

## Modello di descrizione ed annotazione semantica dei Contenuti Intelligenti

### Survey

Il presente documento riassume i contenuti del deliverable **D2.2\_1 "Modello di descrizione ed annotazione semantica dei Contenuti Intelligenti"** relativo all'attività di ricerca *RI 2.2* "Definizione di modelli di descrizione ed annotazione semantica di contenuti multimediali e servizi interattivi", nell'ambito del secondo Obiettivo Realizzativo (OR 2) "Studio di servizi e contenuti intelligenti" di SAPI.

Per contenuto intelligente, in SAPI, si intende un oggetto multimediale con un limitato livello di interattività. Possiamo distinguere tali contenuti in due diverse categorie:

- **contenuti informativi** i.e. la descrizione di un servizio offerto da Poste;
- **contenuti formativi** i.e. le istruzioni per l'utilizzo di un determinato servizio o di particolari operazioni afferenti ad un servizio.

In altre parole i contenuti informativi legati ad un determinato servizio sono pensati per utenti del sistema SAPI che non sono ancora utenti dello specifico servizio e che necessitano solo di una descrizione di massima dello stesso. I contenuti formativi, di contro, sono pensati per trasferire all'utente di uno specifico servizio la conoscenza necessaria per utilizzare appieno il servizio o una parte di esso.

Mentre i contenuti informativi sono relazionati solo con il servizio che descrivono attraverso una relazione di tipo uno a uno, i contenuti formativi possono essere relazionati con uno o più servizi attraverso una relazione molti a molti e possono, ancora, essere relazionati tra loro indicando eventuali dipendenze (e.g. approfondimenti e prerequisiti). La *Figura 1* schematizza la differenza tra le due categorie di contenuto.

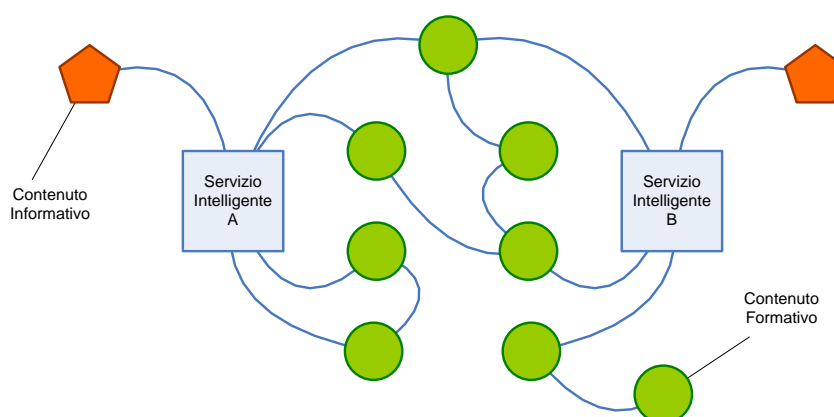


Figura 1: categorie di contenuti intelligenti e relazione con i servizi.

Come si può notare, ciascun servizio intelligente ha uno e un solo contenuto informativo correlato (pentagono) che lo descrive e uno o più contenuti formativi (cerchi). Lo stesso contenuto formativo può essere inoltre correlato a zero o più servizi (possono esistere contenuti formativi non strettamente correlati a servizi) e a zero o più ulteriori contenuti formativi.

Una funzionalità interessante che offrono i contenuti formativi, a differenza di quelli informativi, è la possibilità di essere sequenzializzati dato un obiettivo formativo che si vuole raggiungere. Una sequenza di contenuti formativi è detta **Percorso d'Istruzione**. Ad esempio, se l'utente richiede di essere istruito sull'utilizzo del servizio di pagamento bollettino, un primo contenuto formativo potrebbe introdurre il prodotto bollettino postale, un secondo spiegare le varie fasi del servizio di pagamento on-line del bollettino, un terzo spiegare le forme di pagamento dello stesso (e.g. addebito su conto Banco Posta, addebito su carta di credito, ecc.).

