Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Modello architetturale di un sistema per l'estrazione di informazioni da documenti sanitari

Angelo Esposito, Mario Sicuranza, Mario Ciampi

RT-ICAR-NA-2019-06 ***Data:*** **luglio 2019**

Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR) – Sede di Napoli, Via P. Castellino 111, I-80131 Napoli, Tel: +39-0816139508, Fax: +39-0816139531, e-mail: napoli@icar.cnr.it, URL: *www.icar.cnr.it*

Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni

Modello architetturale di un sistema per l'estrazione di informazioni da documenti sanitari

Angelo Esposito, Mario Sicuranza, Mario Ciampi

Rapporto Tecnico N: **RT-ICAR-NA-2019-06** ***Data:*** **luglio 2019**

Modello architetturale di un sistema per l'estrazione di informazioni da documenti sanitari

Angelo Esposito, Mario Sicuranza, Mario Ciampi
Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche
Via Pietro Castellino, 111 – 80131 Napoli, Italia

E-mail: {angelo.esposito, mario.sicuranza, mario.ciampi}@icar.cnr.it

Abstract

L'ICT negli ultimi ha pervaso il mondo della sanità proponendo standard e tecnologie capaci di supportare i professionisti sanitari nelle loro attività di cura dei pazienti. Uno degli aspetti maggiormente rilevanti nel contesto della sanità elettronica è l'utilizzo di standard per la codifica e la rappresentazione di contenuti informativi sanitari. Gli standard CDA e FHIR di HL7 ad oggi rappresentano un punto di riferimento per la codifica e strutturazione delle informazioni sanitarie a livello mondiale ed in Italia il trend, relativo all'adozione di tali standard, è in continua crescita. L'utilizzo di tali standard semplifica il processo di ricerca ed estrazioni delle informazioni dai documenti sanitari semplificando notevolmente il processo di cura del paziente (i professionisti sanitari possono consultare i documenti sanitari strutturati e ricercare le informazioni di interesse mediante l'ausilio di sistemi informativi). Tuttavia, questi standard consentono di definire documenti sanitari dove la parte narrativa non è opportunamente codificata, in tali casi l'estrazione delle informazioni dai documenti risulta particolarmente articolata e non semplice come nel caso di contenuti completamente codificati. Pertanto in questo rapporto tecnico si propone un modello architetturale di un sistema capace di estrarre contenuti informativi di interesse da documenti clinici semi-strutturati e non strutturati e di proporli in un formato strutturato secondo lo standard FHIR di HL7. Il modello architetturale proposto è definito da un insieme di componenti in grado di identificare specifiche informazioni mediche a partire da un documento clinico, ed offrire al medico la possibilità di richiedere e ottenere solo informazioni strettamente utili, in funzione alle specifiche esigenza informative. Il modello consente di selezionare facilmente i concetti medici da recuperare in un documento. L'obiettivo principale è quello di ricevere domande su informazioni cliniche, per recuperarle da un documento, e di conseguenza costruire risorse FHIR basate sulle informazioni acquisite ed estratte da un documento clinico.

Keywords: Modello Architetturale, Estrazioni di Informazioni Sanitarie di Interesse.

1. Introduzione

Il presente rapporto tecnico propone un modello architetturale di un sistema capace di indicizzare e strutturare automaticamente documenti clinici testuali attraverso l'applicazione di tecniche di Natural Language Processing e Information Extraction [1]. Allo scopo di rendere il modello capace di identificare ed estrarre informazioni cliniche di interesse è stata svolta una analisi delle entità mediche maggiormente rilevanti per il processo di cura del paziente presenti nei documenti clinici sanitari. Il modello è capace di individuare ed estrarre tali entità da documenti scritti in linguaggio naturale. In particolare il risultato dell'attività di analisi ha portato all'individuazione di nove entità mediche rilavanti per la cura del paziente presenti in una moltitudine di documenti clinici sanitari sulla base delle quali è stato definito un modello informativo, presentato nel prossimo paragrafo. Il modello informativo definito è costituito da nove entità mediche di

interesse, tra cui le informazioni demografiche del paziente, le informazioni sulle problematiche, sulle allergie, sulle procedure, sulle osservazioni, sulle immagini diagnostiche, sulle prescrizioni, sulle dispensazioni e sui piani di cura relativi al paziente.

Il modello architetturale di analisi per l'indicizzazione e la strutturazione di documenti testuali (paragrafo 3) è capace di analizzare un documento clinico dato in input e l'espressione della necessità informativa, il riconoscimento delle entità e delle informazioni mediche di interesse ed estrarre le stesse attraverso l'applicazione di tecniche di NLP e Information Extraction (IE) [2, 16]. Le informazioni estratte sono codificate e strutturate secondo il formato FHIR [3,14] in modo da poter essere elaborate automaticamente ed indicizzate.

2. Modello informativo delle risorse

In questo capitolo è descritto il modello informativo definito a partire dall'individuazione delle entità informative di interesse che potenzialmente possono essere processate in maniera automatica mediante l'applicazione di algoritmi di estrazione ed analisi testuale.

