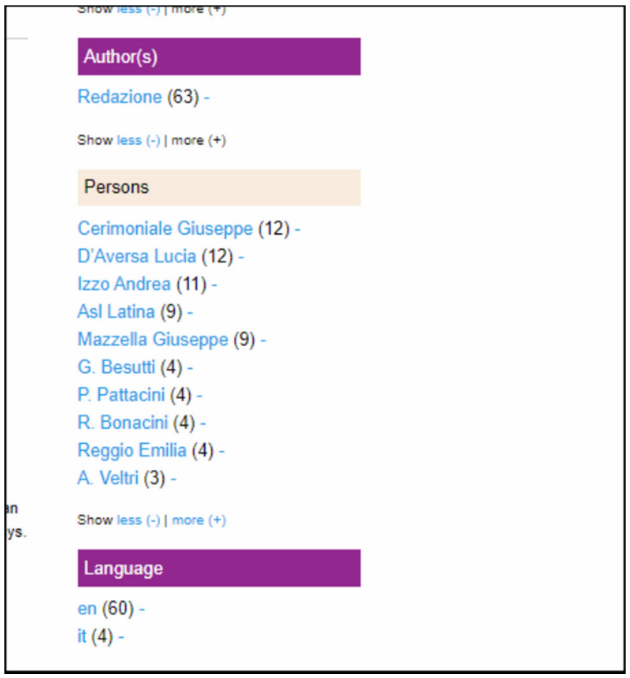
[2] https://tika.apache.org/

l'indicizzazione dei motori di ricerca, l'analisi dei contenuti, la traduzione e molto altro.

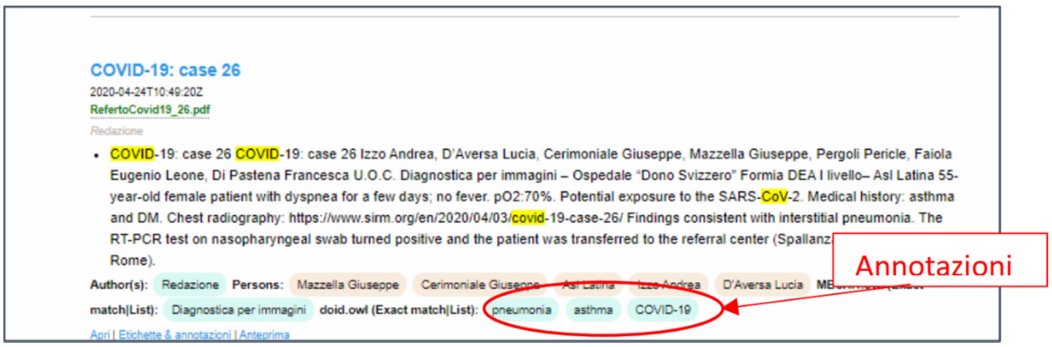
- Named entity recognition: questa componente è deputata all'estrazione delle named entity come persone, organizzazioni o luoghi mediante l'utilizzo di algoritmi di machine learning deputati al named entity recognition. Tali entità vengono poi utilizzate per la navigazione dei risultati mediante filtri (ricerca per facet) e per effettuare analisi specifiche come ad esempio determinare quali persone, organizzazioni o luoghi occorrono insieme e in quanti documenti. Per la realizzazione di questo componente sono state utilizzate le funzionalità di NER di **SpaCy**[3] e/o lo **Stanford Named Entities Recognizer** (Stanford NER)[4]. Nello specifico Spacy è una libreria open source per l'elaborazione del linguaggio naturale, scritta in Python. La libreria è rilasciata sotto licenza MIT ed attualmente implementa modelli statistici di reti neurali in inglese, tedesco, spagnolo, portoghese, francese, italiano, olandese e greco; inoltre offre funzionalità di NER e di tokenizzazione per diverse altre lingue.

[3] https://spacy.io/

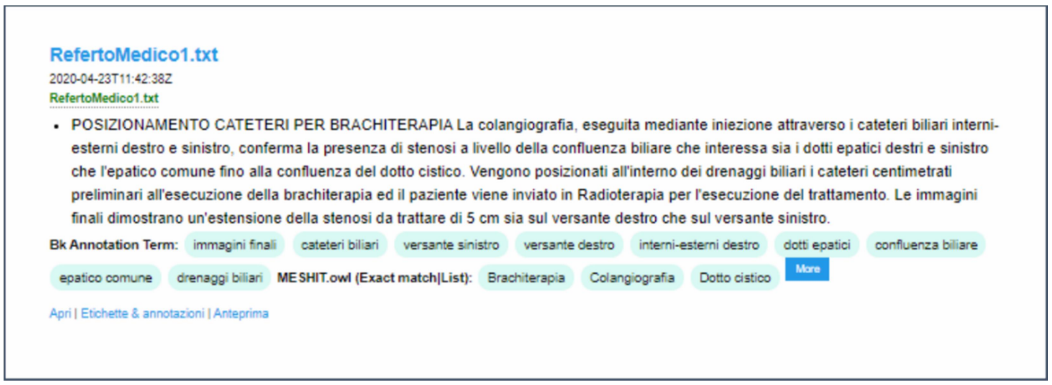
[4] https://nlp.stanford.edu/software/CRF-NER.html



- **Data enrichment plugins e enhancer**: attraverso questa componente è possibile migliorare ed arricchire i contenuti con dati aggiuntivi come metadati (ad esempio tag o annotazioni) o analisi (ad esempio OCR). Di particolare rilievo è questo componente per la sperimentazione oggetto del presente documento in quanto comprende.
  - il già citato **Solr Ontology Tagger** utilizzato per l'annotazione e la classificazione i referti medici in inglese relativi ai pazienti malati di COVID19, in relazione all'ontologia Doid; in figura 18 viene fornito un esempio di annotazioni relative all'ontologia Doid per un referto medico Covid19.