Ovviamente i contenuti erogati, essendo descritti attraverso Entity SAPI, dovranno passare tutte le fasi di personalizzazione della logica di presentazione in modo da essere adattati allo specifico utente e contesto di fruizione. Tuttavia, a differenza dei servizi intelligenti che sono composti da una logica di business e una logica di presentazione, i contenuti intelligenti sono privi di logica di business ovvero hanno una logica di business composta da una singola operazione elementare (del tipo "visualizza contenuto").

Inoltre, nel caso di percorso d'istruzione, la sequenza stessa oltre che i singoli componenti potrà essere adattata sulla base del profilo dell'utente che utilizza il sistema. Nell'esempio precedente, se un utente ha già appreso le forme di pagamento correlate ad altri prodotti di Poste (ad esempio per l'acquisto di materiale filatelico), è inutile riesplorare le stesse cose in relazione al pagamento del bollettino. In tal senso il sistema dovrà tener traccia di tutte le conoscenze apprese in passato dall'utente e spiegare solo quelle ritenute non conosciute al momento della richiesta di un percorso d'istruzione per un dato servizio.

È importante notare che ad ogni servizio intelligente possono inoltre essere associati in SAPI uno o più prodotti di Poste e uno o più operazioni possibili. Ciò consente di sotto-classificare i contenuti formativi associati ad un servizio sulla base delle specifiche operazioni componenti che tali contenuti descrivono. Ovviamente possono esistere contenuti che non riguardano nessuna operazione e contenuti associati ad una o più operazioni diverse. La

Figura 2 seguente mostra un esempio di associazione.

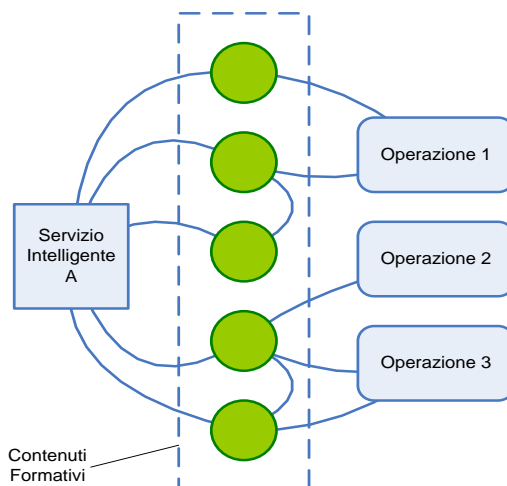


Figura 2: associazione dei contenuti formativi alle operazioni di un servizio.

Come vedremo nel seguito, l'associazione diretta dei contenuti formativi alle operazioni che compongono i singoli servizi consente la realizzazione di un help personalizzato e sensibile al contesto: alla pressione di un pulsante di help, SAPI genererà automaticamente un percorso d'istruzione personalizzato relativo all'operazione corrente.

Il **modello di descrizione dei contenuti intelligenti** consente di descrivere entrambe le tipologie di contenuto (formativa ed informativa) in maniera intellegibile alla macchina. Da un punto di vista formale, per meglio operare con essi, possiamo scollegare il contenuto fisico vero e proprio dalle relazioni che collegano contenuti e servizi in tre diversi livelli di astrazione.

Il livello più basso è costituito dal *contenuto* vero e proprio ovvero dai moduli di formazione elementari che possono essere usati nella costituzione di un percorso d'istruzione. Il secondo livello è costituito dal *metadata* il cui compito è di descrivere formalmente i contenuti elementari attraverso un insieme di attributi. Ad un terzo livello di astrazione abbiamo il *grafo degli argomenti* che descrive le relazioni tra gli argomenti correlati ad un dato servizio intelligente (vedi Figura 3).

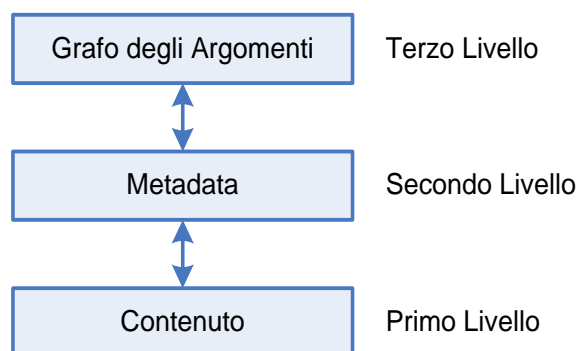


Figura 3: I tre livelli del modello di descrizione dei contenuti intelligenti.

