

SISTEMA DI CLASSIFICAZIONE

NICOLA MORETTO (MATR. 578258)

10 settembre 2012

Il documento presenta i risultati delle fasi di analisi e di progettazione dei nuovi criteri di classificazione.

VERSIONE	DATA	MODIFICHE
0.1	10-09-2012	Prima stesura del documento.
0.2	11-09-2012	Aggiunto il capitolo CONTENUTI INFORMATIVI.
0.3	12-09-2012	Aggiunto il capitolo SCENARI D'USO.
0.4	13-09-2012	Ampliato il capitolo SCENARI D'USO.
0.5	14-09-2012	Aggiunto il capitolo REQUISITI.

Tabella 1: Registro delle modifiche

INDICE

1	CONTENUTI INFORMATIVI	6
1.1	Classi di contenuti	6
1.2	Relazioni tra contenuti	6
2	SCENARI D'USO	7
2.1	Gestione delle etichette	7
2.2	Inserimento di un contenuto	8
2.3	Consultazione dei contenuti	9
3	REQUISITI	11
3.1	Gerarchia delle etichette	11

ELENCO DELLE FIGURE

ELENCO DELLE TABELLE

Tabella 1	Registro delle modifiche	3
-----------	--------------------------	---

CONTENUTI INFORMATIVI

1.1 CLASSI DI CONTENUTI

1.1.1 *Documento*

1.1.2 *Domanda*

1.1.3 *Evento*

1.1.4 *Multimedia*

1.1.5 *Pensiero*

1.1.6 *Risposta*

1.2 RELAZIONI TRA CONTENUTI

SCENARI D'USO

2.1 GESTIONE DELLE ETICHETTE

2.1.1 *Aggiunta di un sinonimo ad un'etichetta primaria*

Gli utenti possono scegliere etichette differenti per riferire la stessa entità (concreta o astratta): ove si trascurino i legami sinonimici tra le etichette, tale duplicità determina la parzialità dei risultati di una ricerca sui contenuti a seconda dell'etichetta scelta, essendo restituito il sottoinsieme di contenuti nei quali l'entità sia identificata precisamente da tale etichetta.

Duplicità sintattica

L'esito desiderato della ricerca consiste invece nell'insieme di contenuti in cui l'entità sia riferita, a prescindere dalla specifica etichetta utilizzata: in altre parole, si desidera che la ricerca venga trasferita dal piano puramente sintattico (l'etichetta specifica) a quello semantico (l'entità indicata dall'etichetta).

Sintassi e semantica

Per evitare la proliferazione di etichette duplicate, ossia sintatticamente differenti ma riferenti la medesima entità, è utile, per ogni entità, 1) definire un'etichetta che la identifichi chiaramente all'interno della piattaforma e 2) tenere traccia dei sinonimi utilizzati dagli utenti per riferire tale entità.

Entità e etichette

I sinonimi vengono quindi mantenuti per garantire una maggior copertura sintattica, aiutando a stabilire se un'etichetta cercata o scelta dall'utente sia già presente nel dizionario, ma il loro impiego è scoraggiato a favore delle corrispettive chiavi primarie, al fine di rendere più efficiente la catalogazione, la ricerca, la navigazione e la consultazione dei contenuti stessi.

Copertura sintattica

Tali esigenze ci portano a distinguere due classi di etichette:

Classi di etichette

ETICHETTA PRIMARIA

Ciascuna etichetta primaria identifica univocamente un'entità specifica del dominio.

ETICHETTA SINONIMICA

Ciascuna etichetta sinonimica rappresenta un sinonimo per una (e una sola) etichetta primaria.

2.1.2 *Aggiunta di un'accezione ad un'etichetta*

Ciascuna etichetta può riferirsi a entità differenti a seconda del contesto in cui viene utilizzata: nell'ambito della piattaforma, diventa cruciale poter precisare facoltativamente l'accezione dell'etichetta (e quindi l'entità specifica) rispetto alla quale si vuole effettuare una ricerca sui contenuti informativi.

