Indice (parte comune)

3 Servizi REST applicati ai casi d’uso

3.1 Introduzione alle architetture RESTful

[…]

Segue parte relativa ai casi d’uso di ciascuno.

1. Sistemi basati su conoscenza non strutturata

La maggior parte delle informazioni attualmente disponibili in formato digitale, siano esse di pubblico dominio o appartenenti a un'organizzazione, sono organizzate in forma non strutturata, ossia in testi o frammenti di testo. L'informazione e la conoscenza contenute in queste fonti sono semanticamente e strutturalmente diverse dall'informazione codificata in forma strutturata e dalla conoscenza presente in una base di conoscenza. Il testo rappresenta la realtà in modo diverso rispetto alle tuple di un database o alle regole in una base di conoscenza, e va dunque processato in modo opportuno affinché il suo contenuto informativo possa essere acquisito e utilizzato automaticamente e non soltanto tramite la lettura. Esistono due livelli di rappresentazione digitale del testo: il più superficiale è basato solo sulla *forma* del testo e non sul suo *significato* (ossia, ad esempio, parole con Io stesso significato sono considerate diverse perché hanno una forma diversa); al contrario, la rappresentazione più profonda dei testi tiene conto del significato delle parole. Le possibilità aperte dal poter operare automaticamente sui testi sono significative: diventa possibile gestire grandi collezioni di testi, contenenti anche decine di milioni di parole, cercando al loro interno tramite parole chiave o analizzando una rappresentazione sintetica del loro contenuto, se questo è rappresentanto preservandone la semantica.

1.1 Elaborazione del linguaggio naturale

I passi compiuti per preparare il testo ad essere rappresentato in forme appropriate - ad esempio - per le operazioni di ritrovamento sono integrati in un processo detto di "elaborazione del linguaggio naturale" o Pipeline di Natural Language Processing (NLP). Esso comprende un

insieme di passi che hanno il fine di aumentare l'efficacia delle tecniche di rappresentazione e analisi del testo, tanto eliminando le ridondanze tipiche di questo genere di rappresentazione, quanto mettendone in risalto la peculiare ricchezza informativa. Le operazioni eseguite in questo processo sono alla base sia della rappresentazione "superficiale" del testo, sia di quella che tiene conto del significato dei termini.

1.1.1 Sentence Detection

Il primo passo nel processo di elaborazione del testo è l'individuazione delle frasi. Ciò viene fatto generalmente considerando che queste sono solitamente separate dal punto fermo, tuttavia è necessario distinguere questo da altri usi del punto che possono essere presenti nei testi considerati (ad esempio, per abbreviare termini, per separare porzioni di un numero, eccetera).

1.1.2 Tokenization

Il secondo passo nella pipeline di NLP è la suddivisione delle frasi in parole o "token", eliminando i segni di punteggiatura e i caratteri speciali. Anche in questo caso, l'euristica base con la quale si identificano i token è che essi sono generalmente separati da spazi; bisogna tuttavia tener conto di token in forme particolari, come le date e l'ora, i nomi propri di persona, eccetera.

1.1.3 Stop Word Removal

Le stopword sono parole molto comuni in una lingua o in un ambito, al punto tale che il loro contenuto informativo è considerato molto basso o nullo; per questa ragione un testo si può rappresentare efficacemente senza comprenderle risparmiando spazio di rappresentazione e tempo di computazione. Per le lingue più comuni esistono liste di stopwords

generali e specifiche per alcuni domini, che vengono usate per filtrare le parole rilevanti nei documenti.

1.1.4 POS Tagging

Il Part Of Speech Tagging è l'operazione che associa a ciascun token la sua categoria grammaticale all'interno della frase. Questo permette di ottenere dei risultati più precisi e compiere ricerche più dettagliate all'interno di una collezione, in particolar modo quando i testi sono rappresentati codificando la semantica delle parole.

1.1.5 Lemmatization & Stemming

Nei documenti sono spesso presenti diverse forme di una parola che hanno dei significati simili (democrazia, democratico, democratizzazione). La lemmatizzazione è un'operazione di semplificazione dei token che trasforma ciascuno di essi nella sua forma grammaticale di base (ossia il suo lemma) trasformando il suffisso proprio della specifica inflessione presente nel testo in quello della forma base dell'elemento grammaticale. Il processo di trasformazione non è banale in quanto va identificata correttamente la forma base di ogni token; per fare ciò, gli algoritmi di lemmatizzazione possono ricondurre i termini che hanno la stessa forma ma funzioni grammaticali differenti (es. un'àncora, egli àncora, ancòra qui) grazie al contesto e al risultato del processo di POS tagging.