- Inoltre per la sperimentazione sui referti medici in italiano è stato altresì implementato un plugin ad hoc denominato **BkAnnotationPlugin** che viene richiamato a fronte di una indicizzazione per arricchire i contenuti richiamando l'apposito servizio REST di BigKnowledge deputato all'annotazione del documento in relazione all'ontologia MeSH, e una volta ottenuto il risultato dell'annotazione in formato json invia le annotazioni a Solr per la reindicizzazione dei documenti in modo che le nuove annotazioni possano essere rese fruibili attraverso le funzionalità di ricerca. La figura seguente mostra un esempio di referto medico in italiano con le annotazioni associate all'ontologia MeSH e la terminologia estratta attraverso il componente di InformationExtraction di BigKnowledge: nello specifico le annotazioni prodotte da BigKnowledge sono denominate **Bk Annotations**.



- Un altro importante plugin utilizzato in questa sperimentazione è il plugin per l'esportazione delle entità e delle associazioni tra esse in Neo4j[5].

[5] https://neo4j.com/

Oltre a quelle sopra citate, la piattaforma OpenSemanticSearch comprende anche le seguenti componenti:

- **Trigger**: il CMS o il file system (attraverso il monitoraggio del file system) notificherà al servizio web (API) quando ci sono nuovi dati o quando il contenuto viene modificato; in queso modo non sarà necessario utilizzare risorse aggiuntiva per frequenti nuove scansioni delle directory finalizzate all'indicizzazione dei nuovi contenuti.
- **Servizi Web** (**REST-API**): disponibili tramite il protocollo di rete standard HTTP e attivabili attraverso azioni specifiche dell'utente (mediante utilizzo dell'interfaccia di amministrazione Web) o un altro servizio (ovvero utilizzando l'API REST) per la scansione e l'indicizzazione di una directory o di una pagina Web.
- **Gestore code** (**Celery on RabbitMQ**[6]): gestione della coda delle attività e avvio dei job di estrazione del testo, analisi, arricchimento dei dati e indicizzazione mediante un giusto equilibrio di workers attivi contemporaneamente in parallelo.
- **Scheduler**: gestione dell'avvio dei job di indicizzazione pianificati. Questo può fatto mediante crontab per l'avvio dei Cron da riga di comando. Oppure è possibile pianificare i job di ri-scansionamento o di importazione dei dati mediante l'apposita funzionalità dell'interfaccia web di ManifoldCF[7].

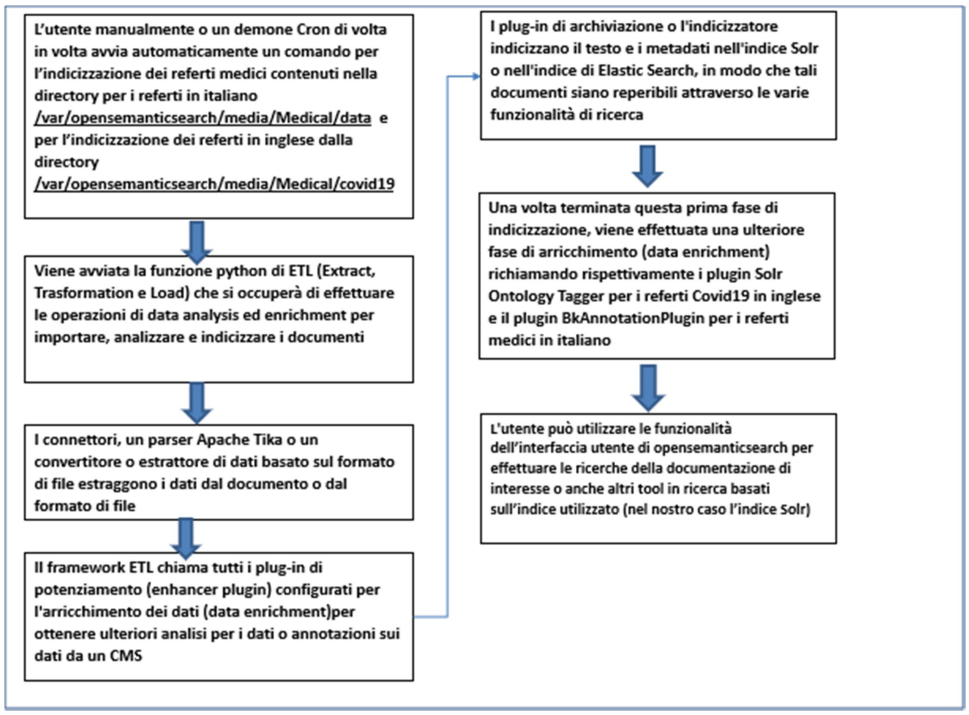
3.2.2. *Flusso operativo*

Nella figura che segue viene mostrato il flusso operativo che è stato adottato nella sperimentazione qui descritta, per gestire di volta in volta i nuovi referti medici con

[6] https://www.rabbitmq.com/

[7] https://manifoldcf.apache.org/

questi componenti di ETL (estrazione, trasformazione, caricamento), elaborazione dei documenti, analisi dei dati e arricchimento dei dati.