Ambiguità semantica

Ad ogni etichetta è possibile associare, sotto forma di metadati, diverse accezioni, ciascuna delle quali identifica univocamente un possibile significato o entità.

*Accezioni*2.1.3 *Aggiunta di un'etichetta*

Il dizionario rappresenta - nella visione più elementare - un insieme di etichette primarie e sinonimi: più in dettaglio, è possibile immaginarlo come una partizione, ciascuno dei cui sottoinsiemi corrisponde ad un'entità

Partizione

del dominio e contiene esattamente un'etichetta primaria, che identifica univocamente il sottoinsieme/entità, e gli eventuali sinonimi (in numero arbitrario, anche nullo).

L'aggiunta di una nuova etichetta consiste innanzi tutto nello stabilire se si tratti di:

Inserimento

ETICHETTA PRIMARIA

La nuova etichetta contribuisce a formare un sottoinsieme logico distinto, contenete un unico elemento (la chiave stessa).

ETICHETTA SINONIMICA

La nuova etichetta viene inserita nel sottoinsieme logico identificato dall'unica etichetta primaria di cui è sinonimo.

Non si dà il caso che una nuova etichetta S_1 possa essere sinonimo di due (o più) etichette primarie. Si considerino ad esempio due etichette primarie, E_1 e E_2 : per la proprietà transitiva, se E_1 è sinonimo di S_1 e E_2 è sinonimo di S_1 , allora E_1 e E_2 sono a loro volta sinonimi; ma allora, in accordo ai principi sopra illustrati, l'ultima tra E_1 e E_2 ad essere stata aggiunta doveva essere inserita nel sottoinsieme dell'altra, contraddicendo le ipotesi iniziali.

Uno-a-molti

2.1.4 Eliminazione di un'etichetta

In maniera simile a quanto visto per l'aggiunta di un'etichetta, occorre distinguere due casi, a seconda che si tratti di:

ETICHETTA PRIMARIA

L'etichetta viene eliminata e un sinonimo dell'etichetta viene promosso a etichetta primaria.

ETICHETTA SINONIMICA

Il sinonimo viene semplicemente eliminato.

2.2 INSERIMENTO DI UN CONTENUTO

2.2.1 Assegnazione di un'etichetta

L'assegnazione di un'etichetta ad un contenuto consiste nell'individuazione di parole o brevi espressioni chiave, che identifichino un'entità concreta (luogo, persona, oggetto, ...) o astratta (concetto, argomento, ...) riferita o citata all'interno del contenuto stesso.

Una volta individuata la parola o espressione, il sistema deve verificare se essa sia già stata utilizzata in precedenza e quindi presente nel dizionario interno: in caso affermativo, possono verificarsi due casi:

Etichetta esistente

ETICHETTA PRIMARIA

L'etichetta viene associata al contenuto.

ETICHETTA SINONIMICA

L'etichetta viene automaticamente rimpiazzata con la corrispondente etichetta primaria.

In caso contrario, viene indagata la presenza nel dizionario interno di etichette aventi significato analogo a quella prescelta dall'utente: a seconda dell'esito della ricerca possono verificarsi due casi:

Nuova etichetta

NESSUN RISULTATO

La parola o espressione viene memorizzata nel dizionario come etichetta primaria.

ETICHETTA PRIMARIA

La parola o espressione viene memorizzata nel dizionario come sinonimo dell'etichetta primaria.

In ogni caso, al termine della procedura al contenuto viene assegnata un'etichetta primaria e l'utente ha facoltà di specificare - ove disponibile - un'accezione.

2.2.2 Rimozione di un'etichetta

La rimozione di un'etichetta assegnata in precedenza ad un contenuto non modifica in alcun modo il dizionario interno, anche qualora non risultasse assegnata ad altri contenuti.