Lo stemming è un'operazione di semplificazione dei token più radicale: essa tronca il suffisso di ogni token riducendolo alla radice del termine. Uno degli algoritmi più usati per compiere questa operazione è l'algoritmo di Porter.

1.1.6 Keyphrase Detection

Questo passo di analisi del testo consiste nell'identificare concetti più rilevanti espressi con più di una parola all'interno del testo. Un approccio semplice, che impiega tecniche di apprendimento non supervisionato, si basa sull'identificazione delle coppie, triple o n-uple di termini che co-occorrono con una frequenza significativa.

1.2 Motori di ricerca classici

Uno degli strumenti più comuni e più utili che si può costruire su una collezione di documenti digitalizzati e trasformati con le operazioni descritte è il motore di ricerca. Questo è un sistema che ha l'obiettivo di restituire all'utente, data una query, il sottoinsieme di documenti della collezione che sono più rilevanti rispetto ad essa. Questo task cambia significativamente nel caso in cui i documenti sono rappresentati tenendo conto della semantica dei termini al loro interno; in questo caso, infatti, i documenti più rilevanti per una query saranno quelli che contengono parole dal significato più simile a quelle date in input, e non semplicemente quelli che ne contengono una certa percentuale, magari con una certa frequenza.

Sono considerati in questo paragrafo i motori di ricerca "classici", ossia basati solo sulla forma dei termini e non sul loro significato. Esistono diversi approcci per creare un motore di ricerca classico che differiscono principalmente per il modo in cui sono rappresentati i documenti al loro interno e come questi sono confrontati con le query dell'utente. Sono presentate di seguito le strutture dati e le assunzioni comuni a diversi tipi di motori di ricerca; viene poi brevemente presentato il modello Booleano e il più diffuso Modello a Spazio Vettoriale.

1.2.1 Indice invertito

L'operazione più basilare ed intuitiva che un motore di ricerca deve compiere per risolvere una query è individuare quali sono i documenti che contengono le parole presenti nella query. Per compiere questa operazione in modo efficiente i motori di ricerca operano su una struttura dati chiamata "indice invertito". Il suo nome deriva dal confronto con il normale indice di un libro, che mette in relazione i capitoli o i paragrafi di un testo con le pagine in cui si trovano, e quindi con le parole contenute in essi. L'indice invertito è una struttura dati che associa ad ogni parola i documenti nei quali essa è contenuta, consentendo così di individuare i sottoinsiemi di documenti in cui sono presenti tutte o molte delle parole in una query. L'insieme di tutti i termini presenti in tutti i documenti della collezione associati alla loro frequenza assoluta è chiamato vocabolario della collezione.

1.2.2 Modello Bag of Words

Il modello Bag of Words è un metodo di rappresentazione di un testo che tiene conto del numero di occorrenze di ogni parola nel testo ma non delle posizioni che occupa in esso; quindi, testi che contengono le stesse parole ma in diverso ordine, e quindi con significati diversi, saranno rappresentati nello stesso modo. Il motivo per cui si adotta questa rappresentazione è che la posizione delle parole in un testo non è considerata una informazione di grande rilevanza per il ritrovamento di documenti rilevanti rispetto a una query.

1.2.3 Boolean Model

Adottando la rappresentazione secondo il modello Bag of Words, un documento sarà dunque rappresentato come un insieme di parole. Il

modello booleano offre la declinazione più semplice di questa rappresentazione, ossia quella che considera la semplice presenza o assenza di un termine in un documento, senza considerarne la frequenza. Un documento è rappresentato, quindi, come un vettore nel quale ogni elemento rappresenta una parola del vocabolario e può assumere il valore *vero* o *falso*. Una query per un motore di ricerca costruito secondo questo modello può essere espressa collegando i termini con operatori booleani (es. *t1* AND *t2* OR *t3*) e la risposta alla query corrisponde al sottoinsieme di documenti della collezione la cui rappresentazione coincide con quella espressa in forma logica nella query. Ogni documento sarà dunque rilevante o non rilevante per una query e non vi sarà associata una misura di rilevanza; per questo i risultati di una ricerca non sono restituiti secondo un ordinamento significativo.

