

# SISTEMA DI CLASSIFICAZIONE

NICOLA MORETTO (MATR. 578258)

11 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 REQUISITI.
0.4	13-09-2012	Ampliato il capitolo REQUISITI.
0.5	14-09-2012	Rivisto il capitolo REQUISITI.

Tabella 1: Registro delle modifiche

## INDICE

---

1	CONTENUTI INFORMATIVI	6	
1.1	Introduzione	6	
1.2	Classi di contenuti	6	
1.2.1	Documento	6	
1.2.2	Domanda	6	
1.2.3	Evento	6	
1.2.4	Multimedia	6	
1.2.5	Pensiero	6	
1.2.6	Risposta	6	
1.3	Relazioni tra contenuti	6	
1.3.1	Gerarchia	6	
1.3.2	Discussione	6	
2	REQUISITI	7	
2.1	Gestione delle etichette	7	
2.1.1	Sinonimi	7	
2.1.2	Accezioni	8	
2.1.3	Dizionario	9	
2.1.4	Gerarchia (padri e figli)	10	
2.2	Gestione dei contenuti	10	
2.2.1	Assegnazione etichette	10	
2.3	Consultazione dei contenuti	11	
2.3.1	Esplorazione gerarchica delle etichette	11	
2.3.2	Ricerca di un'etichetta	11	
2.3.3	Ricerca di contenuti mediante etichette	11	
2.3.4	Ricerca di contenuti affini	12	

## ELENCO DELLE FIGURE

---

Figura 1	Dizionario delle etichette	7
Figura 2	Ambiguità sintattica	7
Figura 3	Etichette primarie e sinonimiche	8
Figura 4	Ambiguità semantica	8
Figura 5	Accezioni e sottoinsiemi	8

## ELENCO DELLE TABELLE

---

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

## CONTENUTI INFORMATIVI

---

### 1.1 INTRODUZIONE

Il patrimonio di conoscenza della piattaforma è dato dai contenuti generati dagli utenti e pubblicati nella piattaforma: essi condividono i medesimi meccanismi e criteri di classificazione, alcune proprietà essenziali (autore, data di pubblicazione, visibilità, ...), un contenuto informativo vero e proprio (di lunghezza variabile a seconda della classe)

### 1.2 CLASSI DI CONTENUTI

#### 1.2.1 *Documento*

#### 1.2.2 *Domanda*

#### 1.2.3 *Evento*

#### 1.2.4 *Multimedia*

#### 1.2.5 *Pensiero*

#### 1.2.6 *Risposta*

### 1.3 RELAZIONI TRA CONTENUTI

All'interno della piattaforma il generico contenuto riveste un ruolo essenziale poiché rappresenta l'astrazione fondamentale di tutti i tipi di contenuti, sul quale converge la maggior parte delle relazioni, sia interne (tra i contenuti stessi) sia esterne (criteri di classificazione, ...), al fine di semplificare l'eventuale integrazione di nuovi criteri di classificazione o classi di contenuti nel futuro.

*Contenuto generico*

#### 1.3.1 *Gerarchia*

#### 1.3.2 *Discussione*

## REQUISITI

### 2.1 GESTIONE DELLE ETICHETTE

Il dizionario della piattaforma rappresenta - nella visione più elementare - un insieme di etichette: più in dettaglio, è conveniente immaginarlo e trattarlo come l'unione di sottoinsiemi, ciascuno dei quali corrisponde ad un'entità del dominio e contiene esattamente un'ETICHETTA PRIMARIA, che identifica univocamente il sottoinsieme/entità in questione, e gli eventuali SINONIMI (in numero arbitrario, anche nullo).

*Modello concettuale*



Figura 1: Dizionario delle etichette

#### 2.1.1 Ciascuna etichetta primaria può avere 0...n sinonimi

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

*Ambiguità sintattica*



Figura 2: Ambiguità sintattica: un significante, molti significati

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:

*Etichette primarie e sinonimiche*

1. definire un'etichetta che la identifichi chiaramente all'interno della piattaforma (ETICHETTA PRIMARIA);
2. tenere traccia dei sinonimi utilizzati dagli utenti per riferire tale entità (ETICHETTE SINONIMICHE).

I sinonimi vengono mantenuti per garantire una maggior copertura sin-

*Copertura sintattica*



Figura 3: Etichette primarie e sinonimiche

tattica, aiutando a stabilire se un'etichetta cercata o scelta dall'utente sia già presente nel dizionario, sebbene ai contenuti vengano assegnate le corrispondenti chiavi primarie, al fine di rendere più efficiente la catalogazione, la ricerca, la navigazione e la consultazione dei contenuti stessi.