2.1 Entità mediche di interesse

In questo paragrafo sono descritte le entità mediche di interesse che è possibile estrarre da documenti sanitari automaticamente mediante l'applicazione di tecniche e di algoritmi descritti nel prossimo capitolo. L'individuazione ha previsto l'analisi della tipologia e la strutturazione delle informazioni presenti nei documenti clinici sanitari. Dall'analisi effettuata sono stati individuate le seguenti entità di interesse:

- **Paziente:** Questa entità è costituita da attributi demografici relativi al paziente a cui si riferisce il documento sanitario. Tali informazioni risultano propedeutiche alle procedure amministrative, finanziarie e logistiche svolte dalla piattaforma. I campi che costituiscono tale entità possono variare tra le varie tipologie di documenti clinici, ma risultano abbastanza simili per essere mappati in un'unica entità.
- **Allergia:** Questa entità è costituita da attributi che riguardano una valutazione (o evidenza) clinica di un'allergia o intolleranza; una propensione o un potenziale rischio per un individuo di presentare una reazione avversa all'esposizione di una sostanza (allergene) specificata o alla classe di sostanze specificata. Questa entità rappresenta una condizione di suscettibilità ad una sostanza, con un insieme di attributi che caratterizzano eventi e / o sintomi di supporto e non ha alcuna relazione diretta.
- **Problema:** Questa entità è costituita da attributi che riguardano informazioni dettagliate su una condizione, un problema, una diagnosi o un altro evento clinico, situazione, problema o concetto clinico che ha raggiunto un livello di preoccupazione.
- **Procedura:** Questa entità è costituita da attributi che riguardano procedure cliniche eseguite su un paziente. Una procedura è un'attività che viene eseguita con o su di un paziente come parte della assistenza clinica. Gli esempi includono procedure chirurgiche, procedure diagnostiche, procedure endoscopiche, biopsie, consulenze, fisioterapia, esercizio fisico, ecc. Le procedure possono essere eseguite da un operatore sanitario o in alcuni casi dal paziente stesso.
- **Osservazione:** Questa entità è costituita da attributi che riguardano le osservazioni cliniche effettuate da un professionista sanitario. Queste osservazioni possono includere segni vitali come peso corporeo,

pressione sanguigna e temperatura; dati di laboratorio come glicemia o la velocità di filtrazione glomerulare (GFR) stimata; caratteristiche personali: come il colore degli occhi etc.

- **Immagine Clinica:** Questa entità è costituita da attributi che forniscono informazioni su uno studio di immagini DICOM [4] e sugli oggetti della serie e di immagini nel particolare studio DICOM. Questa entità fornisce una mappatura dei suoi elementi con gli attributi DICOM.
- **Prescrizione:** Questa entità è costituita da attributi che forniscono informazioni relative ad una prescrizione effettuata per un paziente. Le informazioni relative alla prescrizione riguardano inoltre la posologia nel caso in cui essa si riferisca ad un farmaco.
- **Dispensazione:** Questa entità è costituita da attributi che caratterizzano il farmaco dispensato al paziente. L'entità può essere utilizzata in diversi contesti che includono l'erogazione e il prelievo da una farmacia ambulatoriale, erogazione di farmaci specifici (intera confezione) per il paziente dalla farmacia di reparto ospedaliero, così come emissione di una singola dose dal reparto per il paziente.
- **Piano di cura:** Questa entità è costituita da attributi che caratterizzando un piano di cura che un paziente deve seguire durante un percorso di assistenza. Questa entità indica le azioni sanitarie da intraprendere per il trattamento di un paziente in relazione a vincoli temporali e allo stato del paziente nei diversi step di cura dello stesso.

2.2 Modello informativo dei dati

In questo paragrafo è riportato il modello informativo dei dati definito per gestire in maniera automatica l'indicizzazione e il trattamento delle informazioni socio sanitarie di interesse per la cura del paziente mediante le informazioni estratte dai documenti sanitari. Il modello informativo è rappresentato mediante un diagramma UML [7] delle classi (entità). Il diagramma UML mostra l'indicazione delle molteplicità nelle associazioni tra le entità informative. Con 0…1 si indica che tale relazione è opzionale e al più esiste un unico elemento associato alla relazione tra classi.

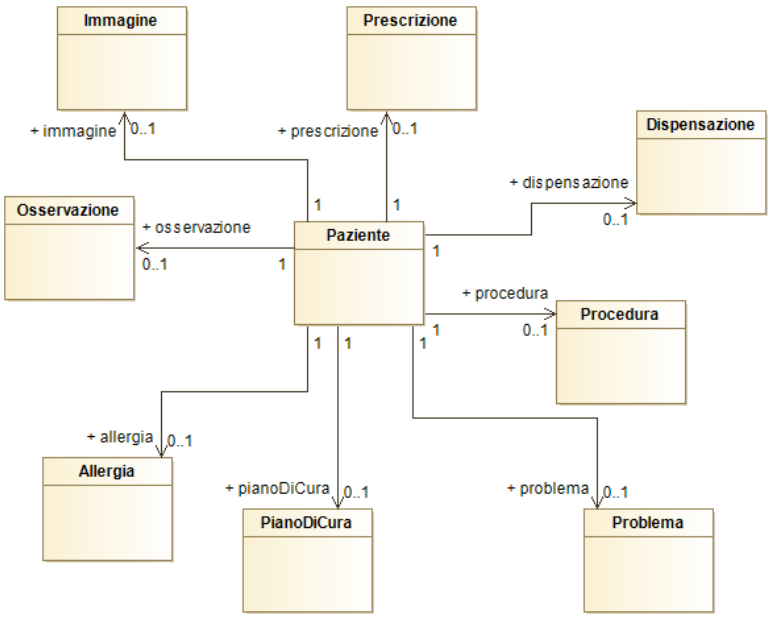

Figure 1 - Class Diagram modello informativo

3. Modello per l'estrazione delle informazioni di interesse da documenti clinici

Il modello architetturale proposto è definito da un insieme di componenti in grado di identificare specifiche informazioni mediche a partire da un documento clinico, ed offre al medico la possibilità di richiedere e ottenere unicamente informazioni strettamente utili, in funzione alle specifiche esigenza informative.

Il modello consente di selezionare facilmente i concetti medici da recuperare in un documento. L'obiettivo principale è quello di ricevere domande su informazioni cliniche, per recuperarle da un documento, e di conseguenza costruire risorse FHIR basate sulle informazioni acquisite ed estratte da un documento clinico [15].