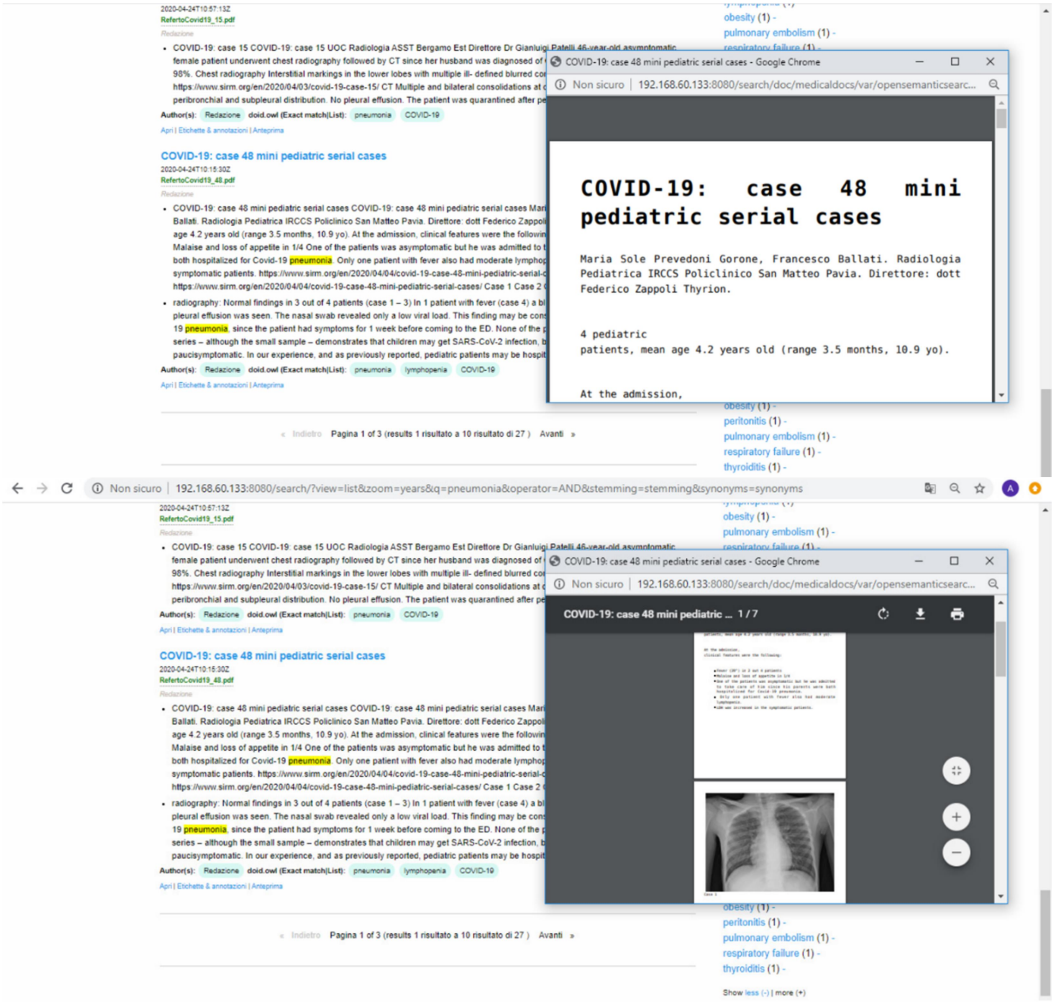


3.3. Casi d'uso

In questo paragrafo presentiamo alcuni esempi applicativi pratici di particolare rilievo per la classificazione dei referti in italiano e in inglese.

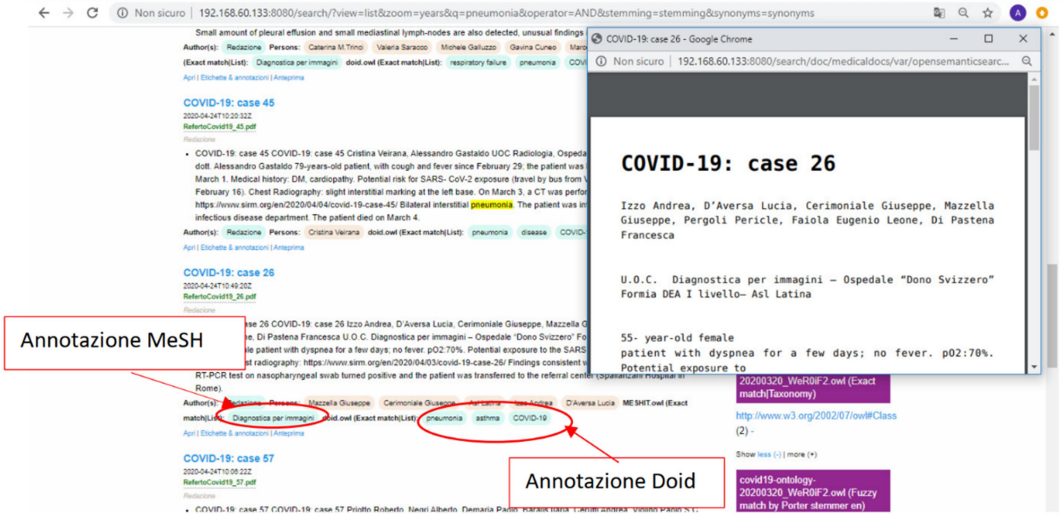
3.3.1. *Caso d'uso 1- Annotazione e classificazione referti Covid19*

Nelle figure che seguono viene mostrato il caso di una ricerca per keyword (pneumonia). Il documento mostrato è un referto in inglese di un caso COVD19 e tra le sue annotazioni la voce pneumonia presente nell'ontologia Doid.