Pur essendo un modello che consente di esprimere interrogazioni con precisione, la forma logica delle query costituisce sia un ostacolo sintattico per l'utente, che deve saper esprimere il proprio bisogno informativo in questa forma, sia un vincolo spesso troppo – o troppo poco – stringente nel processo di individuazione dei risultati rilevanti, in quanto i documenti che corrispondono precisamente alla query possono essere molto pochi, quando la query è molto stringente, o troppi, se consiste di molte condizioni disgiunte.

1.2.4 Term Frequency e Inverse Document Frequency

Per arricchire la rappresentazione di un documento con il modello bag of words si può tenere conto di due parametri legati ad ogni termine: la frequenza del termine in ciascun documento e la "rarità" di un termine nell'intera collezione. L'idea alla base dell'introduzione di queste

metriche è che un documento sarà più rilevante rispetto ad una query se uno o più termini nella query sono molto frequenti nel documento, oppure se uno o più termini che si trovano nella query sono presenti solo in quel documento o in pochi altri. La frequenza di una parola in un documento è detta Term Frequency; matematicamente è definita come la frequenza assoluta del termine nel documento. Generalmente, questa misura non viene direttamente impiegata per calcolare la rilevanza di un documento rispetto a una query, poiché costituirebbe un fattore moltiplicativo troppo forte (un documento in cui un termine appaia 10 volte sarebbe 10 volte più significativo di uno in cui esso appare una volta sola). Per questa ragione, si impiega il logaritmo per ridurre l'impatto della crescita del Term Frequency sul prodotto; l'operazione applicata è detta *sublinear tf-scaling* ed è definita matematicamente come:

$$sub\_tfs(t,d) = \begin{cases} 1 + log_{10} tf_{t,d}, & tf_{t,d} > 0 \\ 0, & tf_{t,d} \leq 0 \end{cases}$$

La "rarità" di un termine in una collezione è definita matematicamente sulla base del concetto di Document Frequency di un termine, ossia il numero di documenti in cui il termine appare. Questa misura è, ovviamente, inversamente proporzionale alla rarità di un termine; la rarità viene quindi misurata dall'Inverse Document Frequency, che è definito matematicamente come:

$$idf_t = log(\frac{N}{df}),$$

dove *N* è il numero di documenti nella collezione e *df* è definito come il Document Frequency del termine *t*.

Ad ogni termine in un documento può essere associato, dunque, un peso che è direttamente proporzionale sia alla frequenza del termine nel

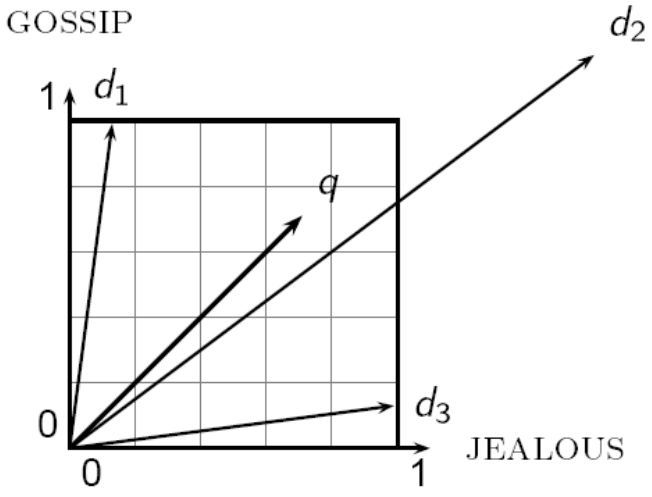
documento stesso, sia alla rarità del termine nella collezione; esso è il prodotto tra Term Frequency e Inverse Document Frequency del termine:

$$w_{t,d} = sub\_tfs(w,d) \cdot idf_t$$

1.2.5 Vector Space Model

Il modello a spazio vettoriale è un modello utilizzato nei sistemi di ritrovamento dell'informazione nel quale i documenti di una collezione e le query su di essa sono rappresentati come vettori. L'idea alla base del modello è di rappresentare query e documenti nello stesso spazio multidimensionale, nel quale ogni dimensione rappresenta un termine nel vocabolario della collezione. L'i-esimo elemento di un vettore che rappresenta un documento (o una query) ha come valore il peso *tf-idf* associato al termine corrispondente ali'i-esima dimensione dello spazio, calcolato rispetto al documento considerato.