Dato un documento clinico da cui estrarre le informazioni di interesse, il modello è in grado, di ricevere una query, tramite la componente **Interfaccia** e di elaborala per comprendere le necessità informative del richiedetene. Dopo la comprensione della query il documento ricevuto in input viene elaborato con l'obiettivo di estrarre le informazioni richieste, e presentare tali informazioni estratte come risorse FHIR. La Figura seguente mostra il modello architetturale completo, e illustra le modalità di interazione tra le sue componenti che sono:

- **La componente di interfaccia**: prende come input un documento semi strutturato in CDA [8] o in linguaggio naturale e una query che indica la necessità informativa;
- **La componente di mappatura**: contiene varie tabelle che definiscono una mutua associazione tra un elemento FHIR e un altro rappresentato mediante lo standard CDA o mediante un concetto/relazione individuato mediante l'analisi del linguaggio naturale;
- **La componente di estrazione**: identifica le informazioni richieste nel documento mediante la componente di mappatura e si occupa dell'estrazione delle informazioni di interesse dal documento;
- **La Componente di costruzione**: aggrega i valori estratti dalla componente di estrazione in una o più risorse FHIR e restituisce le informazioni in risposta alla query.

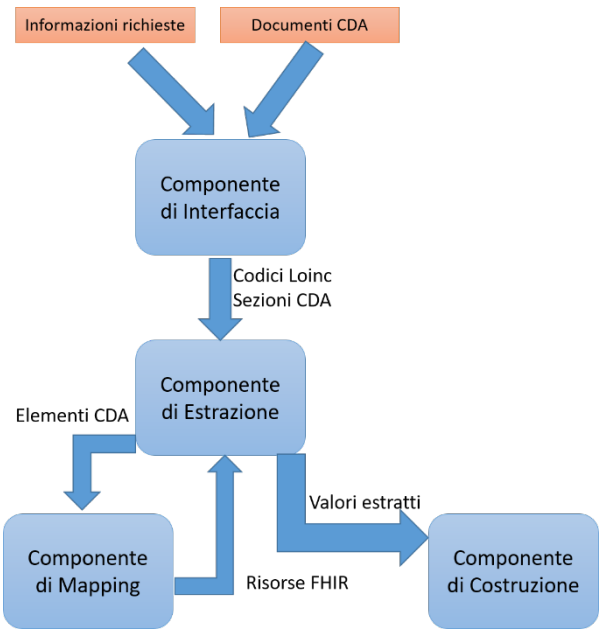

Figure 2 - Architettura del modello proposto

3.1 La componente interfaccia

La componente di interfaccia consente all'utente di selezionare la tipologia di documenti da analizzare (testo libero o semi strutturati). Inoltre, l'utente può scegliere le informazioni cliniche da recuperare che sono mappati dal sistema con una specifica risorsa FHIR. Nel caso di documenti a testo libero indirizza l'individuazione di concetti e relazioni verso il contesto (discorso) di interesse. Le risorse FHIR identificate possono fare riferimento a concetti amministrativi, come il concetto paziente, il concetto professionista e il concetto organizzazione, così come una varietà di concetti clinici, come la lista di problemi, dei farmaci, e la diagnostica, e così via. Grazie a questa componente, il professionista può utilizzare un semplice menu a tendina per selezionare i tipi di input:

- Tipologia di documenti in input (testo libero o semi strutturato) dal quale estrarre le risorse;
- Risorsa FHIR che deve essere recuperata;

Al termine del recupero delle informazioni, la componente di Interfaccia consente al medico di osservare facilmente il risultato della query come un semplice elenco di risorse. Gli elementi delle risorse saranno valorizzati con le informazioni cliniche estratte dai documenti in input.

3.2 La componente di mapping

Il nucleo del modello architetturale è rappresentato dalla componente di Mapping. Essa contiene diversi schemi di trasformazione che permettono di mappare un concetto / relazione in una risorsa FHIR. La componente di mapping contiene i principali concetti di sanità associate ad un paziente, come le allergie, i farmaci, i problemi, i segnali vitali, i piani di cura e così via. In particolare il sistema proposto è capace di trasformare i concetti e le relazioni contenuti nelle sezioni del CDA o estratti dai documenti in linguaggio naturale in risorse equivalenti FHIR. Questa risorsa può esprimere la stessa informazione medica mediante la rappresentazione FHIR. L'implementazione della componente di Mapping richiede la conoscenza della struttura del modello estratto mediante l'NLP dall'insieme di documenti in input. Inoltre, le risorse FHIR nelle quali mappare i concetti e le relazioni individuati mediante l'analisi devono essere preliminarmente identificate.

La proposta è costituita da una serie di schemi di mapping tra le informazioni presenti nelle sezioni di un CDA, una entità medica o relazione individuata mediante l'analisi NLP in una risorsa FHR, tramite una associazione di attributi.

La componente di mapping riceve gli elementi dalla componente di Interfaccia, trova la corrispondenza in FHIR mediante gli schemi implementati. Le informazioni relative al posizionamento degli oggetti informativi di interesse nel documento sono fornite alla componente di estrazione che si occupa di estrarre tali informazioni.

Nella fase preliminare, è stato identificato un modello di costruzione del Patient Summary [9] in FHIR. Per supportare la continuità di cura, il Patient Summary in italiano deve contenere elementi capaci di offrire un quadro di insieme generale della storia clinica del paziente. Lo scopo del documento è migliorare la qualità e garantire la continuità di cura. Il medico medicina generale produce il documento PS, ma diversi professionisti sono abilitati a consultarlo. Il Patient Symmary contiene le informazioni maggiormente rilevanti del soggetto, con dettagli circa le reazioni avverse, i farmaci utilizzati dal paziente, e i vaccini, informazioni circa le diagnosi e la condizione del paziente e i risultati di laboratorio.

Grazie al nostro modello le risorse FHIR possono essere estratte a runtime dal documento CDA, mediante gli schemi della componente Mapping. Il clinico può ottenere e gestire la granularità delle informazioni di

interesse dell'intero documento.

