



OR5 - Modelli e metodologie per l'Extended Adaptive Enterprise Architecture

RI5.1 - STUDIO E ANALISI DI TECNICHE DI KNOWLEDGE EXTRACTION

RI5.1 si è posto come obiettivo quello di illustrare ed analizzare tecniche di Knowledge Extraction utili ad estrarre la conoscenza da sessioni di comunicazione/collaborazione realizzate con strumenti del Web 2.0 sia all'interno del contesto aziendale, sia in ottica Open Enterprise. La conoscenza estratta in questa fase può risultare di supporto alle decisioni relative alla riconfigurazione dei processi aziendali.

All'interno delle organizzazioni, esiste una grossa quantità di conoscenza tacita che risulta difficile da trasferire, che viene trasferita da un soggetto all'altro mediante la socializzazione, in accordo al Modello SECI (oggetto dell'attività 4.2). La socializzazione può avvenire o verbalmente o mediante l'utilizzo di strumenti del Social Web. Questi strumenti permettono la condivisione della conoscenza tra i dipendenti nell'azienda, promuovendo un approccio collaborativo e cooperativo.

E' quindi importante estrarre tale conoscenza, analizzarla e elaborarla per migliorare i processi aziendali.

Estrarre informazioni dal testo comporta la trasformazione di un documento semi-strutturato (o non strutturato) in un formato idoneo al *text mining*, che ha costituito un elemento fondamentale per RI5.1. Tale trasformazione consiste in una preelaborazione linguistica, avente fra le principali finalità, la riduzione delle parole caratterizzanti per il documento (*bag-of-words*).

Il *bag-of-words* iniziale è ottenuto mediante tokenizzazione che prevede, essenzialmente, la suddivisione del testo in un elenco di elementi detti *token* eliminando, ad esempio, spazi vuoti e simboli che non sono considerati rilevanti.

Ai token è applicata la rimozione delle stop word ossia l'identificazione ed eliminazione di quelle parole troppo generiche per poter fornire informazioni utili. Tali parole includono:

- elementi non lessicali precedenti un nome, quali gli articoli (determinativi e indeterminativi), gli aggettivi dimostrativi, possessivi e indefiniti;
- congiunzioni;
- preposizioni;
- parole appartenenti ad un eventuale stop-list, ossia un elenco di parole definite dall'utente che possono essere rimosse in quanto poco significative per il contesto in esame.

Spesso è prevista anche un'attività di normalizzazione che ha lo scopo di eliminare quelle differenze poco significative tra i *token*. Ad esempio, può essere opportuno trasformare tutti i caratteri in minuscolo.

Dal momento che spesso, all'interno dei documenti, le stesse parole compaiono sotto



diverse varianti morfologiche, si applicano processi di *stemming* e lemmatizzazione per ricondurle alla loro forma base. Lo *stemming* è un processo euristico atto a ricondurre parole flesse o derivate alla loro forma base, generalmente realizzato mediante la rimozione dei suffissi. La lemmatizzazione invece è un processo algoritmico per la determinazione del lemma (o forma canonica) di una parola.

Un'attività di fondamentale importanza per i processi successivi nell'estrazione della conoscenza da testo consiste nel *part-of-speech tagging* (PoSTagging). Essa associa ai *token* estratti la relativa parte del discorso (sostantivo, verbo, ecc.)

Una ulteriore attività di *tagging* potrebbe essere finalizzata alla disambiguazione che prevede l'assegnazione automatica del significato più appropriato di un termine polisemico, ossia di un termine che presenta più significati.

RI5.1 ha affrontato anche il problema della *Sentiment Analysis*, nota in letteratura anche con il nome di *Opinion Mining*: una disciplina che studia le opinioni, i sentimenti, le valutazioni e le emozioni delle persone relativamente a prodotti, servizi, aziende, individui, ecc., con l'ausilio di processi di analisi del testo e linguistica computazionale.

Il problema affrontato dalla *Sentiment Analysis* consiste sostanzialmente nel riconoscimento della eventuale positività o negatività di una specifica opinione. È, dunque, fondamentale comprendere come si struttura un'opinione. I sentimenti positivi, negativi o neutrali sono detti *sentiment* (o *opinion*) *orientation* (o *polarity*).