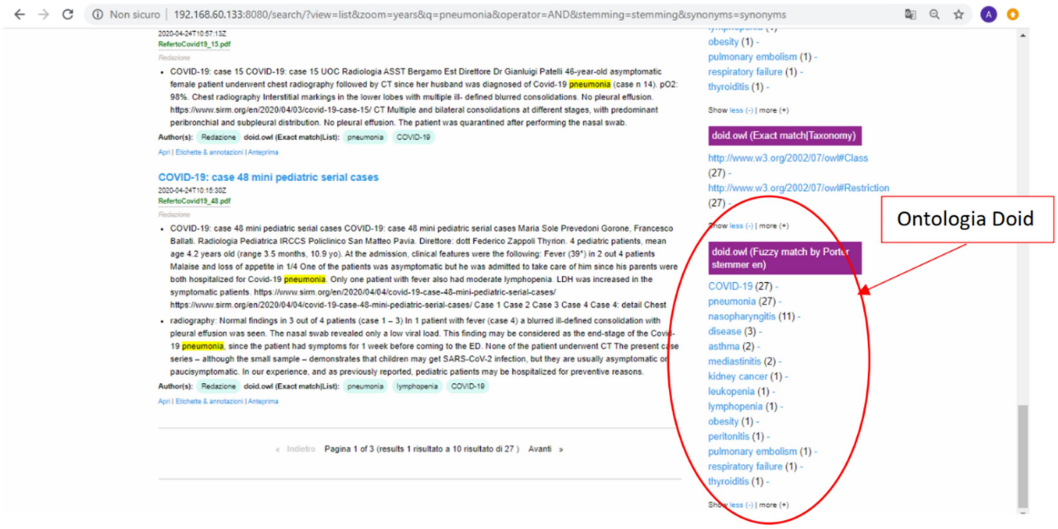


Nella figura seguente viene mostrato un referto in inglese sempre relativo ad un caso COVID19, in cui compaiono annotazioni sia per l'ontologia MeSH (voce diagnostica per immagini) che per l'ontologia Doid.

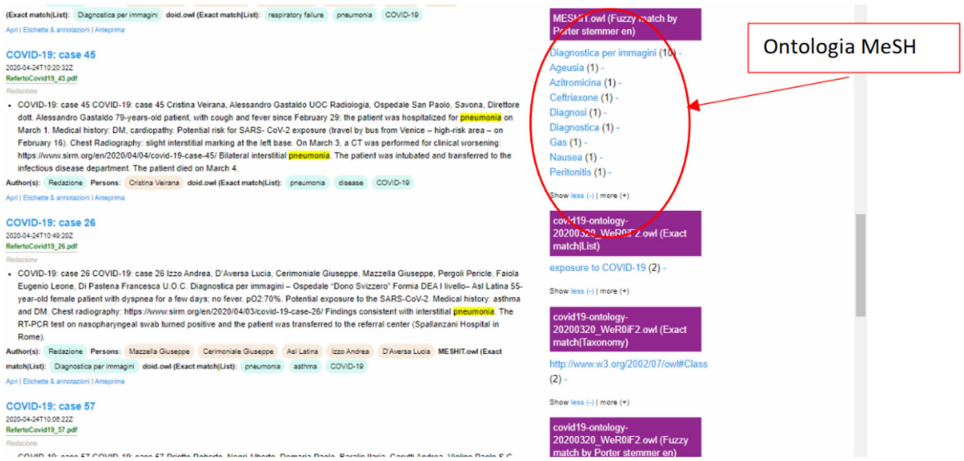
Infatti, il documento è un referto contenente delle immagini di diagnostica.