Nella figura seguente è mostrato la corrispondenza tra le compenti dell'header CDA e le componenti delle risorse FHIR, inoltre è mostrato a titolo esemplificativo il mapping tra la sezione degli alert presente body del CDA e la risorsa Allergia Intolleranza in FHIR.

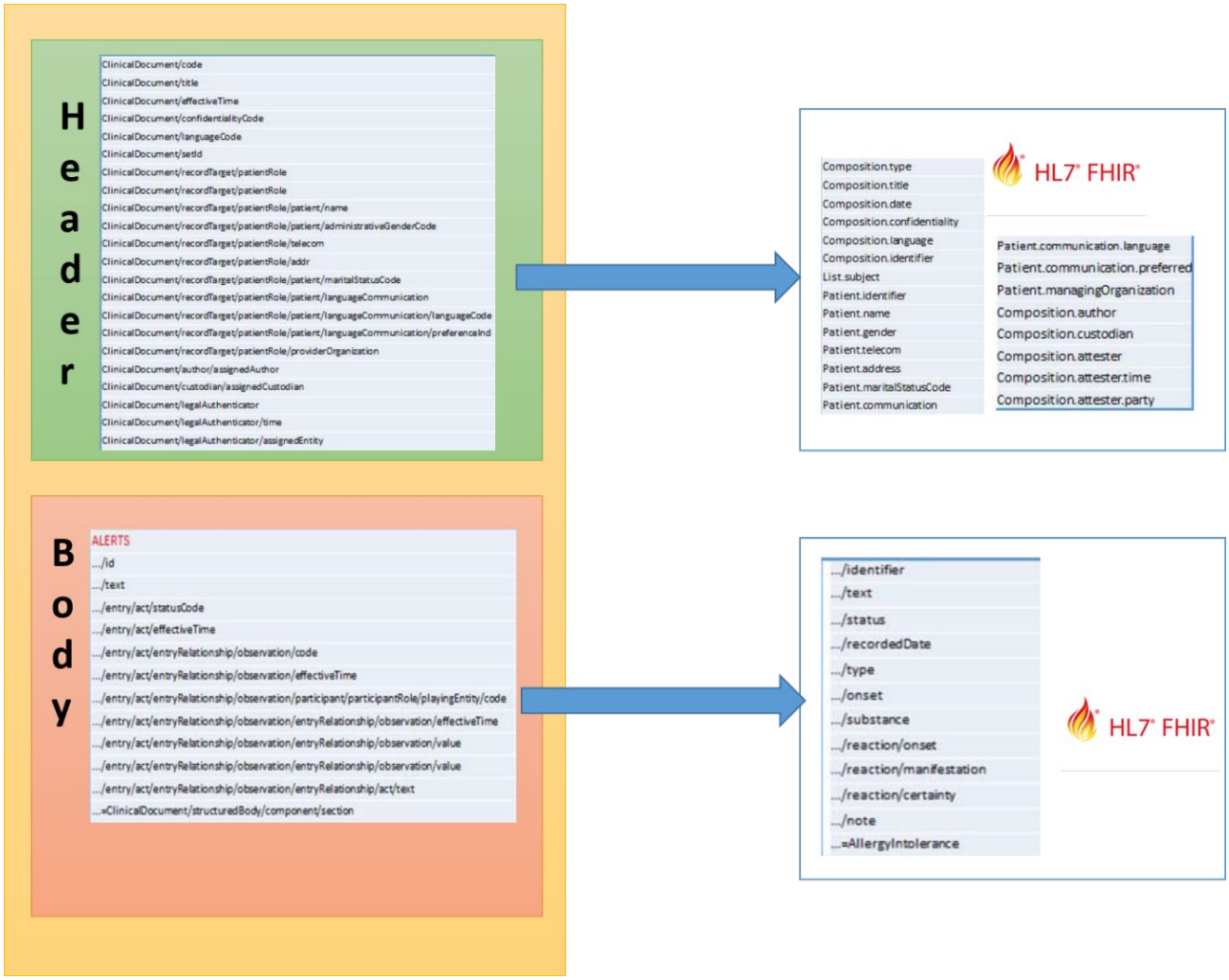

Figure 3 - Corrispondenza tra CDA e FHIR

3.3 La componente di estrazione

La componente di estrazione elabora la richiesta di un utente, inserita mediante la componente di interfaccia. Il primo step effettuato dalla componente riguarda l'analisi dei documenti in linguaggio naturale dati in input. Il secondo step riguarda l'individuazione della necessità informativa espressa dal professionista sanitario mediante il modulo di Query dell'interfaccia. La componente di estrazione applica tecniche di NLP allo scopo di individuare, mediante tecniche di tokenizzazione [10], lemmizzazione e mediante l'utilizzo di dizionari, vocabolari e UMLS [11], i concetti chiave e le relazioni tra essi. In dettaglio, la componente di Estrazione identifica i concetti clinici di interesse espressi mediante la query o identificati mediante l'applicazione di tecniche di NLP. In caso di input strutturato o semi struttura il modulo conosce le differenti sezioni possibili e associa ad essa un codice LOINC [12], ad esempio il codice "10160-0" identifica la sezione rappresentativa i farmaci assunti dal paziente in un documento strutturato nello standard CDA.

Nel caso di documento strutturato CDA, la componente di Estrazione deve trovare la sezione CDA contenente il codice LOINC associato alla informazione medica richiesta. Una volta identificata la sezione nel documento, è invocata la componente di mapping al fine di individuare l'equivalente concetto FHIR degli elementi della sezione CDA. Grazie a questa componente, i valori necessari a creare la risorsa FHIR possono essere estratti e possono divenire input per la componente di costruzione. La componente di estrazione utilizza query di tipo XPath per analizzare il Documento CDA, identificare il codice LOINC associato alla richiesta clinica, estrarre tutti gli elementi della sezione di interesse, e identificare la corrispondenza con FHIR attraverso gli schemi di mappatura.