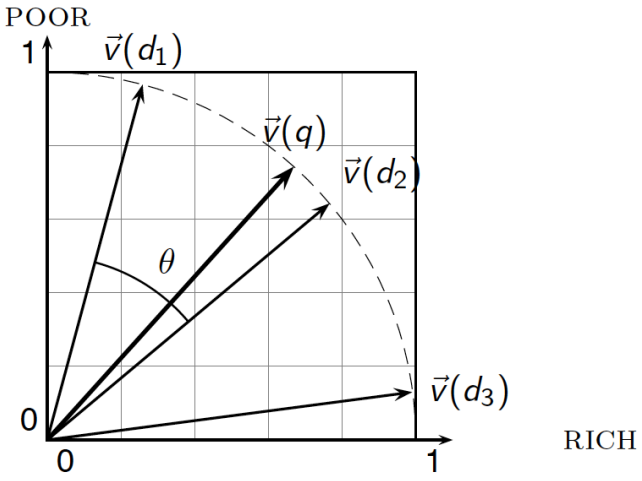
Utilizzando il VSM è possibile comparare documenti e query utilizzando operazioni vettoriali. Una prima possibilità per calcolare la distanza tra due documenti è considerare la loro distanza Euclidea nello spazio



Rappresentazione dei documenti come vettori di lunghezza non normalizzata

multidimensionale; questa è definita come segue: $\sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}$, dove q_i e p_i sono gli elementi in posizione i-esima di generici vettori $\vec{q}$ e $\vec{p}$. Usando questa metrica, però, due documenti che contengono gli stessi termini ma hanno lunghezza diversa – poiché alcuni termini sono ripetuti – saranno considerati diversi perché distanti nello spazio. È necessario, quindi, compiere una normalizzazione della dimensione dei vettori. Questa operazione è compresa nella misura di similarità del coseno, che valuta la distanza tra documenti sulla base dell'angolo compreso tra i vettori che li rappresentano. La metrica è definita come segue:

$$\cos(\vec{q}, \vec{d}) = \frac{\vec{q} \bullet \vec{d}}{|\vec{q}||\vec{d}|} = \frac{\vec{q}}{|\vec{q}|} \bullet \frac{\vec{d}}{|\vec{d}|}$$



Rappresentazione della distanza tra vettori normalizzati basata sull'angolo compreso tra essi

Il Vector Space Model presenta comunque alcune problematiche:

1. La posizione dei termini non viene presa in considerazione;
2. Documenti molto lunghi potrebbero avere risultati scarsi;
3. Il modello è limitato dalla sparsità dei dati.

Inoltre, il modello a spazio vettoriale non può rilevare la similarità tra due documenti che abbiano un contenuto semanticamente simile ma

contengano parole differenti, poiché la rappresentazione dei documenti è basata solo sulla forma delle parole che contengono e non sul loro significato.

1.3 Modelli Semantici Distribuzionali

La rappresentazione del testo che tiene conto del significato delle parole ha un impatto significativo sulle operazioni di ritrovamento di documenti rilevanti rispetto a una query. Avere a disposizione una rappresentazione del significato delle parole in usate in una collezione di documenti permette di calcolare la similarità tra una query e un documento non soltanto in base alle parole della query che occorono nel documento, ma anche considerando le parole in esso che hanno un significato simile a quelle date in input.

Per creare una rappresentazione del significato delle parole si utilizzano spesso modelli matematici che rappresentano le parole come vettori in uno spazio multidimensionale, nel quale le parole con un significato simile sono rappresentate tramite vettori "vicini" tra loro; lo spazio multidimensionale viene detto Word Space o Semantic Space. Il significato di ogni parola in questo tipo di modelli è fortemente correlato a quello dei termini che occorrono più frequentemente insieme ad essa. Infatti, la rappresentazione di una parola sottoforma di vettore nello spazio si basa sull'analisi del contesto d'uso di quella parola nella collezione considerata, poiché si assume che un gruppo di parole frequentemente co-occorrenti abbiano un significato simile perché appartnenenti allo stesso "campo semantico".

Esistono diversi metodi per creare lo spazio vettoriale per rappresentare le parole; di seguito è presentata la tecnica chiamata Random Indexing.