Pertanto in tale attività di ricerca sono state studiate e analizzate le suddette tecniche (statistiche e semantiche) di analisi ed estrazione della conoscenza. In riferimento alla trattazione dei principali tool utilizzati per l'estrazione della conoscenza sono stati descritti in maniera sintetica ma esaustiva:

- LanguageTool;
- Apache OpenNLP;
- Apache Tika;
- Mallet;
- Weka;
- MLPACK;
- Mahout;
- NLKT;
- Scikit-learn.

Tali *tools* sono stati inclusi nella trattazione a seguito di una preliminare analisi delle features implementate e messe a disposizione. Inoltre, sono state considerate anche caratteristiche di popolarità, linguaggio di programmazione di implementazione delle API, e disponibilità del codice sorgente (open source) per eventuali estensioni, integrazioni.

RI5.2- DEFINIZIONE DI UNA METODOLOGIA A SUPPORTO DELLA RICONFIGURABILITÀ ADATTIVA DEI PROCESSI



L'attività RI5.2 ha avuto l'obiettivo di definire una metodologia originale per la riconfigurazione dei processi a livello utente.

L'attenzione si è focalizzata sui cosiddetti processi adattivi, cioè quella categoria di processi aziendali caratterizzata dalla mancanza di predicibilità circa la propria evoluzione. Infatti, sono sempre più frequenti situazioni in cui l'evoluzione di un dato processo è determinata 'al momento' in base allo stato attuale dello stesso; tutto questo perché lo stato di un processo è caratterizzato da un numero indefinito di variabili, il cui valore, non può essere previsto a priori. Sulla base delle precedenti considerazioni, l'utente che segue il processo acquisisce un ruolo fondamentale: è lui, infatti, che deve stabilire come il processo deve eseguire, facendo leva sulla sua conoscenza ed esperienza. Quindi, diviene sempre più importante, in un contesto aziendale, condividere le informazioni o comunque individuare la figura in grado di fornire le informazioni necessarie. Anche le figure esterne all'azienda possono fornire un contributo alla evoluzione dei processi o comunque necessitano di prendere visione di tali processi (stakeholder).

Riassumendo, è possibile sintetizzare gli obiettivi in tre punti principali:

- coinvolgere gli stakeholders affinché prendano visione dell'evoluzione dei processi di interesse e/o forniscano un supporto agli stessi;
- favorire la comunicazione tra le varie funzioni aziendali al fine di diffondere la conoscenza;
- Fornire supporto a situazioni di imprevedibilità di un processo.

Il punto di partenza per la definizione della metodologia è stato ACM (Adaptive Case Management), che fornisce l'infrastruttura per gestire i processi basati sulla conoscenza, difficilmente gestiti dai sistemi attuali, in quanto troppo dinamici, variabili e non strutturati.

Così com'è presentato, ACM si presta a sostegno dei processi non deterministici. Tuttavia, durante l'attività, sono stati dettagliati gli approcci che garantiscono le contromisure da intraprendere ogniquale volta si presenta una nuova situazione: cioè la domanda alla quale si è voluto dare risposta è "a quali strumenti/tecniche/approcci è possibile ricorrere per risolvere un 'case'?"

Di seguito sono riportate le possibili risposte che sono state opportunamente approfondite:

Sistemi di raccomandazione: le informazioni necessarie alla risoluzione di un determinato case possono essere ricercate dall'utente, che esprime in modo dichiarativo al sistema la propria esigenza di ricerca, o essere proposte dal sistema all'utente in modo proattivo ed in questo secondo caso si parla di Recommender System (RS). L'elemento fondamentale su cui si basa la metodologia di adattività di processo, consiste proprio nella definizione di un sistema di raccomandazione (Recommender System RS), illustrato in Figura 1.