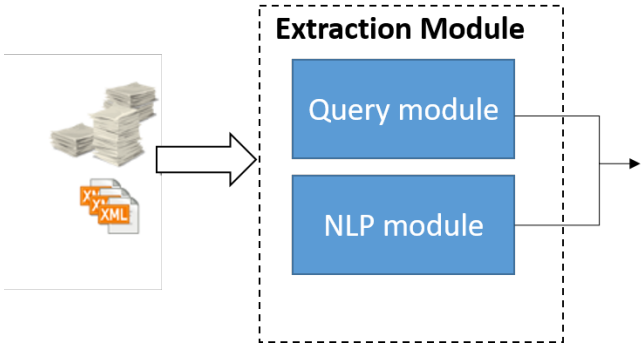

Figure 4 - Modulo di Estrazione

Nel caso di insiemi di documenti in linguaggio naturale e testo libero, il modulo NLP, individua tutti i concetti e mediante la componente di mapping è capace di individuare l'equivalente risorsa FHIR.

Il modulo NLP è costituito da 5 sotto moduli differenti:

- **Modulo di identificazione del linguaggio**: consente di identificare il linguaggio relativo al documento fornito in input (mediante la componente interfaccia).
- **Modulo di tokenizzazione**: si occupa di dividere il testo in token significativi e di conseguenza individuare i concetti di interesse.
- **Modulo di analisi lessicale**: consente l'assegnazione di tag ai token, i tag sono sia label grammaticali che label di dominio medico.
- **Modulo di applicazione delle regole di Parsing**: consente di annotare le combinazioni di token di interesse, con espressioni positive, espressioni negative e relazioni tra concetti.
- **Modulo di riconoscimento delle entità mediche**: consente, partendo dalle espressioni, concetti e relazioni, di evidenziare le entità mediche presenti nei documenti narrativi.

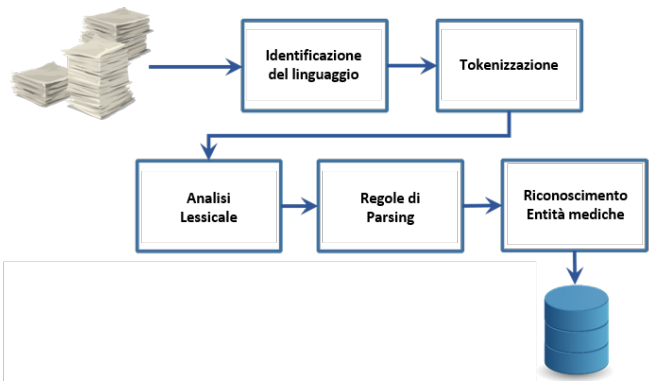

Figure 5 - Modulo NLP

3.4 La componente di raggruppamento

Il sottosistema per il raggruppamento dei dati, permette una fase di filtraggio e aggregazione dei dati capace di raggruppare i concetti, e le relazioni relative ad un unico paziente. Il sottosistema è costituito da due moduli differenti, il modulo detto di **Filtraggio** e il modulo detto di **Aggregazione dei dati.**

Questo sottosistema consente di prendere come input parole chiavi cliniche ed effettuare le seguenti operazioni:

- mappare e riconoscere le parole chiave all'interno del testo grazie alla componente di mappatura all'interno dei documenti;
- assegnare un identificatore univoco al documento (contenente un'etichetta con il nome del paziente);
- memorizzare un elenco contenente tutte le entità mediche individuate;
- associare all'entità mediche individuate una etichetta e ad un valore;
- aggregare informazioni per tipo presentando un riassunto visivo ai medici.

I concetti e le informazioni ottenute, possono essere classificate ad esempio in problemi, procedure mediche, codifiche ICD di malattie, farmaci etc.

3.5 La componente di costruzione

La componente di costruzione riceve le informazioni estratte dai documenti clinici di interesse per la costruzione di una risorsa FHIR equivalente da un punto di vista informativo. La risorsa FHIR così costruita rappresenterà l'informazione medica richiesta o individuata dal modulo NLP come concetto (o relazione) di interesse e nel caso di documento strutturato (CDA) conterrà gli stessi dati clinici della sezione CDA. Pertanto, la componente di costruzione restituirà in risposta alla richiesta una o più risorse FHIR che mappano le informazioni di interesse estratte. Questa componente può essere basata su una serie di funzioni scritte in XSLT [13] per realizzare la trasformazione degli oggetti contenuti nel documento CDA in risorse FHIR. Infine, la componente Costruzione restituisce la risposta alla componente di Interfaccia, che presenta in modo leggibile un riepilogo della risorsa informativa al professionista sanitario.

In caso di concetto restituito dal modulo NLP, la componente di costruzione consente di associare al concetto una specifica risorsa FHIR. Nel prossimo paragrafo è illustrata la modalità in cui opera il modello architetturale proposto. In particolare, la sezione mostra una possibile richiesta su di una specifica informazione di interesse estratta dalla storia clinica del paziente presente nel documento di Patient Summary.

3.6 Strutturazione delle entità mediche

In questo paragrafo è descritta la strutturazione delle entità mediche descritte nel paragrafo 2. La strutturazione dell'entità riguarda l'insieme di attributi che costituiscono l'entità stessa e la tipologia degli stessi. Allo scopo di rappresentare in modo formale le entità individuate e descritte nel precedente paragrafo è stato scelto di utilizzare la sintassi XML [5] ed in particolare XSD [6].