1.3.1 Random Indexing

Il Random Indexing è una delle tecniche di creazione di un Word Space su una collezione di documenti. Essa ha il vantaggio di generare una rappresentazione del significato delle parole senza utilizzare fonti di conoscenza esterne (come dizionari o ontologie) e in modo incrementale, ossia riuscendo ad aggiornare la rappresentazione in modo semplice all'aggiunta di nuovi documenti nella collezione; l'unica operazione di preprocessing del testo necessaria per costruire lo spazio è la tokenizzazione.

La tecnica consiste in due passi principali: il primo consiste nell'assegnare a ogni termine del vocabolario di una collezione un vettore ternario e sparso generato casualmente, ossia un vettore i cui elementi possono assumere i valori -1, 0 e 1, nel quale la maggior parte degli elementi hanno valore 0 e nel quale le posizioni degli elementi non nulli sono scelte in modo casuale. Il secondo passo consiste nel generare un "vettore contesto" per ogni termine nel vocabolario sommando i vettori precedentemente associati alle parole che occorrono frequentemente insieme ad esso nella collezione, in base ad una soglia di vicinanza prefissata. Il vettore contesto di un termine è calcolato secondo la formula seguente:

$$\overrightarrow{cv_i} = \sum_{d \in C} \sum_j \vec{r_j}, \ -c < j < c, i \neq j,$$

dove C è la collezione di documenti, c è la soglia di vicinanza dei termini prefissata che delimita il contesto di ogni parola e $\vec{r_j}$ è il vettore casuale assegnato ad ogni parola presente nel contesto.

La creazione di un vettore contesto per ogni parola corrisponde formalmente ad un'operazione di riduzione della dimensionalità della matrice di co-occorrenza dei termini in una collezione. Questa operazione genera uno spazio di dimensionalità ridotta nel quale le distanze tra i vettori che rappresentano le parole restano proporzionali alle distanze che le parole avevano nello spazio orignale. [Johnson-Lindenstrauss lemma]

1.3.2 Motori di ricerca semantici

Usando la tecnica di Random Indexing è possibile rappresentare termini e documenti nello stesso spazio vettoriale multidimensionale. Si può generare, infatti, una rappresentazione vettoriale di un documento tramite la somma pesata di tutti i vettori che rappresentano le parole che contiene, usando come pesi gli indici di Inverse Document Frequency associati ad ogni parola; in questo modo è aumentata la rilevanza della rappresentazione di un documento dei termini che sono meno comuni nella collezione.

Questa rappresentazione vettoriale di un documento permette di calcolare la similarità semantica tra un termine e un documento, oltre che tra due termini. Questa funzionalità è alla base dell'implementazione di un motore di ricerca semantico, ossia che tiene conto del significato dei termini. Attraverso la rappresentazione vettoriale dei termini, infatti si può rappresentare anche la query di un utente come il vettore somma dei termini presenti in essa. Si può, dunque, calcolare la rilevanza dei documenti presenti nella collezione rispetto alla query come similarità del coseno tra il vettore che rappresenta la query e quelli che rappresentano i documenti; si può, come di consueto, ordinare i

documenti in base alla rilevanza rispetto alla query, mostrando i primi *k* risultati all'utente.

2. Il Progetto TALIA

Il progetto TALIA – acronimo di "Territorial Appropriation of Leading-edge Innovation Actions" – è progetto trasversale finanziato dalla Commissione Europea nel contesto di una call del programma Interreg-Med, del quale è capofila la Regione Puglia. Gli obiettivi più ampi della call erano la promozione dello sviluppo sostenibile supportato dalla tecnologia nell'area del Mediterraneo, e, più specificatamente, aumentare la capacità di comunicazione e cooperazione tra gli attori principali nei più importanti settori socioeconomici nell'UE.

Per raggiungere questi obiettivi, nell'ambito del progetto TALIA è stato sviluppato il prototipo per un sistema di supporto decisionale basato sull'analisi intelligente del contenuto testuale dei documenti scritti in ogni progetto finanziato nell'ambito del programma Interreg-Med. Il sistema di supporto decisionale, che mette a disposizione tre servizi all'utente finale, è basato sul Semantic Framework, una piattaforma per l'indicizzazione e l'analisi semantica dei documenti. Il progetto TALIA, grazie al sistema di supporto alle decisioni, mira a costruire e sviluppare la comunità *Social&Creative* del programma Interreg-Med, la quale lavora per promuovere cluster di innovazione fornendo strumenti che permettono la connessione di progetti modulari con le comunità locali a partire dai partner regionali di ogni progetto.