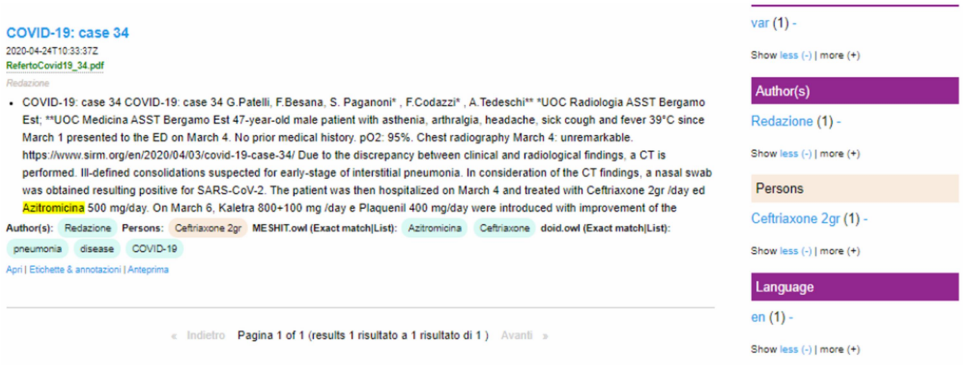
Visualizzazione ontologia Doid



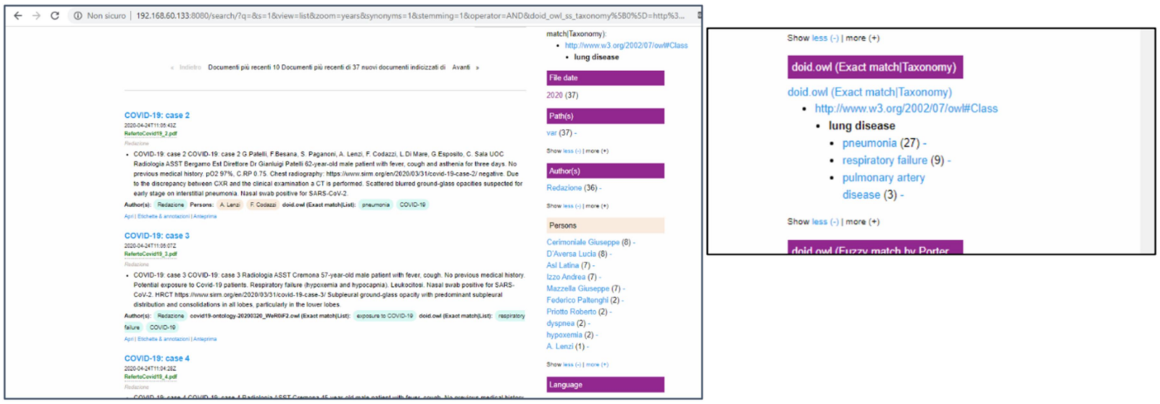
Visualizzazione ontologia Mesh



Nella figura che segue è presente un referto medico, che è stato classificato sia per l’ontologia mesh, in particolare con evidenza di annotazione per 2 medicinali (Azitromicina e Cifrìaxone) sia per l’ontologia Doid con evidenza in questo caso di annotazioni associate alle patologie pneumonia e COVID19. Questo esempio è particolarmente interessante perché permette di evidenziare come l’utilizzo combinato delle due ontologie ha permesso di ricavare una correlazione tra i farmaci adottati e le specifiche patologie rilevate nei pazienti COVID19.

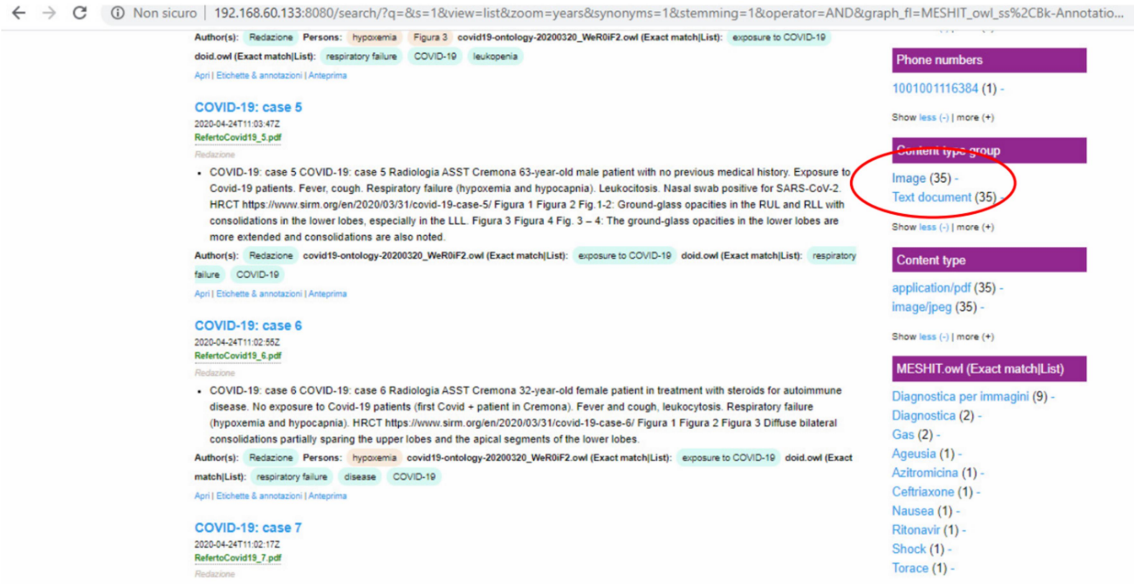


La figura che segue mostra come selezionando una voce dell’ontologia doid (lung disease) il sistema è in grado di ricercare tutti i documenti associati alle voci dell’ontologia sottostanti (in questo caso pneumonia, respiratory failure, pulmunary artery disease).



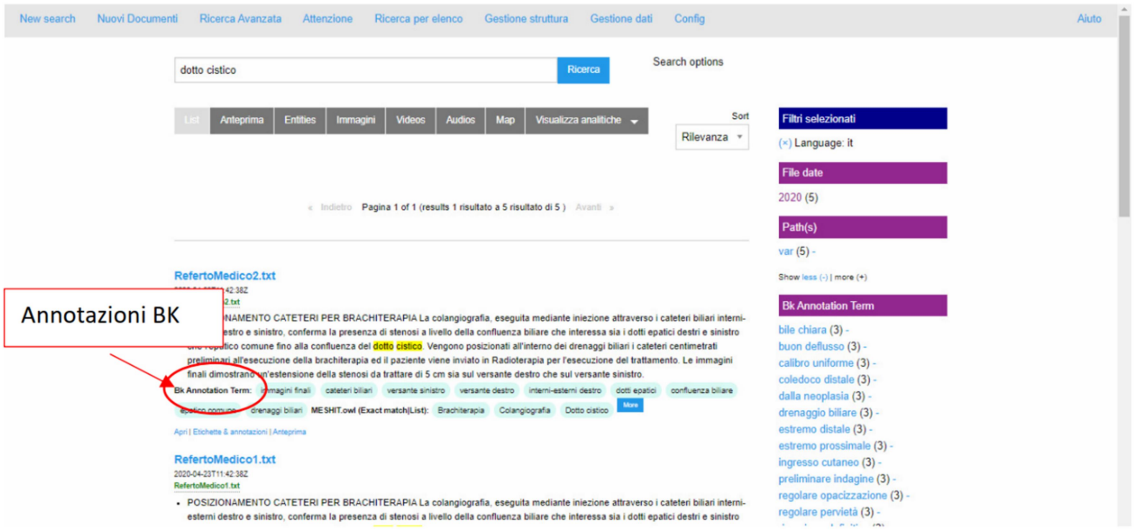
Tra i filtri è possibile selezionare anche quello relativo alla presenza di immagini all’interno del testo.