- **Entità Paziente**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
            xmlns:tns="http://schema/paziente.xsd">
```

```
    <xsd:complexType name="Paziente">
        <xsd:sequence>
            <xsd:element name="CodiceFiscale" type="xsd:string"/>
            <xsd:element name="Nome" type="xsd:string"/>
            <xsd:element name="Telefono" type="xsd:string"/>
            <xsd:element name="Sesso" type="xsd:string"/>
            <xsd:element name="DataDiNascita" type="xsd:date"/>
            <xsd:element name="Indirizzo" type="xsd:string"/>
            <xsd:element name="LuogoDiNascita" type="xsd:string"/>
            <xsd:element name="MedicoCurante" type="xsd:string"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Allergia**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
           xmlns:tns="http://schema/allergia.xsd">
    <xsd:complexType name="Allergia">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="statoClinico" type="xsd:string"/>
            <xsd:element name="tipoAllergia" type="xsd:string"/>
            <xsd:element name="categoriaAllergia" type="xsd:string"/>
            <xsd:element name="criticità" type="xsd:string"/>
            <xsd:element name="codiceAllergia" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="dataInizio" type="xsd:dateTime"/>
            <xsd:element name="dataFine" type="xsd:dateTime"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="reazione" type="xsd:string"/>
            <xsd:element name="sostanza" type="xsd:string"/>
            <xsd:element name="minifestazione" type="xsd:string"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Problema**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
           xmlns:tns="http://schema/allergia.xsd">
    <xsd:complexType name="Problema">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="statoClinico" type="xsd:string"/>
            <xsd:element name="tipoProblema" type="xsd:string"/>
            <xsd:element name="codiceProblema" type="xsd:string"/>
            <xsd:element name="gradoSeverità" type="xsd:string"/>
            <xsd:element name="categoria" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="manifestazione" type="xsd:date"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="minifestazione" type="xsd:date"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Procedura**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
           xmlns:tns="http://schema/procedura.xsd">
    <xsd:complexType name="Procedura">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="parteInteressata" type="xsd:string"/>
            <xsd:element name="stato" type="xsd:string"/>
            <xsd:element name="ragione" type="xsd:string"/>
            <xsd:element name="parteDi" type="xsd:string"/>
            <xsd:element name="azioneCompiuta" type="xsd:string"/>
            <xsd:element name="categoria" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="followUp" type="xsd:string"/>
            <xsd:element name="chirurgo" type="xsd:string"/>
            <xsd:element name="complicazioni" type="xsd:string"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="dataProcedura" type="xsd:date"/>
```

```
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Osservazione**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
           xmlns:tns="http://schema/osservazione.xsd">
    <xsd:complexType name="Osservazione">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="tipologia" type="xsd:string"/>
            <xsd:element name="stato" type="xsd:string"/>
            <xsd:element name="codice" type="xsd:string"/>
            <xsd:element name="contesto" type="xsd:string"/>
            <xsd:element name="interpretazione" type="xsd:string"/>
            <xsd:element name="commento" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="dataOsservazione" type="xsd:date"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Immagine Clinica**

```
<xsd:schema xmlns:xsd="http:/www.w3.org/2001/XMLSchema"
           xmlns:tns="http:/schema/immagine.xsd">
    <xsd:complexType name="Immagine">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="idImmagine" type="xsd:string"/>
            <xsd:element name="disponibilità" type="xsd:string"/>
            <xsd:element name="contesto" type="xsd:string"/>
            <xsd:element name="tipologia" type="xsd:string"/>
            <xsd:element name="ragione" type="xsd:string"/>
            <xsd:element name="ParteInteressata" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="codiceOperatore" type="xsd:string"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="dataImmagine" type="xsd:date"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Prescrizione**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
           xmlns:tns="http://schema/prescrizione.xsd">
    <xsd:complexType name="Prescrizione">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="stato" type="xsd:string"/>
            <xsd:element name="principioAttivo" type="xsd:string"/>
            <xsd:element name="contesto" type="xsd:string"/>
            <xsd:element name="informazioniAggiuntive" type="xsd:string"/>
            <xsd:element name="posologia" type="xsd:string"/>
            <xsd:element name="codiceFiscalePrescrittore" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="dataPrescrizione" type="xsd:date"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Dispensazione**

```
<xsd:schema xmlns:xsd="http:/www.w3.org/2001/XMLSchema"
           xmlns:tns="http:/schema/dispensazione.xsd">
    <xsd:complexType name="Dispensazione">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="stato" type="xsd:string"/>
            <xsd:element name="principioAttivo" type="xsd:string"/>
            <xsd:element name="codiceFarmaco" type="xsd:string"/>
            <xsd:element name="contesto" type="xsd:string"/>
```

```
            <xsd:element name="riferimentoPrescrizione" type="xsd:string"/>
            <xsd:element name="posologia" type="xsd:string"/>
            <xsd:element name="codiceFiscaleDispensatore" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="dataDispensazione" type="xsd:date"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

- **Entità Piano di cura**

```
<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema"
           xmlns:tns="http://schema/pianodicura.xsd">
    <xsd:complexType name="PianoDiCura">
        <xsd:sequence>
            <xsd:element name="identificativo" type="xsd:string"/>
            <xsd:element name="stato" type="xsd:string"/>
            <xsd:element name="tipologia" type="xsd:string"/>
            <xsd:element name="descrizione" type="xsd:string"/>
            <xsd:element name="periodoDA" type="xsd:date"/>
            <xsd:element name="periodoA" type="xsd:date"/>
            <xsd:element name="obiettivo" type="xsd:string"/>
            <xsd:element name="attività" type="xsd:string"/>
            <xsd:element name="codiceFiscaleMedico" type="xsd:string"/>
            <xsd:element name="codiceFiscalePaziente" type="xsd:string"/>
            <xsd:element name="note" type="xsd:string"/>
            <xsd:element name="motivazione" type="xsd:date"/>
            <xsd:element name="DescrizioneAttività" type="xsd:string"/>
            <xsd:element name="PianificazioneIncontri" type="xsd:string"/>
        </xsd:sequence>
    </xsd:complexType>
</xsd:schema>
```