Formalmente possiamo definire il modello di descrizione dei contenuti intelligenti come la terna

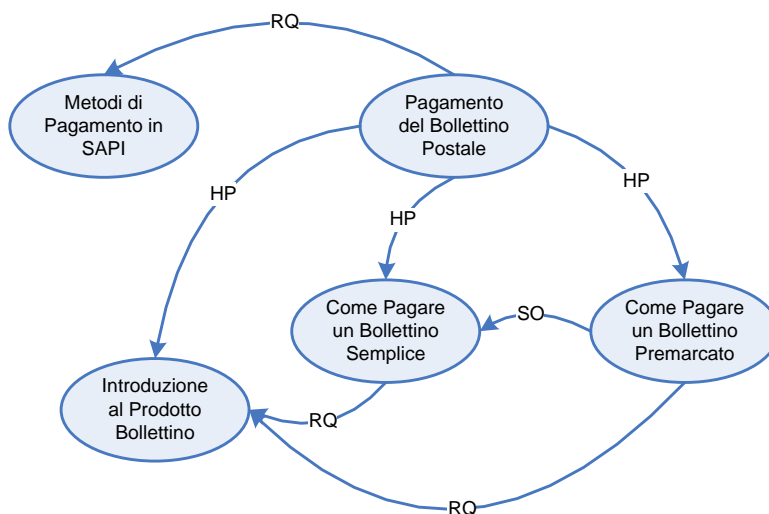
$M = (GA, MD, CO)$  dove  $GA$  è un grafo di argomenti,  $MD$  è un insieme di metadata ed  $CO$  è un insieme di contenuti.

I **Grafi degli Argomenti** consentono di descrivere un insieme di argomenti ed un insieme di relazioni tra di essi. Esso può essere visto dunque come la coppia  $GA = (Ga, Gr)$  dove  $Ga$  è l'insieme di argomenti e  $Gr$  è l'insieme delle relazioni tra argomenti composto da terne  $(tipo, a_1, a_2)$  dove  $tipo \in \{HP, RQ, SO\}$  è il tipo di relazione mentre  $a_1$  e  $a_2$  sono argomenti di  $Ga$ . Le tipologie di relazioni supportate sono le seguenti.

- Relazione *HP* (Has Part) per definire una gerarchia di argomenti. La relazione  $(HP, c, d)$  significa che l'argomento  $d$  è parte dell'argomento  $c$ . Sia  $D = \{d \in Ga \mid (HP, c, d) \in Gr\}$  è possibile dire che, per apprendere  $c$ , è necessario e sufficiente apprendere tutti gli argomenti appartenenti a  $D$ .

- Relazione  $RQ$  (Requires) per definire la propedeuticità.  $(RQ, c, d)$  significa che  $d$  è pre-requisito per  $c$ . Sia  $D = \{d \in Ga \mid (RQ, c, d) \in Gr\}$  è possibile dire che, per apprendere  $c$ , è necessario aver appreso preliminarmente tutti gli argomenti appartenenti a  $D$ .
- Relazione  $SO$  (Suggested Order) per definire una propedeuticità più lasca.  $(SO, c, d)$  significa che, se occorre apprendere  $c$  e  $d$ , è conveniente apprendere  $c$  dopo  $d$ .

La *Figura 4* mostra un esempio di Grafo di Argomenti calato nel dominio di Poste per apprendere gli argomenti correlati al corretto utilizzo del servizio di pagamento bollettino.



*Figura 4: Un esempio di grafo di argomenti.*

Nell'esempio riportato, per apprendere l'argomento "pagamento del bollettino postale" (e quindi imparare ad utilizzare il servizio corrispondente) sarà necessario apprendere gli argomenti "introduzione al prodotto bollettino", "come pagare un bollettino semplice" e "come pagare un bollettino premarcato". A ciascuno di questi argomenti sarà associato uno specifico contenuto formativo che verrà presentato all'utente.