Il programma Interreg-Med comprende diverse decine di progetti raggruppati in nove gruppi tematici chiamati Community; a loro volte le Community sono raggruppate in Assi che comprendono le Community

con obiettivi comuni: Low Carbon Economy Axis, Natural and Cultural Resources Axis, Innovation Axis.

2.1 Semantic Framework: architettura e funzionalità

Lo scopo del Semantic Framework è quello di rendere accessibile la conoscenza e le informazioni presenti nei deliverable di progetto prodotti nei progetti del programma MED. Il Framework si può considerare la parte di back-end del progetto TALIA, poiché offre le funzionalità per l'estrazione di informazioni che verranno poi utilizzate nel sistema di supporto alle decisioni per soddisfare i bisogni informativi dei policy makers.

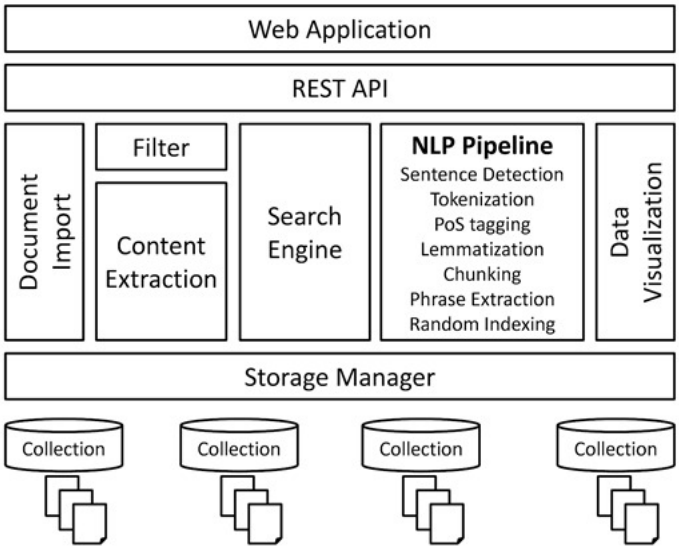
Il Framework consente di gestire collezioni di documenti e di applicare ad ogni documento le operazioni di elaborazione del linguaggio naturale e di rappresentazione vettoriale e semantica descritte nel capitolo precedente. Il Framework è stato sviluppato secondo un'architettura che permette l'accesso ai risultati del processo di elaborazione dei documenti tramite Web API sviluppate secondo il protocollo REST.

2.1.1 Gestione Collezioni

La componente alla base del Semantic Framework è quella che consente di creare e gestire collezioni di documenti, implementata grazie alla libreria Java Apache Lucene. Nell'ambito del progetto TALIA, le collezioni di documenti corrispondono alle Community relative al programma Interreg-Med e contengono tutti i deliverable di progetto che sono stati scritti nell'ambito dei progetti afferenti a ciascuna Community. I documenti sono stati recuperati attraverso l'ausilio di un servizio software esterno al Semantic Framework che ha effettuato il crawling e lo scraping delle pagine web ufficiali di ogni progetto, scaricando i file

dei deliverable e i metadati relativi a ogni progetto codificandoli in formato json.

Attraverso i servizi REST esposti dal Framework, ogni collezione può essere creata specificandone il nome e la lingua, quest'ultima necessaria durante il processo di NLP. Oltre a poter creare collezioni completamente nuove, se ne possono generare di nuove attraverso la fondendone due o più già esistenti. Sono state create, in questo modo, le collezioni di documenti relative agli Assi tematici del programma MED. Per ogni collezione possono essere aggiunti o rimossi documenti; al momento del caricamento in una collezione, ogni documento può essere strutturato in sezioni (come "titolo", "autore", "corpo", ecc.), grazie alle funzionalità messe a disposizione dalla libreria Lucene.