3.7 Casi d'uso

Lo scopo di questa sezione è illustrare un caso d'uso reale, nel quale un professionista sanitario a partire da un documento ottiene una risorsa FHIR che rappresenta il suo bisogno informativo. In particolare, dato un documento clinico di tipo Patient Summary, un medico di medicina generale può consultarlo e ottenere il nucleo delle informazioni sanitarie di un paziente. In accordo con l'Implementation Guide di HL7 Italia, il Patient Summary è caratterizzato dalle seguenti sezioni:

- Lista degli Alert - sezione richiesta;
- Lista di farmaci - sezione richiesta;
- Lista delle Vaccinazioni - sezione richiesta;
- Lista di Problemi - sezione richiesta;
- Storia familiare - sezione opzionale;
- Storia sociale - sezione richiesta;
- Storia delle gravidanze - sezione opzionale;
- Segni vitali - sezione opzionale;
- Attrezzatura medica - sezione richiesta;
- Piano di cura - sezione opzionale;
- Procedure - sezione raccomandata;
- Incontri - sezione opzionale;
- Stato funzionale - sezione richiesta;
- Risultati - sezione consigliata;
- Direttive avanzate - sezione facoltativa;
- Motivo per esenzione dal pagamento dei pagamenti - sezione richiesta;
- Rete patologica - sezione richiesta.

Include un quadro completo della storia medica di un paziente. A seconda della contingenza medica, il medico potrebbe non aver bisogno di consultare il documento completo, bensì avere bisogno solo ad un insieme limitato di dati. Con il modello architetturale proposto, è possibile estrarre informazioni strutturate come risorse FHIR da un documento in formato HL7 CDA 2.

Supponiamo di trovarci in una situazione reale in cui il medico deve prescrivere un farmaco. In questo caso, è realistico immaginare che lui sia interessato alle precedenti terapie farmacologiche e alle possibili allergie del paziente, al fine di valutare il potenziale rischio sanitario per il paziente.

In questo caso, il medico può richiedere informazioni collegate a Medication Statement e alla risorsa FHIR allergia e intolleranze. In particolare, il medico può facilmente ottenere unicamente le informazioni di interesse estraendole da un insieme di sezioni cliniche contenute nel Patient Summary.

Utilizzando il modello architetturale proposto, il professionista sanitario può, fornendo in input il documento CDA ed esprimendo una necessità informativa mediante una query, ottenere le informazioni di interesse contenute nel documento.

Grazie alla componente Interfaccia, l'operatore può in primo luogo indicare il farmaco che intende prescrivere e richiedere al sistema l'individuazione dello stesso all'interno della sezione allergia del patient summary.

La componente di mapping si occupa di individuare all'interno del documento l'informazione richiesta, la componente di estrazione si occuperà di estrarre le informazioni riguardanti i farmaci utilizzati e le allergie del soggetto.

Dopo l'estrazione delle informazioni di interesse le stesse sono passate alla componete di costruzione che definisce risorse FHIR con equivalente contenuto informativo rispetto alle informazioni estratte contenute nel documento.

Allo scopo di effettuare il mapping delle informazioni presenti nel CDA rispetto alle risorse FHIR in fase di definizione la componente di estrazione richiedere gli schemi di mappatura alla componente di Mapping.

La tabella 1 e la tabella 2 mostrano rispettivamente la corrispondenza tra la sezione dei farmaci del CDA e la risorsa Medication Statement FHIR e la sezione degli alert CDA e la risorsa FHIR Allergia e Intolleranza.

La componente di costruzione riceve le informazioni estratte e le mappa nelle corrispettive risorse FHIR di riferimento.

Infine, la componente Interfaccia mostra la risposta, presentando al professionista tutti i dati contenuti nelle risorse FHIR.

CDA	FHIR
ClinicalDocument/structuredBody/component/section/id	MedicationStatement/identify
ClinicalDocument/structuredBody/component/section/text	MedicationStatement/text
ClinicalDocument/structuredBody/component/section/entry /substanceAdministration/statusCode	MedicationStatement/status
ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/effectiveTime	MedicationStatement/effectiveDateTime
ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/effectiveTime	MedicationStatement/dosage/timing
ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/routeCode	MedicationStatement/dosage/route
ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/doseQuantity	MedicationStatement/dosage/quantityQuantity
ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/approachSiteCode	MedicationStatement/dosage/siteCodeableConcept
ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/rateQuantity	MedicationStatement/dosage/rateRatio

ClinicalDocument/structuredBody/component/section /entry/substanceAdministration/consumable /manufacturedProduct/manufacturedMaterial/code	MedicationStatement/medicationCodeableConcept

Tabella 1 – Sezione Medication mappatura CDA FHIR

CDA	FHIR
ClinicalDocument/structuredBody/component/section/id	AllergyIntolerance/identifier
ClinicalDocument/structuredBody/component/section/text	AllergyIntolerance/text
ClinicalDocument/structuredBody/component/ section/entry/act/statusCode	AllergyIntolerance/status
ClinicalDocument/structuredBody/component/ section/entry/act/effectiveTime	AllergyIntolerance/recordedDate
ClinicalDocument/structuredBody/ component/section/entry/act/entryRelationship/ observation/code	AllergyIntolerance/type
ClinicalDocument/structuredBody/component/section /entry/act/entryRelationship/observation/effectiveTime	AllergyIntolerance/onset
ClinicalDocument/structuredBody/component/section/entry/act/entryRelationship /observation/participant/participantRole/playingEntity/code	AllergyIntolerance/substance
ClinicalDocument/structuredBody/component/section/entry/act /entryRelationship/observation/entryRelationship/observation/ effectiveTime	AllergyIntolerance/reaction/onset
ClinicalDocument/structuredBody/component/section/entry/act /entryRelationship/observation/entryRelationship/observation/ value	AllergyIntolerance/reaction/manifestation
ClinicalDocument/structuredBody/component/section/entry /act/entryRelationship/observation/entryRelationship/observation/value	AllergyIntolerance/reaction/certainty
ClinicalDocument/structuredBody/component/section/entry /act/entryRelationship/observation/entryRelationship/act/text	AllergyIntolerance/note

