

TITOLO

Big4H - Big Data Analytics for e-Health Applications

Soggetti Attuatori

- SA Documents S.r.l
- CINI - Consorzio Interuniversitario Nazionale per l'Informatica

L'obiettivo del progetto **Big4H (Big Data for e-Healthapplications)** riguarda l'industrializzazione e il potenziamento tecnologico di un prototipo di ricerca **CLINSIC** per la gestione (estrazione, indicizzazione e ricerca) "semantica" di informazioni da cartelle cliniche, frutto di una precedente collaborazione tra l'Organismo di Ricerca CRdC Tecnologie Scarl e l'impresa SA Documents s.r.l. nell'ambito di un progetto di ricerca presentato ai sensi dell'art. 5 del D.M. 8 agosto 2000 n.593 ed approvato dal Ministero dell'Istruzione, dell'Università e della Ricerca con Decreto n.496/Ric. Del 07/05/2008.

In particolare, le funzionalità del prototipo saranno da un lato ingegnerizzate ricorrendo alle più moderne tecnologie informatiche e, dall'altro, estese con la finalità di "integrare" al suo interno altre tipologie informazioni provenienti dalle varie sorgenti di Big Data (es. dati amministrativi, dati sociali, dati del territorio, etc.). Sarà pertanto possibile "alimentare" una vera e propria *Base di Conoscenza* per le differenti strutture sanitarie (ospedali, cliniche, aziende sanitarie, etc.) in cui le informazioni, rappresentate sotto forma di *Linked Open Data*, potranno essere eventualmente "aperte" all'esterno (es. cittadino) e collegate anche ad ontologie, vocabolari e tassonomie (malattie, farmaci, etc.) di dominio utilizzate a livello internazionale.

Saranno poi messe a disposizione degli utenti finali avanzate funzioni di *Analytics* per il supporto alle decisioni (basate su algoritmi evoluti di *data mining*) per estrarre "valore" dai dati, tenendo conto della semantica associata a questi ultimi e supportando le più disparate analisi:

- monitoraggio dei ricoveri,
- correlazione tra patologie e caratteristiche del territorio,
- statistiche sull'uso di determinati farmaci per determinate malattie,
- feedback degli utenti su una struttura ospedaliera,
- suggerimenti del percorso clinico-sanitario ottimale,
- supporto alla diagnosi,
- etc.

utili per tutti gli enti interessati (es. strutture sanitarie, Ministero della Salute, Regioni).

Il prodotto così realizzato potrà sia costituire un sistema autonomo per la gestione di cartelle cliniche sia andare a complementare con funzionalità di *Big Data Analytics* gli attuali applicativi di gestione presenti sul mercato e già a disposizione di strutture sanitarie.

Il capofila aziendale, SA Documents, mette in campo la sua esperienza nella progettazione e realizzazione di sistemi di gestione documentale in ambito sanitario per fornire il necessario supporto alla industrializzazione del prototipo in tutte le sue parti.

L'ente di ricerca, CINI, di contro, metterà a disposizione le sue expertise nell'ambito di ricerca dei Big Data e del Web Semantico per progettare metodologie e tecniche che consentiranno, da un lato, l'integrazione delle sorgenti dati nella Base di Conoscenza e, dall'altro, la relativa gestione ed analisi attraverso meccanismi di indicizzazione, retrieval semantico e reasoning sui dati.

Il progetto si baserà sul modello americano dell'Electronic Health Report (**EHR**), una raccolta sistematica di informazioni sanitarie elettroniche su un singolo paziente o una popolazione. Si tratta di un record in formato digitale che è teoricamente in grado di essere condiviso tra i diversi contesti sanitari. In alcuni casi questa condivisione può avvenire per mezzo di sistemi informativi aziendali connessi in rete e di altre reti di informazione o scambi. Il sistema sarà progettato per rappresentare i

dati che riflettono accuratamente lo stato del paziente e le prestazioni effettuate dal personale sanitario e/o assistenziale in ogni momento.

Lo sviluppo e la modifica del modello EHR presenta due componenti determinanti di ricerca, il primo afferente alla sfera sanitaria e il secondo a quella informatica.