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

Ogni qualvolta un utente propone una nuova etichetta  $E_1$ , che risulti sinonimo di un'altra esistente  $E_0$ , essa viene aggiunta al dizionario interno della piattaforma come sinonimo di  $E_0$ : da quel momento, qualora un utente provi ad assegnare quella medesima parola o espressione ad un contenuto della piattaforma, il sistema la identificherà come sinonimo di  $E_0$  e assegnerà la corrispondente etichetta primaria al contenuto.

#### 2.1.2 Ciascuna etichetta può avere 0...n accezioni

Ciascuna etichetta può riferirsi a entità differenti a seconda del contesto in cui viene utilizzata, perciò diventa cruciale poter precisare l'accezione dell'etichetta, ossia l'entità cui si riferisce.

*Ambiguità semantica*



Figura 4: Ambiguità semantica: un significato, molti significanti

Con l'introduzione di tale concetto, il dizionario della piattaforma acquisisce una nuova dimensione poiché ciascuna etichetta può appartenere contemporaneamente a diversi sottoinsiemi: poiché infatti ciascuno di essi corrisponde ad un'entità distinta, ogni accezione di un'etichetta, rimandando ad un'entità distinta, colloca l'etichetta nel sottoinsieme corrispondente.



Figura 5: Entità distinte: accezioni e sottoinsiemi di un'etichetta

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

L'aggiunta di un'accezione ad un'etichetta consiste nel definire il contesto o ambito in cui essa assuma un significato univoco e non equivocabile.



*Eliminazione di un'accezione associata ad un'etichetta*

L'eliminazione di un'accezione associata ad un'etichetta prevede due possibili casi:

**ETICHETTA PRIMARIA**

Se l'etichetta è primaria l'accezione viene eliminata e un sinonimo viene promosso in sua vece ad etichetta primaria del sottoinsieme.

**ETICHETTA SINONIMICA**

Se l'etichetta è sinonimica si procede direttamente alla cancellazione dell'accezione.

*2.1.3 Il dizionario contiene 0...n etichette*

Il dizionario rappresenta l'insieme delle etichette e delle reciproche relazioni. L'inserimento di una nuova etichetta implica l'assegnazione di una singola accezione, implicita o esplicita.

*Aggiunta di un'etichetta primaria*

L'aggiunta di un'etichetta primaria consiste - da un punto di vista logico - alla formazione di un sottoinsieme (della partizione del dizionario) distinto, contenente un unico elemento, ossia l'etichetta stessa.

*Aggiunta di un'etichetta sinonimica*

L'aggiunta di un'etichetta sinonimica consiste nell'associare tale etichetta ad una corrispondente primaria, ossia - da un punto di vista logico - all'inserimento della stessa nel sottoinsieme (della partizione del dizionario) identificato dall'unica etichetta primaria di cui è sinonimo.

Non si dà il caso che nuova etichetta  $S_1$  possa essere sinonimo - rispetto ad una specifica accezione - di due (o più) etichette primarie, ma può essere sinonimo di etichette primarie in numero al più pari alle relative accezioni.

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 così le ipotesi iniziali.

*Accezioni e sinonimi*

*Uno-a-molti*

*Eliminazione di un'etichetta*

L'eliminazione di un'etichetta richiede di considerare separatamente ogni possibile accezione, valutando caso per caso:

**ETICHETTA PRIMARIA**

Se l'etichetta è primaria viene eliminata e un sinonimo viene promosso in sua vece ad etichetta primaria del sottoinsieme.

**ETICHETTA SINONIMICA**

Se l'etichetta è sinonimica si procede semplicemente alla sua cancellazione.

#### 2.1.4 *Ciascuna etichetta può avere 0...n padri e 0...m figli*

##### *Aggiunta di un'etichetta padre*

L'aggiunta di un'etichetta padre consiste nell'inserire una nuova etichetta, che rappresenti un tema più generico rispetto a quello corrente.

##### *Aggiunta di un'etichetta figlio*

L'aggiunta di un'etichetta padre consiste nell'inserire una nuova etichetta, che rappresenti un tema più specifico rispetto a quello corrente.

##### *Eliminazione di un'etichetta padre*

##### *Eliminazione di un'etichetta figlio*

## 2.2 GESTIONE DEI CONTENUTI

### 2.2.1 *A ciascun contenuto possono essere assegnate 0...n etichette*

#### *Assegnazione di un'etichetta primaria ad un contenuto*

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.

#### *Eliminazione di un'etichetta primaria associata ad un contenuto*

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 criteri o 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 gerarchica 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 di contenuti mediante etichette

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.