Tabella 2 – Sezione Alert mappatura CDA FHIR

4. Conclusioni

Il modello architetturale proposto consente di estrarre risorse FHIR a partire da un documento semi strutturato o costituito da una serie di concetti e relazioni individuate mediante la fase di analisi NLP. Il modello prevede quattro componenti: i) una interfaccia, per la selezione della tipologia di informazione relativa ad una risorsa FHIR da recuperare partendo da un documento selezionato; ii) una componente di estrazione, per identificare la sezione di interesse; iii) schemi di mapping che consentono di trasformare un elemento di tipo concetto/ relazione in una risorsa di tipo FHIR; iv) una componente di costruzione atta a produrre una risorsa FHIR a partire da informazioni estratte da documenti clinici di interesse.

L'obiettivo principale è ottenere benefici legati all'utilizzo dello standard FHIR, che assicura attenzione all'implementazione, mediante specifiche chiare, risorse adattabili, e solide basi nell'utilizzo di standard per il web e il supporto ad architetture RESTful. In particolare, il grande vantaggio offerto dal modello riguarda la possibilità di estrarre informazioni granulari di interesse clinico da un documento clinico.

In particolare, il modello è stato applicato al documento clinico di tipo Patient Summary, un documento contenente dettagli sulla storia clinica completa del paziente. Questa scelta è dovuta al fatto che, una rapida estrazione delle informazioni mediche contenuto in questo documento è cruciale per la cura del paziente soprattutto in scenari di emergenza. In Italia, il Patient Summary è strutturato in conformità con lo standard

CDA. Il modello è basato sul mapping secondo le specifiche FHIR, delle risorse informative e dei concetti clinici presenti nelle diverse sezioni del Patient Summary.

Attraverso il modello architetturale proposto, le risorse FHIR possono essere facilmente estratte dalla struttura del CDA. Il nucleo del modello è costituito dalla componente di Mapping, in grado di mappare le trasformazioni di risorse informative da elementi CDA a risorse FHIR.

La soluzione proposta in definitiva consente l'estrazione di dati rilevanti da relazioni mediche narrative e documenti semi-strutturati, raggruppando automaticamente le informazioni cliniche più importanti relative a un paziente mediante l'utilizzo di dizionari e vocabolari.

5. Ringraziamenti

Il presente lavoro è stato parzialmente finanziato dal progetto "Big Data Architecture for Personal Health Record", presentato nell'ambito del Bando MISE grandi progetti R&S – PON". Si ringrazia la sig. Stefania Marra per il supporto amministrativo-gestionale fornito.

6. Riferimenti

[1] Kaplan, R. M. (1973). A general syntactic processor. In R. Rustin (Ed.), Natural Language Processing, pp. 193–241. Algorithmics Press, New York.

[2] Banko, M., M. Cafarella, S. Soderland, M. Broadhead, and O. Etzioni (2007). Open information extraction from the web. In Proceedings of the Joint Conference on Artificial Intelligence (IJCAI-2007), Seattle, WA.

[3] Fast Healthcare Interoperability Resources: https://www.hl7.org/fhir/

[4] Digital Imaging and COmmunications in Medicine: https://www.dicomstandard.org/

[5] eXtensible Markup Language: https://www.w3.org/XML/

[6] XML Schema Definition: https://www.w3.org/TR/xmlschema11-1/

[7] Unified Modeling Language: https://www.uml.org/

[8] Clinical Document Architecture: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=7

[9] Profilo Sanitario Sintetico: https://www.fascicolosanitario.gov.it/sites/default/files/public/media/Specifiche%20HL7%20CDA%20R2_%20Profilo%20Sanitario%20Sintetico-v1.2-S.pdf

[10] Grefenstette, G. and P. Tapanainen (1994). What is a word, What is a sentence? Problems of Tokenization. In The 3rd International Conference on Computational Lexicography (COMPLEX 1994), Budapest, Hungary.

[11] Unified Medical Language System: https://www.nlm.nih.gov/research/umls/

[12] Logical Observation Identifiers Names and Codes: https://loinc.org/

[13] eXtensible Stylesheet Language Transformations: https://www.w3.org/TR/xslt20/

[14] C. Diomaiuta, M. Sicuranza, M. Ciampi, e G. De Pietro, "A FHIR-based System for the Generation and Retrieval of Clinical Documents", in Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health - Volume 1: ICT4AWE, 135-142, 2017, Porto, Portugal

[15] Angelo Esposito, Mario Sicuranza, Mario Ciampi , "Progettazione e sviluppo di un servizio terminologico basato su FHIR per l'accesso univoco a risorse terminologiche sanitarie", 2017 disponibile al seguente link: https://intranet.icar.cnr.it/wp-content/uploads/2017/11/RT-ICAR-NA-2017-06.pdf

[16] M. Sicuranza, A. Esposito, M. Ciampi, "Semantic Information Retrieval from Patient Summaries", 11th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 2016. Lecture Notes on Data Engineering and Communications Technologies, vol 1. Springer, Cham

[17] M. Ciampi, G. De Pietro, C. Esposito, M. Sicuranza, P. Donzelli, “A federated interoperability architecture for health information systems”, International Journal of Internet Protocol Technology, vol. 7, no. 4, pp. 189-202, 2013

[18] M. T. Chiaravalloti, M. Ciampi, E. Pasceri, M. Sicuranza, G. De Pietro, R. Guarasci, “A model for realizing interoperable EHR systems in Italy”, in the proc. of the 15th International HL7 Interoperability Conference, pp. 13-22, 2015

[19] M. Sicuranza, M. Ciampi, “A semantic access control for easy management of the privacy for EHR systems”, in the proc. of the Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), IEEE, pp. 400-405, 2014