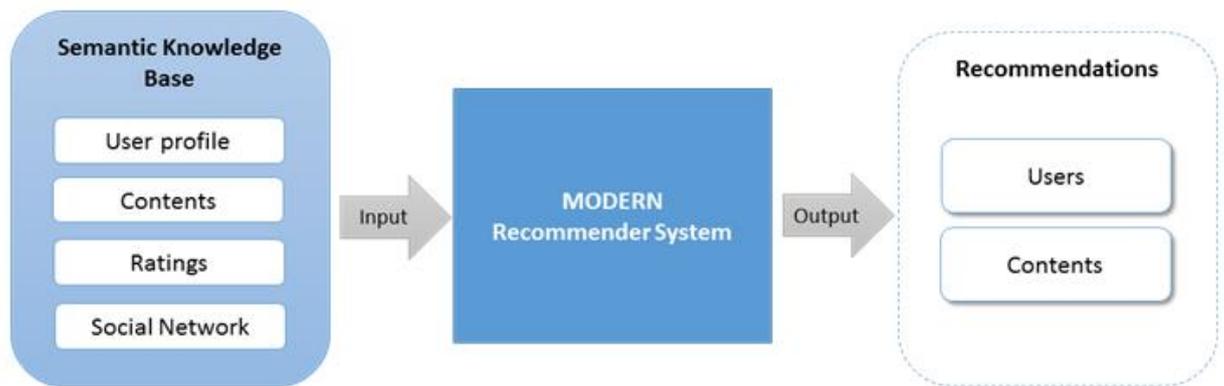


Figura 1: Input/Output del MODERN RS.

Diffusione della conoscenza e identificazione dei knowledge-worker: esistono situazioni (case), che sono state affrontate in passato da altri utenti dell'azienda. In questo caso la soluzione del problema sta nel diffondere le informazioni o consentire una facile individuazione di chi le detiene; in questo modo la soluzione del case avviene in tempi rapidi. Per perseguire tale obiettivo bisogna trovare il modo di facilitare le comunicazioni tra gli utenti di un'azienda.

I principi guida da prendere in considerazione sono quelli dettati da BPM 2.0 (o social BPM): l'integrazione del software sociale e di BPM può aiutare le organizzazioni a sfruttare il valore delle relazioni informali, senza compromettere le pratiche commerciali consolidate. Social BPM fonde BPM con i social software, con l'obiettivo di incrementare le performance potenziando la partecipazione degli utenti aziendali al fine di massimizzare la diffusione della conoscenza. Inoltre, come già accennato in precedenza, esistono situazioni in cui utenti esterni all'azienda possono voler partecipare ad un processo al fine di contribuire alla sua evoluzione o semplicemente per prenderne visione. Anche in questo caso social BPM consente una partecipazione controllata degli stakeholder.

La metodologia messa a punto in RI5.2 sfrutta, quindi, il connubio tra gestione dei casi, Recommender System e Web 2.0 (Figura 2). Il processo, per sua natura, è dinamico e non deterministico, in quanto il suo sviluppo dipende dal caso che si verifica e dalla decisione presa dall'utente che si trova a doverlo affrontare. Il workflow, pertanto, si adatta e cambia in funzione della decisione presa. A supporto dell'utente nella fase di decisione intervengono i meccanismi di raccomandazione e gli strumenti per la diffusione e la strutturazione della conoscenza, che supportano tali scelte, velocizzando ed ottimizzando l'esecuzione del processo stesso.