Il sistema visualizzerà i soli documenti che contengono immagini.

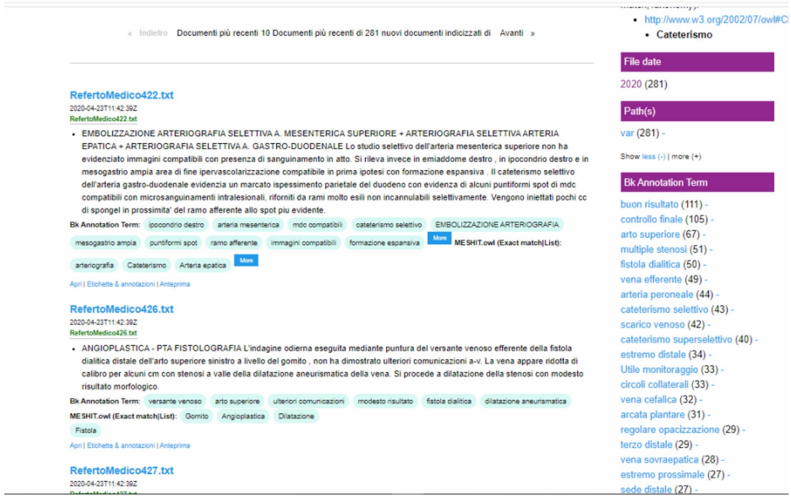


3.3.2. *Caso d'uso 2- Annotazione e classificazione referti medici in italiano*

Nella figura che segue viene mostrata la presenza per i referti in italiano di annotazioni relative sia all'ontologia MESH che estratte tramite Big Knowledge (identificate dalle annotazioni BkAnnotationTerm).

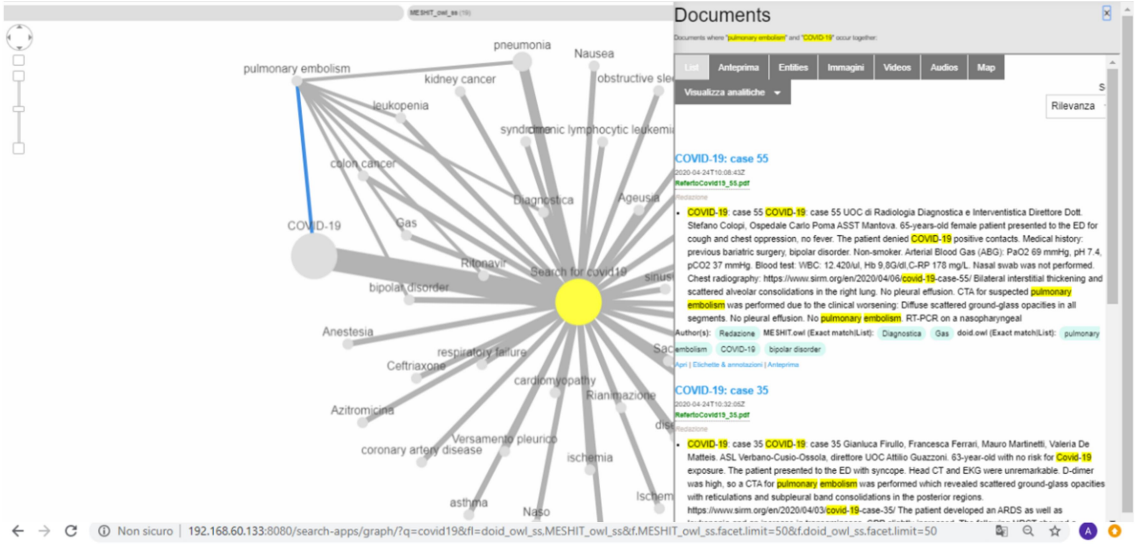


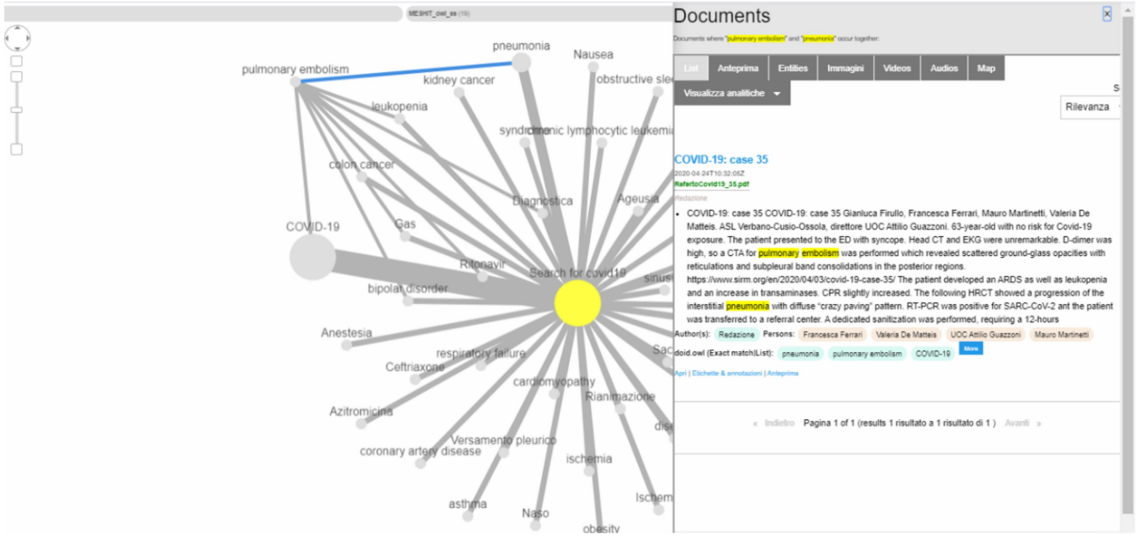
Visualizzazione dei documenti che sono associati a voci sottostanti dell'ontologia (ho selezionato cateterismo e il sistema visualizza anche documenti associati a categorie "figlie" di cateterismo nello specifico angioplastica).