Architettura del Semantic Framework

2.1.2 Motori di ricerca

Per poter estrarre informazioni e conoscenza dai documenti presenti nelle collezioni, su ogni documento è possibile eseguire tutti i passi della pipeline di NLP descritti nel paragrafo 1.1, oltre che le operazioni di indicizzazione basate sul modello a spazio vettoriale o sul Random

Indexing. In seguito alle operazioni di preprocessing ed indicizzazione dei documenti è possibile creare e associare a ciascuna collezione un motore di ricerca classico o semantico.

2.1.3 Gestione dei metadati

Nell'ambito del progetto TALIA, oltre a beneficiare della conoscenza estratta dai documenti, è risultato utile sfruttare le informazioni strutturate relative ad ogni community e progetto. Per questa ragione, è stato integrato nel Semantic Framework un modulo di gestione dei metadati relativi ad ogni collezione, che rende disponibili informazioni più dettagliate riguardo ogni documento presente in essa. Ad esempio, sono disponibili informazioni relative all'autore o agli autori di un deliverable, al progetto cui ogni deliverable afferisce e ai partner coinvolti in esso.

La disponibilità di queste informazioni aggiuntive in forma strutturata è utile ad arricchire i risultati ottenuti tramite l'uso delle funzionalità di ricerca nella Web Application da parte dei policy maker.

2.2 Panoramica sui casi d'uso

Nell'ambito del progetto TALIA, i servizi del Semantic Framework sono stati utilizzati per l'implementazione di due casi d'uso: un motore di ricerca semantico e una matrice di correlazione tra concetti.

2.2.1 Motore di ricerca semantico

Il motore di ricerca semantico ha l'obiettivo di soddisfare i bisogni informativi degli utenti del sistema di supporto alle decisioni, offrendo la possibilità di compiere una ricerca sulle collezioni di documenti relativi a una Community o ad un Asse del programma Interreg-Med.

I risultati di una ricerca sono arricchiti grazie alle funzionalità esposte dal Semantic Framework e grazie alla disponibilità di metadati relativi ai documenti: infatti, l'utente riceve, in risposta ad una query, l'elenco dei documenti più significativi rispetto ad essa, un elenco dei concetti maggiormente correlati a quello cercato e un insieme di metadati che forniscono informazioni più dettagliate relative al documento.

2.2.2 Matrice di correlazione tra concetti

La matrice di correlazione dà all'utente del sistema di supporto alle decisioni la possibilità di esplorare in modo interattivo il contenuto informativo dei documenti di una collezione grazie all'analisi semantica dei testi svolta dal Semantic Framework.

L'esplorazione di una collezione consiste nel poter visualizzare, data una coppia di concetti, un elenco dei concetti maggiormente correlati ad essi all'interno della collezione considerata. Grazie alla presentazione della funzionalità sottoforma di matrice è possibile compiere questa analisi su diverse coppie di concetti contemporaneamente: assegnando un concetto ad ogni posizione sulle righe e sulle colonne, infatti, ciascuna cella della matrice corrisponde alla coppia di concetti presenti sulla riga e sulla colonna corrispondenti.

3. Servizi RESTful e casi d'uso

L'implementazione dei casi d'uso ha richiesto l'utilizzo delle funzionalità offerte dal Semantic Framework esposte tramite servizi RESTful. Ogni caso d'uso ha richiesto l'utilizzo e l'implementazione di servizi specifici, sia per realizzare le funzionalità richieste, che per aumentarne l'efficacia.

3.1 Introduzione alle architetture RESTful

Lo stile architetturale RESTful è una tecnica di sviluppo di servizi Web concepita nel 2000 da Roy Fielding, allora studente di dottorato all'Università della California a Irvine.

Le architetture REST (ossia REpresentational State Transfert), impiegano il protocollo HTTP per trasferire tutte le informazioni in una comunicazione tra host client e server. Nel protocollo HTTP ogni entità che si può richiedere ad un host server è considerata una "risorsa", identificabile e localizzabile univocamente tramite il suo indirizzo URL.

Un servizio web sviluppato secondo quest stile architetturale, dunque, è accessibile a un computer host tramite una semplice richiesta HTTP all'host server che espone il servizio. A seconda che il servizio richiamato abbia lo scopo di modificare o ottenere informazioni presenti sull'host server, la richiesta HTTP dev'essere formulata tramite i metodi HTTP progettati per ciascuno scopo particolare; ad esempio, il metodo GET è specifico per la richiesta di una risorsa senza modifica, mentre i metodi POST e PUT sono specifici per le modifiche delle risorse.