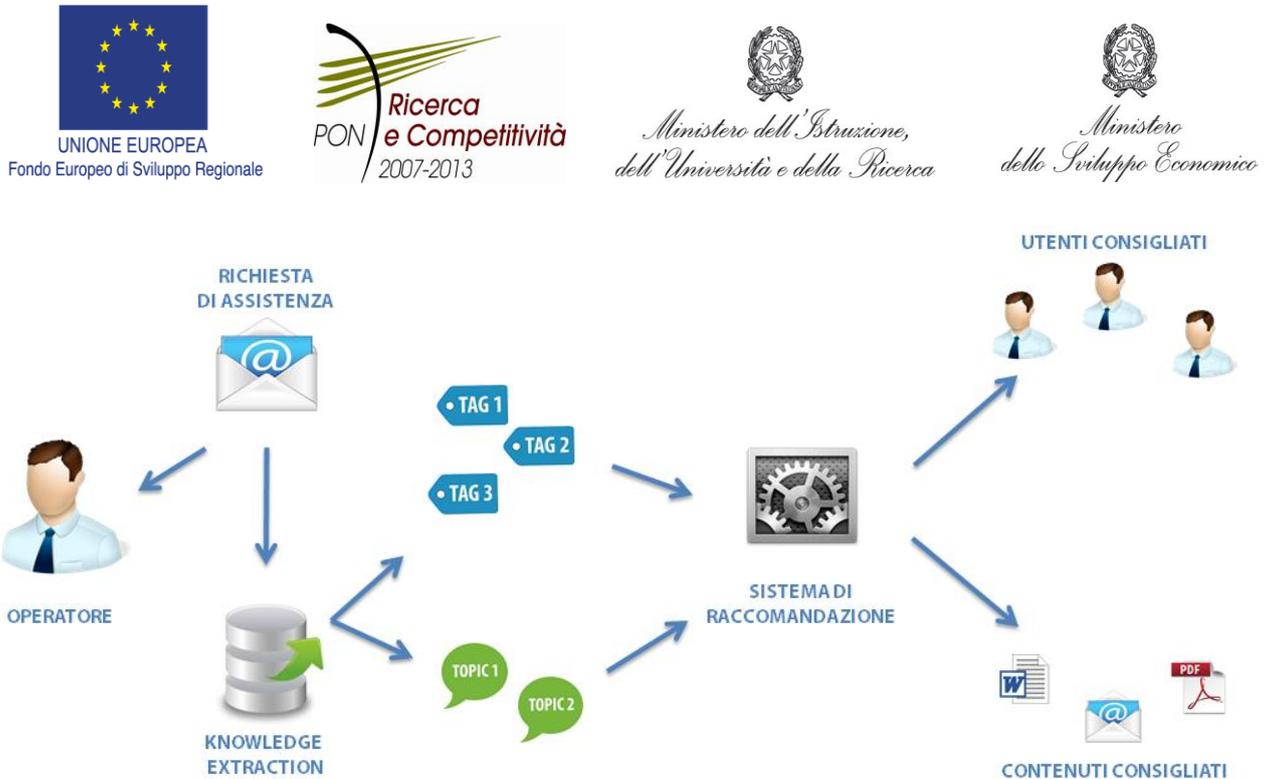


Figura 2: Adattività del processo per mezzo di Web2.0 e Recommender System.

Al fine di rendere la metodologia più chiara, sono stati definiti diversi scenari in cui viene mostrata la validità dell'approccio applicandolo in reali contesti aziendali (Polo Tecnologico di Poste Italiane).

RI5.3 - DEFINIZIONE DELL'ARCHITETTURA GENERALE

L'attività RI5.3 si è occupata di raccogliere ed armonizzare i pattern, le tecnologie e gli strumenti ottenuti come risultato dagli OR1-4 e dalle attività RI5.1 e RI5.2.

In particolare, la base di partenza è stata la metodologia dell'EAEA MODERN definita nell'attività RI1.4, che RI5.3 ha esteso e finalizzato, integrando i risultati provenienti dalle attività di ricerca condotte in OR4 ed OR5.

Lo scopo delle attività associate a RI1.4, infatti, è stato quello di definire un modello di Enterprise Architecture di nuova generazione che attraverso:

- una metodologia di analisi dei processi e dei sistemi informativi aziendali ed
- un modello di intervento sugli stessi processi e sistemi informativi

permetta all'organizzazione di poter calare al suo interno strumenti (metodologici e tecnologici) che si integrano sia con i Sistemi Informativi aziendali che con servizi ed applicazioni esterne attraverso un ecosistema di servizi con caratteristiche di flessibilità e dinamicità, predisposto all'orchestrazione ed alla riconfigurabilità adattiva dei processi informativi, su cui costruire l'esperienza dell'utente sulla base dei paradigmi e dei modelli del Web 2.0.