2.3 CONSULTAZIONE DEI CONTENUTI

La ricerca e la consultazione dei contenuti rappresentano attività cruciali per gli utenti della piattaforma e ci si affida al criterio di classificazione delle etichette per reperire in maniera efficiente le informazioni cercate; l'approccio e lo scopo con cui gli utenti navigano l'insieme di contenuti disponibili all'interno della piattaforma può tuttavia differire sensibilmente.

2.3.1 Esplorazione delle etichette

L'utente alle prime armi o semplicemente interessato a conoscere gli argomenti discussi nella piattaforma esplora i contenuti informativi a partire dalle etichette: per facilitare la navigazione di queste ultime dovrebbe essere disponibile una struttura gerarchica, che le raccolga e le cataloghi in maniera ordinata.

Gerarchia

Tale soluzione permette all'utente di individuare più rapidamente gli argomenti di interesse mediante un PROCESSO DICOTOMICO, che partendo dagli argomenti generali proceda per ulteriori raffinamenti ad individuare con crescente precisione e accuratezza i temi di interesse, senza dover consultare l'insieme delle etichette nella sua completezza.

Dicotomia

Gli elementi della gerarchia sono etichette primarie, i cui sinonimi e accezioni vengono mostrati come metadati associati.

Elementi

2.3.2 Ricerca di un'etichetta

L'utente alla ricerca di informazioni su un tema particolare cerca di individuare le etichette esistenti aventi maggiore attinenza e rilevanza con il tema in questione: la ricerca di corrispondenze nel dizionario prevede che:

1. vengano prese in esame tutte le etichette, poiché contemplando non solo le chiavi primarie ma anche i relativi sinonimi aumenta la probabilità di ottenere riscontri positivi (maggiore copertura sintattica);
2. vengano restituite le chiavi primarie corrispondenti alla ricerca: qualora si abbia riscontro su un sinonimo, viene restituita la corrispondente chiave primaria.

2.3.3 Ricerca contenuti per etichetta

La ricerca di informazioni riguardanti un tema specifico viene effettuata specificando una o più etichette, eventualmente declinate nelle specifiche accezioni, che presentino agli occhi dell'utente particolare attinenza con il tema in questione e siano dunque con maggior probabilità associate ai contenuti di interesse.

Etichette e accezioni

Siano U_e l'insieme delle etichette inserite e C_e l'insieme delle etichette assegnate ad un generico contenuto: il primo passo consiste nel sostituire le etichette sinonimiche con le equivalenti primarie ed estendere l'insieme U_e alle etichette figlie di ogni $e \in U_e$.

Insiemi di etichette

A questo punto si possono distinguere tre casi principali, a seconda del grado di corrispondenza/attinenza dei contenuti rispetto alle etichette cercate:

Corrispondenza

CORRISPONDENZA COMPLETA: $U_e \subseteq C_e$

Al contenuto sono state assegnate tutte le etichette richieste dall'utente e verrà quindi visualizzato in cima ai risultati di ricerca (massima attinenza).

CORRISPONDENZA PARZIALE: $U_e \cap C_e \neq \emptyset$

Al contenuto sono state assegnate parte delle etichette richieste dall'utente (media attinenza).

NESSUNA CORRISPONDENZA: $U_e \cap C_e = \emptyset$

Al contenuto non risulta assegnata alcuna etichetta richiesta dall'utente (attinenza nulla).

I contenuti attinenti vengono visualizzati in ordine decrescente rispetto al numero di etichette assegnate corrispondenti a quelle richieste dall'utente:

Attinenza

$$|U_e \cap C_e|$$

2.3.4 Ricerca di contenuti affini

La ricerca di contenuti affini consiste nell'identificare, a partire da un contenuto dato, altri la cui pertinenza rispetto al tema trattato sia massima: in questo scenario valgono le medesime considerazioni emerse nella sezione precedente, previa sostituzione di U_e con l'insieme delle etichette assegnate al contenuto corrente.

REQUISITI

3.1 GERARCHIA DELLE ETICHETTE