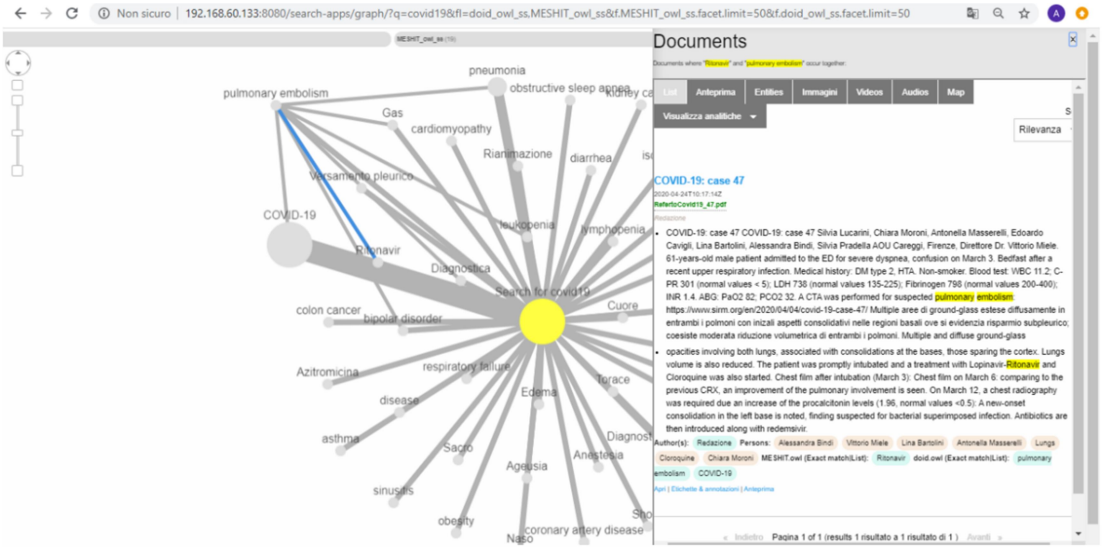
3.3.3. *Graph visualization*

Come mostrano le figure che seguono il sistema è anche in grado di fornire una visualizzazione a grafo delle relazioni presente tra le entità evidenziando anche i documenti in cui queste correlazioni sono presenti.