Nella generazione del miglior percorso d'istruzione, tuttavia, bisogna tener conto anche di eventuali prerequisiti degli argomenti da trattare. Nel caso citato, come unico prerequisito dell'argomento "pagamento del bollettino postale" c'è l'argomento "metodi di pagamento in SAPI" che dovrà essere opportunamente introdotto prima degli altri.

Nel caso in cui l'utente non voglia essere edotto sul completo utilizzo del servizio di pagamento bollettino ma solo su alcune sue funzionalità il sistema, basandosi sul grafo di argomenti su riportato, potrà costruire versioni personalizzate del percorso d'istruzione che tengano conto solo degli argomenti effettivamente necessari caso per caso. Ad esempio se l'utente vuole imparare "come pagare un bollettino premarcato", egli dovrà innanzitutto conoscere i "metodi di pagamento in SAPI", poi una "introduzione al prodotto bollettino" e, solo alla fine, "come pagare un bollettino premarcato".

Come si può notare, gli argomenti di un percorso d'istruzione dovranno necessariamente essere ordinati in maniera inversa alle relazioni di tipo RQ e SO. Inoltre, mentre la relazione RQ partecipa attivamente alla definizione del percorso d'istruzione (un prerequisito viene

inserito nel percorso), la relazione SO serve solo ad ordinare argomenti già inclusi nel percorso.

Passiamo ora alla descrizione del livello immediatamente inferiore per la descrizione dei contenuti intelligenti, ovvero i **Metadata**. Per i nostri scopi possiamo definire un metadata  $m \in MD$  come la tripla  $m = (mco, mt, mc)$  dove:

- $mco \in CO$  è il contenuto indicizzato dal metadata  $m$  (importante notare che ad ogni contenuto appartenente ad  $CO$  corrisponde uno ed un solo metadata appartenente ad  $MD$ );
- $mt$  è la tipologia di contenuto e può assumere il valore *formativo* e *informativo*;
- $mc \subseteq Ga$  è l'insieme di argomenti atomici del Grafo di Argomenti coperti (ovvero spiegati) dal contenuto  $mco$  (è importante che un contenuto spieghi solo degli argomenti atomici).

Sulla base delle informazioni contenute nel Grafo degli Argomento e nei Metadata, un **Percorso d'Istruzione** definisce le modalità ottimali di trasferimento della conoscenza circa un determinato servizio intelligente ad uno specifico utente tenendo conto della conoscenza pregressa dello stesso. Il modello del percorso d'istruzione può formalmente essere espresso come una tripla  $PI = (AO, LP, PR)$  composta dalle seguenti strutture:

- l'insieme  $AO$  degli *Argomenti Obiettivo*, ovvero degli argomenti che l'utente conoscerà alla fine del percorso e sono usualmente (ma non necessariamente) coincidenti con l'utilizzo di un determinato servizio intelligente;
- la sequenza ordinata  $LP$  detta *Learning Path*, ovvero la sequenza di argomenti atomici del grafo di argomenti di riferimento, ordinati in base alle reciproche propedeuticità, che si dovrà trasferire all'utente nell'ordine dato e nel rispetto dello specifico profilo come definito al paragrafo successivo;
- la sequenza ordinata  $PR$  detta *Presentazione* rappresenta la sequenza di contenuti corrispondente al *Learning Path* concettuale che permetterà l'effettivo trasferimento di conoscenza all'utente degli *Argomenti Obiettivo* del corso.

Al fine di utilizzare proficuamente i contenuti intelligenti, in particolare quelli di tipo formativo, il modello del profilo utente di SAPI dovrà essere esteso per poter catturare le conoscenze acquisite dagli utenti nella fruizione di contenuti formativi. Chiameremo questa sezione del modello del profilo Stato Cognitivo.