L'Enterprise Architecture di nuova generazione che si è prodotta organizza, in una visione sistemica, tutti i contributi provenienti dai altri task di OR1 ed è arricchita recependo i



risultati di OR2-OR4. L'EA sviluppata è stata utilizzata in OR5 per definire il Modello di Architettura Software e per realizzare un'istanza di quest'ultimo. Il risultato è la definizione di un prototipo dell'EA stessa, istanziata e sperimentata in OR6.

I risultati ottenuti dall'attività RI5.3, nel dettaglio, possono essere suddivisi tra: il raffinamento e l'estensione della Extended Adaptive Enterprise Architecture (EAEA), e la definizione dell'architettura logica del prototipo software realizzato in SS5.4. La costruzione dell'Enterprise Architecture, definita in MODERN, prevede una modellazione in 3 fasi (Modello di analisi, Modello di intervento e Architettura logica di riferimento) che si concretizzano poi in 2 diversi artefatti: la metodologia di MODERN e l'architettura SAFE4MODERN. Pertanto, i contributi, che RI5.3 ha provveduto ad integrare nella metodologia, sono stati differenziati in funzione del fatto che essi supportino la metodologia o l'architettura di MODERN.

In aggiunta a questi contributi, all'interno del task RI5.3, è stata estesa l'EAEA relativamente agli stream di MIKE (Strategy, Process, Organisation, People), mediante l'integrazione dei pattern dell'Enterprise 2.0 partendo dal modello socio-tecnico. Un pattern consiste in una descrizione strutturata di una soluzione ad un problema ricorrente o ad una classe di problemi, in un determinato contesto e, nel caso di MODERN, sono finalizzati a guidare l'adozione del paradigma Enterprise 2.0 e degli strumenti del Web Semantico in azienda ed a supportare le scelte architeturali legate all'introduzione di tali aspetti all'interno dell'organizzazione. In particolare, sono state identificate 5 aree di interesse per le quali sono stati definiti opportuni pattern:

1. Definizione, orchestrazione e delivery dei servizi: integrazione di sistemi legacy, basi di dati, applicazioni esterne, anche di tipo Web 2.0, seguendo i principi di un'architettura service-oriented; orchestrazione dei servizi per abilitare la riconfigurazione agile dei processi; modelli di delivery dei servizi che possano garantire scalabilità, flessibilità, riduzione dei costi, interoperabilità;
2. Accesso e composizione dei servizi: accesso alle informazioni presenti all'interno dei sistemi legacy e provenienti da flussi informativi esterni;
3. Rappresentazione e organizzazione della conoscenza: gestione della conoscenza collettiva mediante integrazione delle informazioni frammentate, provenienti sia dall'interno dell'organizzazione sia dai confini esterni e supporto alla cooperazione di servizi e applicazioni mediante l'uso di schemi e tecnologie semantiche;
4. Condivisione della conoscenza: modelli di interazione sia tra attori interni che con attori esterni all'azienda, al fine di arricchire la conoscenza collettiva;
5. Riconfigurazione dei processi user-centric: estrazione della conoscenza dai processi di comunicazione e collaborazione a supporto delle decisioni relative alla riconfigurazione dei processi.

Per ogni area sono stati individuati uno o più macro-pattern. Questi pattern di alto livello presentano una descrizione ad ampio raggio del contesto applicativo e della classe di



problemi oggetto di interesse, nonché dell'approccio risolutivo proposto. Per ogni macro-pattern è presentata una serie di pattern di dettaglio, di granularità più fine, mutuati dai diversi paradigmi, pattern e linee guida analizzati durante la fase di revisione della letteratura o comunque individuati a valle di un'analisi critica della stessa e di una rielaborazione in ottica MODERN.

Nel delineare un modello di intervento per l'applicazione della vision MODERN in azienda, l'analista avrà, quindi, a disposizione un catalogo di pattern relativi all'Extended Adaptive Enterprise Architecture, che guidano l'adozione del paradigma Enterprise 2.0 e degli strumenti del Web Semantico in azienda. Essi rappresentano un valido supporto nell'effettuare le dovute scelte architettoniche legate all'introduzione di tali aspetti all'interno dell'organizzazione e sono opportunamente raccolti ed organizzati in modo da facilitarne la fruizione.