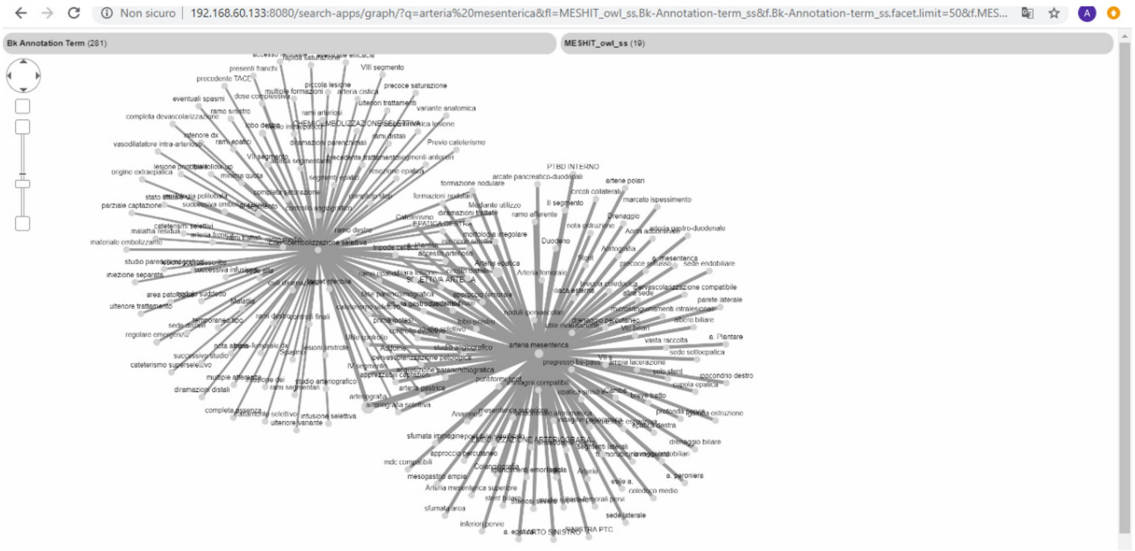


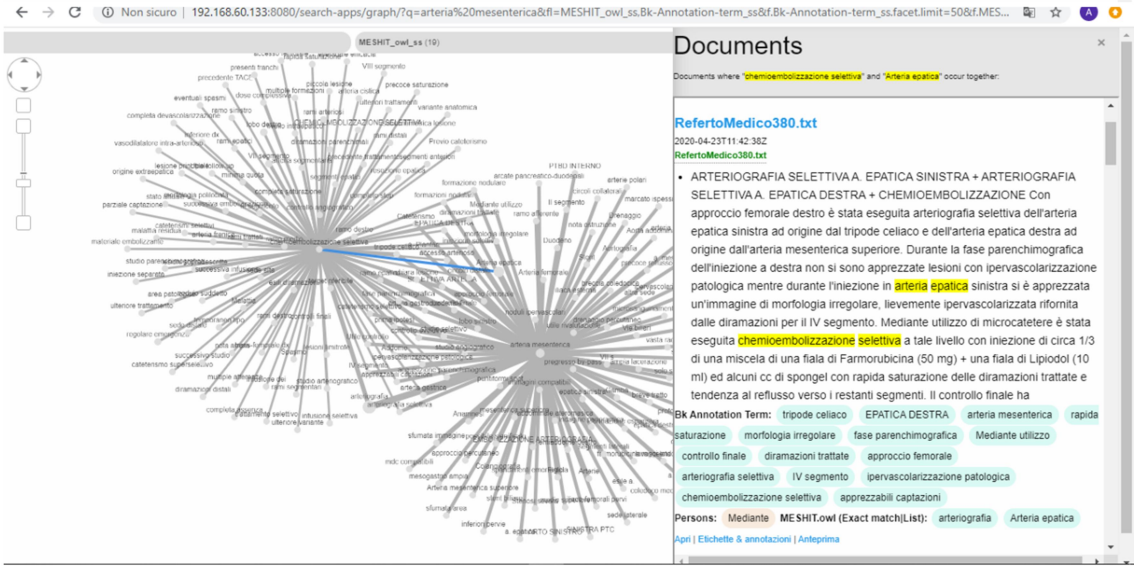


Un esempio molto interessante è rappresentato dalla figura seguente in cui vengono mostrate tramite grafo le associazioni tra i farmaci rilevati dall'ontologia MESH e le patologie rilevate dall'ontologia DOID.



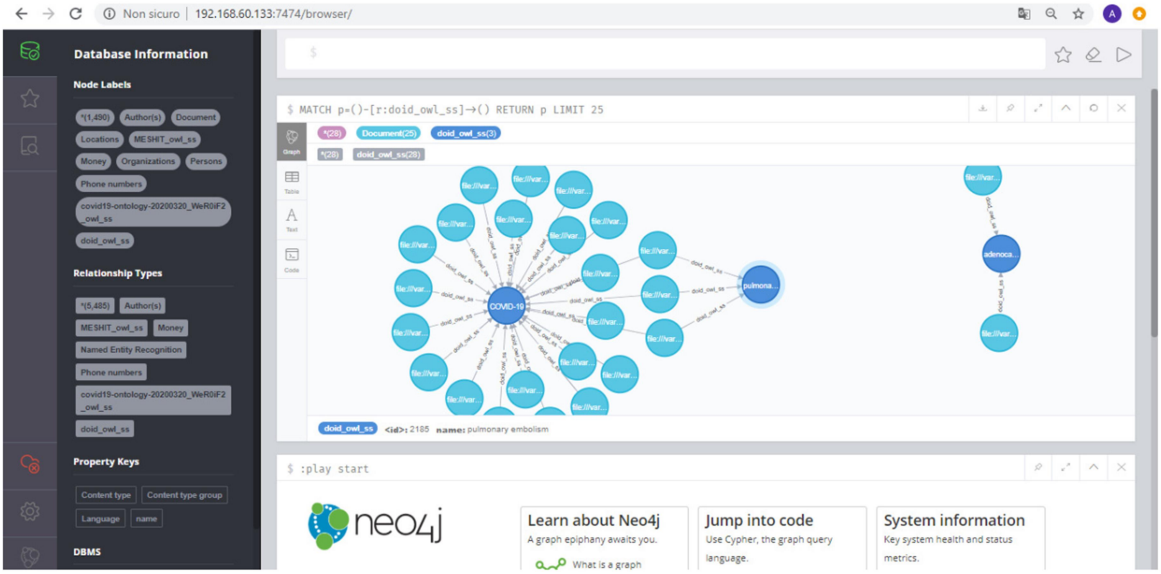
Nelle figure seguenti vengono invece mostrati gli incroci mediante rappresentazione a grafo tra l'ontologia MESH e la terminologia estratta da BigKnowledge (Bk Annotation Term) con evidenza e visualizzazione anche dei documenti in cui tali correlazioni sono presenti.





Infine, nella immagine che segue è presente il risultato dell’export effettuato mediante l’utilizzo dell’apposito plugin delle annotazioni e delle rispettive relazioni in un grafo Neo4J.

Come è possibile evincere osservando il grafo neo4j ottenuto il livello di dettaglio in questo caso è molto inferiore rispetto alla visualizzazione grafica ottenuta mediante il componente di graph visualization interno a opensemanticsearch. In questo caso infatti in relazione al caso COVID19, viene mostrata una sola associazione tra entità, quella credibilmente più marcata all’interno dei referti medici, ossia l’associazione tra la patologia COVID19 e la patologia pneumonia. È possibile tuttavia definire un differente livello di granularità del grafo modificando opportunamente la configurazione.


Non sicuro | 192.168.60.133:7474/browser/
Database Information
Node Labels
Relationship Types
Property Keys
DBMS
$ MATCH p=()-[r:doid_owl_ss]→() RETURN p LIMIT 25
COVID-19
pulmona...
adenoca...
doid_owl_ss <id>: 2185 name: pulmonary embolism
$:play start
neo4j
Learn about Neo4j
A graph epiphany awaits you.
What is a graph
Jump into code
Use Cypher, the graph query language.
System information
Key system health and status metrics.