Lo **Stato Cognitivo** si interesserà di rappresentare, per ciascun utente, le conoscenze possedute in un determinato istante sotto forma di "belief", ovvero di coppie argomento-valutazione. Più formalmente, sia  $M = (GA, MD, CO)$  il modello di descrizione dei contenuti intelligenti, uno *Stato Cognitivo* compatibile con tale modello può essere espresso come  $CS = \{B_1, B_2, \dots, B_n\}$  dove ogni belief  $B_i = (c_i, e_i)$  è composto da un argomento  $c_i \in Ga$  e da un valore  $e_i$  che rappresenta il grado di conoscenza dell'argomento da parte dell'utente che assume valori compresi fra  $0 \leq e_i \leq 1$ .

Si definisce insieme degli **Argomenti Conosciuti** dall'utente a cui il modello si riferisce, l'insieme di tutti gli argomenti che hanno grado di conoscenza maggiore di una costante data  $e_{min}$ . Tale costante è definita empiricamente nell'intervallo  $0 \leq e_{min} \leq 1$  e rappresenta il grado minimo di conoscenza di un argomento per far sì che questo venga considerato conosciuto.

La valutazione degli argomenti conosciuti da un dato utente avviene per aggiornamenti successivi. Inizialmente il grado di conoscenza di ciascun argomento è posto pari a 0. Per ciascun argomento interessato dalle operazioni svolte dall'utente nell'ambito di SAPI, l'aggiornamento avverrà mediando la valutazione attuale dell'efficacia di esecuzione di una data operazione (che si pone per default pari ad 1) con le eventuali valutazioni ottenute in precedenza sullo stesso argomento.

Dato un modello di descrizione dei contenuti intelligenti  $M = (GA, MD, CO)$  ed uno stato cognitivo  $CS$  appartenente al profilo di un dato utente di SAPI, è stato definito l'algoritmo di **Generazione Automatica di un Percorso d'Istruzione** a partire da un dato insieme di Argomenti Obiettivo  $AO$ . La descrizione formale dell'algoritmo è fuori dagli scopi della presente sintesi. Si preferisce dunque fornire una descrizione qualitativa del processo attraverso l'esempio seguente.

Supponiamo di avere il seguente grafo di argomenti e che l'utente decida in un determinato istante di apprendere l'argomento target  $c_c$ .

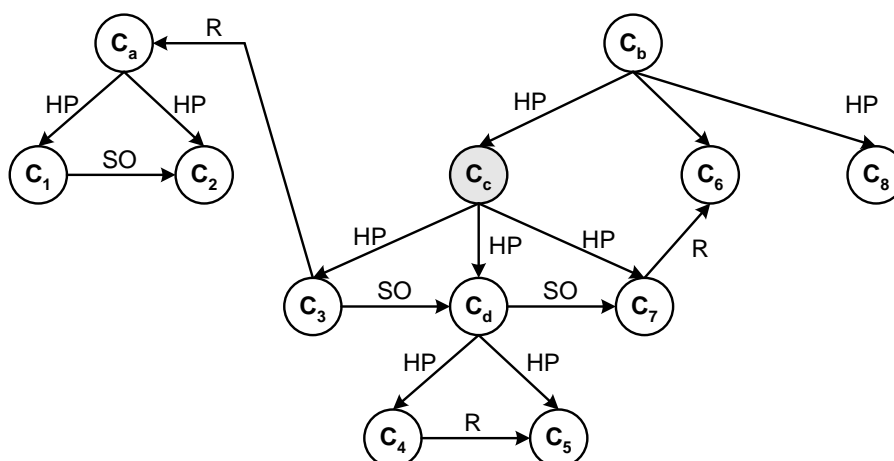


Figura 5: Il Grafo di Argomenti di partenza.

Innanzitutto SAPI dovrà individuare il learning path a partire dall'argomento target del grafo di riferimento e dallo stato cognitivo attuale dell'utente. Supponendo che all'inizio lo stato cognitivo sia vuoto, SAPI genererà il learning path nella maniera seguente:

- $c_c$  si decompone nella sequenza  $(c_3, c_d, c_7)$ ;
- $c_3$  ha per requisito  $c_a$  che viene, dunque fatto precedere a  $c_3$ ;
- la sequenza diventa  $(c_a, c_3, c_d, c_7)$ ;
- $c_a$  si decompone nella sequenza:  $(c_1, c_2)$  che viene sostituita a  $c_a$
- la sequenza diventa  $(c_1, c_2, c_3, c_d, c_7)$ ;
- $c_d$  si decompone nella sequenza:  $(c_4, c_5)$  che viene sostituita a  $c_d$ ;
- la sequenza diventa  $(c_1, c_2, c_3, c_4, c_5, c_7)$ ;
- $c_7$  ha per requisito  $c_6$  che viene, dunque fatto precedere a  $c_7$ ;

– la sequenza finale diventa:  $(c_1, c_2, c_3, c_4, c_5, c_6, c_7)$ ;

Il learning path finale è dunque:

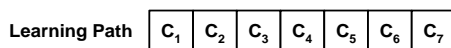


Figura 6: Il Learning Path risultante.

Se l'utente dovesse già conoscere alcuni degli argomenti del learning path (come risultante dallo stato cognitivo), tali argomenti andrebbero rimossi dal learning path ed il procedimento continuerebbe allo stesso modo.

Basandosi, poi, sulle corrispondenze tra contenuti formativi e argomenti, SAPI genererà la presentazione ottimale. Nella figura seguente sono mostrate alcune possibili corrispondenze di contenuti formativi (indicati con  $L_k$  dove  $k$  è un numero progressivo) con gli argomenti del grafo di riferimento. Come si può notare possono esistere contenuti che coprono più argomenti (ad esempio  $L_1$  copre  $c_1$  e  $c_2$ ).

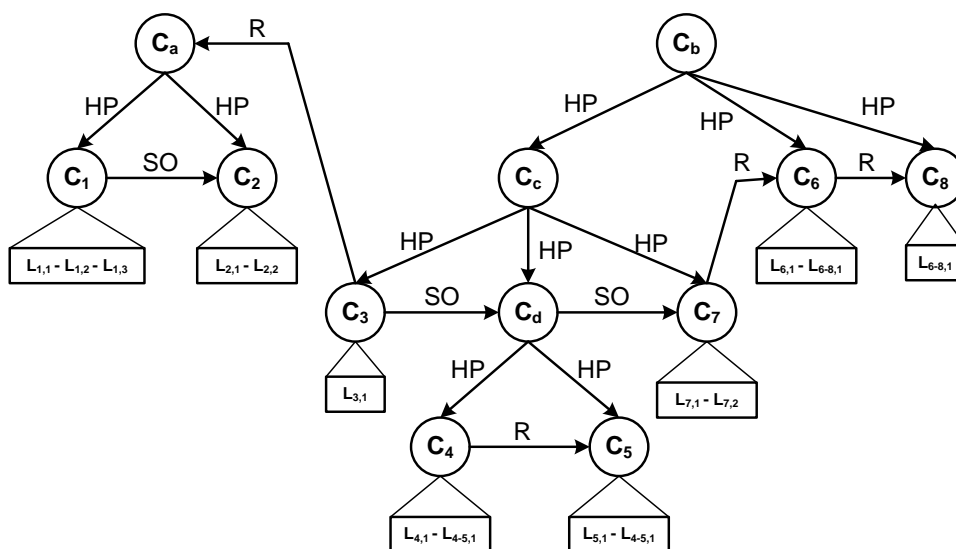


Figura 7: Connessione dei contenuti formativi agli argomenti del grafo.

La presentazione generata conterrà dunque solo contenuti formativi di tipo espositivo scelti in maniera casuale tra quelli appartenenti ai singoli argomenti da spiegare nel rispetto dei vincoli introdotti in 3.1 e cercando di evitare sovrapposizioni (argomenti spiegati più volte nel corso della presentazione). La figura seguente mostra la sequenza finale di contenuti formativi che verrà presentata all'utente per l'esempio discusso.

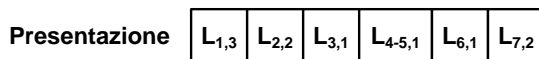


Figura 8: La presentazione risultante.

La presentazione così ottenuta costituirà l'elemento finale del percorso d'istruzione che potrà essere sottoposto all'utente al fine di assicurare la comprensione dell'argomento target iniziale.