In sostanza i Pattern EAEA sono definiti con lo scopo di fornire delle best-practice e linee guida che, in relazione agli ambiti su citati ed in risposta quindi alle rispettive esigenze, rappresentano degli strumenti ausiliari per l'introduzione del modello architettonico MODERN in un'azienda che non adotta il paradigma Enterprise 2.0 e i principi della Semantic Enterprise, oppure che adotta, in parte o in toto, i principi dell'Enterprise 2.0, ma i cui flussi informativi non sono governati.

Per quanto riguarda la definizione dell'architettura logica del prototipo software, è stata disegnata un'infrastruttura software strutturata fondamentalmente su tre layer (Figura 3):

- Presentation Layer
- Orchestration & Integration Layer
- Data Layer

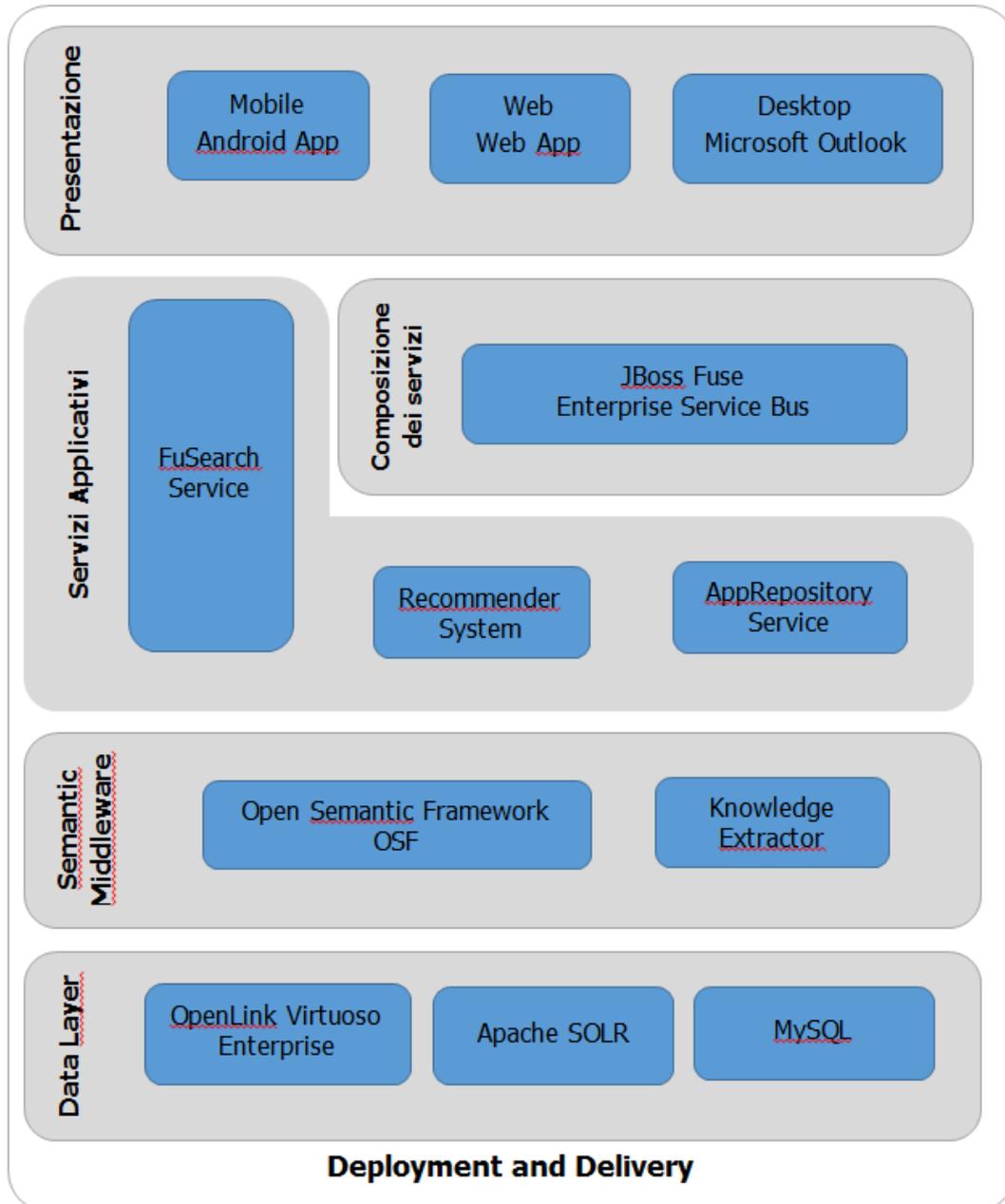


Figura 3: Modello EAEA - Architettura prototipale di dettaglio.

Il Presentation Layer implementa la metafora del PE attraverso tre tipi di APP: MobileAndroid App, WebApp, Desktop App. La MobileAndorid App consente di accedere ai contenuti MODERN in completa mobilità ed ha un'architettura di tipo client-server. La WebApp è un search engine che utilizza tecnologie innovative per fornire risultati di ricerca più accurati e chiari ed anch'essa presenta un'architettura client-server. La Desktop App è plugin di Outlook che permette di rendere sociali le discussioni via email. L'Orchestration & Integration layer è stato ulteriormente suddiviso in tre sotto-layer: layer



di composizione dei servizi, layer dei servizi applicativi e layer del Semantic Middleware. Nel primo è presente l'ESB, che fa da connettore tra il Presentation Layer e il Back-End, nel secondo sono disponibili i servizi associati alle capability di raccomandazione, ricerca e di app repository e nel terzo risiedono il framework l'OSF ed il modulo di KE.

Il Data Layer è composto dalle componenti adibite alla memorizzazione dei dati usati da MODERN, sia essi semantici (attraverso un RDF Store) che non.

Per ogni Layer, precedentemente descritto, sono stati realizzati i rispettivi *Diagram* necessari alla fase implementativa del prototipo prevista nell'attività SS5.4. In particolare sono stati prodotti:

- Deployment Diagram, per la descrizione delle risorse necessarie all'esecuzione del prototipo
- Activity Diagram, per rappresentare le diverse dinamiche degli Use Case
- Class Diagram, per la descrizione delle entità
- Sequence Diagram
- Workflow dei servizi Rest

SS5.4 - IMPLEMENTAZIONE PROTOTIPALE DI ALCUNI ASPETTI ARCHITETTURALI

Obiettivo dell'attività di progetto è stato l'implementazione degli aspetti architetturali individuati nell'attività di ricerca RI5.3 prestando particolare attenzione a tutti gli aspetti ritenuti fondamentali per il raggiungimento degli obiettivi fissati dal progetto MODERN. Il modus operandi adottato è stato di tipo modulare e fortemente incentrato sulle features collegate nella metafora applicativa adottata nel prototipo oggetto di sperimentazione.

Il processo di sviluppo ha riguardato tutti i layer logici/funzionali individuati nell'architettura delineata in RI5.3. In particolare, nell'ambito del **Presentazione layer** sono state realizzate:

- la web application *FuSearch*, di fatto, search engine della piattaforma MODERN e in grado di proporre due tipologie di ricerca:
 - *Semantica*, che trova nelle relazioni ontologiche definite nel modello di conoscenza la base su cui attingere le informazioni ricercate dagli utenti;
 - *Full-Text*, ricerca tradizionale basata su testo libero.
- la web application *App Repository* responsabile della messa in essere del concetto market place progettuale.
- il plug-in *Social Mail*, che consente di corredare le mail con informazioni aggiuntive in modo da favorire la collaborazione e la condivisione della conoscenza mediante processi sociali di emersione del valore. Il modulo software implementa le funzionalità di base che trasformano la mail in un social object. In particolare, attraverso il plugin, è possibile visualizzare il thread di discussione in una forma più fruibile e simile ad un blog post; è possibile commentare gli interventi, aggiungere tag, votare con like e ricercare le discussioni mediante ricerca full-text.