In termini generali il presupposto scientifico è la scomposizione del “dato” in “informazioni”. Nell’ambito della ricerca biostatistica, cui si ispira questo progetto di ricerca, il “dato” è rappresentato da una serie di “informazioni”, tra cui le cosiddette “variabili”, che lo definiscono nella sua qualità di “osservazione” data.

Nell’attività di cura e assistenza le informazioni che compongono un dato sono numerose e influenti nell’esito del processo. Vale a dire che ogni singolo dato, ad esempio una prima diagnosi, è a sua volta scomponibile nelle singole informazioni che la determinano, come ad esempio l’anamnesi, la visita, i reperti di laboratorio e di radiologia cui si aggiungono i singoli passaggi del paziente, dal suo ingresso in ospedale fino alla diagnosi. Altro elemento, ma questo difficilmente traducibile in informazione, è l’individualità del paziente e degli operatori che su di lui intervengono. In definitiva il dato viene scomposto nelle informazioni “possibili”. A latere di queste deve rientrare anche il costo sostenuto per quel determinato dato, anch’esso scomponibile in singole informazioni, come, ad esempio, il costo dei singoli operatori sanitari per categoria specifica, identificabili attraverso un codice di riconoscimento.

Una volta scomposti i numerosi dati che attengono alla gestione di un paziente e di una popolazione (più pazienti), questi possono essere aggregati nella forma classica della cartella clinica oppure le singole informazioni, raccolte in un unico database, possono essere riaggregate secondo modelli biostatistici atti a fornire una serie di informazioni sanitarie, epidemiologiche, economiche, comparative.

In altre parole da ogni singolo paziente i dati, oltre a mantenere le caratteristiche di una tradizionale cartella clinica, sono ridotti a informazioni che, adeguatamente aggregate ed elaborate, offrono sintesi aggiornate in tempo reale utili ad affrontare le problematiche organizzative, cliniche e assistenziali inerenti a strutture sanitarie

Di seguito, infine, si riportano le specifiche quantitative delle principali tecnologie - oggetto dell’industrializzazione - che il progetto Big4H propone in ogni fase del processo di analytics per le informazioni:

- Fase di Estrazione delle Informazioni o Sorgenti Informative di tipo Big Data e alle Risorse Linguistiche disponibili ==> copertura del dominio pari almeno al 50%
 - Meccanismi di accesso ai dati ➤ almeno 2 sorgenti per categoria di Big Data
 - Strumenti ETL ==> analisi delle principali soluzioni open-source
 - Sistemi di gestione documentale => Assessment e ingegnerizzazione del sistema CLINSIC
- Fase di Costruzione/Aggiornamento della Base di Conoscenza o Standard e formati per LOD ==> copertura del dominio pari al 100%
 - NOSQL/Triple Storage & Query Engine => analisi delle principali soluzioni open-source
 - Reasoning Engine => analisi delle principali soluzioni open-source
- Fase di Selezione e Analisi delle Informazioni o ROLAP/MOLAP engine => analisi delle principali soluzioni open-source
 - In Memory Database => analisi delle principali soluzioni open-source
 - Tool di Data Analytics => analisi delle principali soluzioni open-source

Durante il progetto di industrializzazione, si intende realizzare l’Azione Connessa D a supporto dell’innovazione e del trasferimento tecnologico a monte e a valle delle attività di R&S. Tale azione prevede:

- trasferimento verso potenziali Stakeholders dei risultati del progetto; coinvolgimento degli stessi su indicazioni dell’Impresa e/o su indicazioni dell’Organismo di Ricerca e su scouting del

fornitore, organizzazione e gestione di incontri periodici (cadenza trimestrale) per aggiornamenti sullo stato del progetto, azioni mirate di dimostrazione e utilizzo della soluzione Big4H, organizzazione di eventuale evento di divulgazione.

- -incontri di brokeraggio tecnologico, in Italia e in Europa (geografica) per presentare i risultati del progetto e per eventuali pre-accordi di commercializzazione, laddove i soggetti incontrati siano interessati alla soluzione Big4H.
- azioni di disseminazione di base: sito web, materiali di comunicazione.
- supporto alla gestione del progetto.