- il plug-in *Social Network*, che permette di creare e quindi sfruttare a pieno le potenzialità di una rete di contatti (Social Network aziendale) allo scopo di favorire l'emersione di quella tipologia di capitale intellettuale che solitamente viene veicolato attraverso canali comunicativi propri dei rapporti interpersonali. Il modello di SN proposto, di tipo sincrono, consente all'utente di attingere, ma soprattutto iniettare, informazioni utili alle attività lavorative aziendali. Grazie all'utilizzo delle funzionalità implementate, l'utente può decidere di "seguire" delle persone, attivando in questo modo l'accesso alle informazioni pubblicate dalla persona seguita (vero e proprio bridge comunicativo tra i due stakeholders). Attraverso il plugin, inoltre, l'utente può indicare sia competenze professionali e sia interessi personali che aggiunte alle informazioni di natura anagrafico/societario dell'utente vanno a costituire un'importante entità informativa: il profilo social. Il plugin, realizzato, nel rispetto delle direttive emerse nelle attività di ricerca, consente alla piattaforma non solo di fornire validi strumenti di social innovation ma anche di storicizzare dati e/o informazioni utilizzabili in attività di monitoraggio/valutazione della validità degli strumenti proposti.

Il blocco architetturale **Composizione dei Servizi** è stato implementato attraverso l'utilizzo della piattaforma d'integrazione *JBoss Fuse* come environment di esecuzione di una serie di servizi RESTFull, implementanti la logica funzionale della platea applicativa proposta. Il processo di sviluppo ha previsto l'applicazione dei principali enterprise pattern studiati nel progetto.

Nel blocco architetturale **Servizi Applicativi** sono stati collocati i servizi applicativi inerenti feature specifiche presenti nella piattaforma prototipale, es: la ricerca Semantica / Fulltext, servizi di raccomandazione, etc.

Il sistema di raccomandazione permette di effettuare suggerimenti al fine di supportare l'adattività dei processi. Esso si presenta come un modulo web oriented che espone i propri servizi via REST e offre le seguenti raccomandazioni:

- suggerimento utente esperto di un certo topic;
- suggerimento contenuto in base ai tag presenti in un altro contenuto;
- suggerimento contenuto in base a similarità con un altro contenuto;
- suggerimento tag in base ai tag presenti in un certo contenuto.

La funzionalità di *Expert Finding* (suggerimento utente esperto) usufruisce del motore di KNA, opportunamente predisposto per lo scenario di sperimentazione (sono stati implementati una serie di connettori ad hoc tra il motore KNA il Data Layer del prototipo). Le funzionalità di suggerimenti contenuto e tag, invece, sfruttano i risultati prodotti dal motore di KE, anche esso proposto nel forma di web module e dotato di servizi REST per ogni tipologia di suggerimento.



Lo strato **Semantic Middleware** è stato realizzato attraverso l'utilizzo di due moduli e/o framework applicativi:

- *OSF – Open Semantic Framework*, di fatto responsabile della gestione semantica della knowledge aziendale. L'attività d'integrazione ha previsto una fase di customizzazione di alcuni dei servizi standard proposti dalla piattaforma;
- *Knowledge Extractor*, modulo software che consente l'emersione della così detta hidden knowledge presente nelle varie sorgenti informative quali archivi documentali. È basato sull'impiego di linguaggi d'interrogazione semantici (e.g. SPARQL) e rule engine per inferire conoscenza.

Il blocco architetturale **Data Layer**, infine, responsabile della storicizzazione della knowledge aziendale intesa sia come istanze del modello ontologico individuato e sia come repository centralizzato, utilizza:

- Openlink Virtuoso, triple store per la memorizzazione dell'istanza del modello ontologico;
- Apache Solr, utilizzato nell'ambito della ricerca fulltext proposta nella piattaforma;
- MySQL RDBMS, utilizzato per la storicizzazione di tutte le informazioni necessarie per la fase di reportistica e statistica previste dal progetto.