

Analisi e progettazione di un'interfaccia grafica per la
consultazione dei contenuti informativi in una piattaforma
web tematica

Relazione di stage - Bozza

Nicola Moretto (matr. 578258)

19 novembre 2012

Indice

1. Introduzione	6
1.1. Contenuti	6
1.2. Convenzioni tipografiche	6
2. Progetto	8
2.1. Genesi	8
2.2. Reti sociali	8
2.3. Architettura	9
2.3.1. Contenuti informativi	9
2.3.2. Criteri di classificazione	10
3. Stage	11
3.1. Piano di stage	11
3.1.1. Obiettivi e requisiti	11
3.1.2. Pianificazione	12
3.2. Norme di stage	13
3.3. Criterio di classificazione	16
3.3.1. Entità	17
3.3.2. Etichette	19
3.3.3. Contenuti	21
3.3.4. Utente	21
3.3.5. Modello relazionale	23
3.4. Fase 2: interfaccia grafica	26
3.4.1. Casi d'uso	27
3.4.2. Requisiti	29
3.4.3. Progettazione	29
4. Valutazioni finali	30
4.1. Consuntivo	30
4.2. Competenze professionali	30
4.3. Stage e università	30
A. Glossario	31

Indice

B. Criterio di classificazione	34
C. Interfaccia grafica	38

Elenco delle figure

3.1. Diagramma di Gantt	14
3.2. Contenuti informativi e conoscenza	16
3.3. Valore informativo di un contenuto	17
3.4. Dominio di conoscenza della piattaforma	17
3.5. Ambiguità sintattica di un'entità	18
3.6. Sintassi e semantica di un'entità	18
3.7. Relazioni tra entità	19
3.8. Accezioni di un'etichetta	20
3.9. Modello relazione del criterio di classificazione	23
3.10. Modello ad oggetti delle accezioni	25

Elenco delle tabelle

3.1. Pianificazione settimanale delle attività	13
3.2. Configurazione dell'ambiente di lavoro	14

L'attività di stage si è svolta presso l'azienda *Sintesi Sas*, che opera nel settore IT (*Information Technology*) realizzando software ERP e piattaforme Web per aziende, in particolare attive nel settore turistico, e fornendo servizi di consulenza e di formazione di imprenditori nell'ambito del marketing strategico, operativo e del controllo di gestione.

Il prodotto di punta dell'azienda - *Planet Hotel* - costituisce un sistema software per la gestione alberghiera tra i più flessibili, ampi e completi presenti nel panorama italiano, in grado di coprire la maggior parte delle necessità aziendali: oltre alla gestione delle prenotazioni e dei conti, esso offre un insieme di moduli integrati per supportare il controllo di gestione e degli interventi di marketing.

Si tratta di una realtà imprenditoriale a clientela nazionale con sede unica a Mestre (VE), la cui direzione e amministrazione è affidata al solo fondatore, che ha assunto il ruolo di tutor esterno e referente aziendale per l'intera durata dello stage.

Le attività svolte si inseriscono nell'ambito di un progetto esterno rispetto al business dell'azienda. finalizzato alla realizzazione di una piattaforma web tematica per la condivisione di informazioni e la vendita diretta di prodotti alla clientela e affidato ad un team costituito da differenti figure professionali (sociologi, informatici, ingegneri, ...).

1. Introduzione

1.1. Contenuti

Il presente documento costituisce una relazione dettagliata in merito all'attività di stage svolta dallo studente Nicola Moretto presso l'azienda *Sintesi Srl*. I contenuti sono organizzati nei seguenti capitoli:

Introduzione

Il primo capitolo illustra brevemente la struttura del documento e le convenzioni tipografiche utilizzate.

Progetto

Il secondo capitolo illustra le strategie dell'azienda e gli obiettivi, i requisiti e i vincoli del progetto in cui si inseriscono le attività di stage.

Stage

Il terzo capitolo illustra gli obiettivi, i requisiti e l'organizzazione (piano e norme di lavoro) delle attività di stage. A seguire vengono presentate le scelte più rilevanti effettuate e i risultati conseguiti.

Valutazioni finali

Il quarto capitolo presenta un'analisi critica a posteriori dell'attività di stage: raggiungimento degli obiettivi prefissati, competenze professionali acquisite,

1.2. Convenzioni tipografiche

Al fine di agevolare la consultazione del documento, sono state adottate alcune convenzioni tipografiche illustrate di seguito.

Glossario Gli acronimi, le abbreviazioni, i nomi propri e i termini specialistici contenuti nel presente documento sono illustrati nel *Glossario*, consultabile in appendice, al fine di agevolare la lettura e la comprensione degli argomenti trattati. La prima occorrenza di ciascun termine o espressione presente nel glossario è riconoscibile per la sottolineatura.

Terminologia La prima occorrenza di termini propri o di provenienza straniera divenuti di uso corrente nella lingua italiana sono evidenziati in *corsivo*, mentre per le parole o espressioni che assumono particolare significato nel presente contesto si ricorre al MAIUSCOLETTO.

Codice e formule I nomi di tabelle, classi, package, ... sono riportati con un carattere di tipo `sans serif`, mentre i frammenti di codice o formule sono riconoscibili per l'impiego di un carattere a `spaziatura fissa`.

2. Progetto

2.1. Genesi

L'idea della piattaforma *Social (Life) Shuttle* nasce nel 2010 da un progetto concepito per creare una comunità virtuale destinata agli artisti sconosciuti accessibile in mobilità mediante un'applicazione dedicata, *ArtYR*.

Nello stesso periodo una consulenza nell'ambito dei sistemi informativi territoriali ad un'azienda di Bolzano conduce allo sviluppo di un'innovativa piattaforma software: si tratta di un sistema informativo territoriale in cui l'erogazione di informazioni turistiche è integrata con la vendita di servizi turistici.

Il progetto evolve - grazie alla partecipazione di Comuni, Province e Regioni - in una rete tematica di agenzie di viaggio con un'identità comune e finalizzata alla fusione dei sistemi informativi distrettuali e di vendita.

L'architettura di *Social (Life) Shuttle* ne trae ispirazione, integrando tre componenti differenti:

1. *business*: vendita diretta di prodotti alla clientela;
2. *social*: creazione e sviluppo delle relazioni sociali attraverso la condivisione di informazioni e conoscenza;
3. sistema di erogazione di informazioni turistiche e territoriali.

2.2. Reti sociali

Il modello sociologico di rete sociale non ha attualmente riscontro presso le piattaforme web di condivisione dei contenuti (blog, forum, ...) o i *social network* (Facebook, Twitter, ...), che si limitano a considerare o concretizzarne singoli aspetti.

Nelle moderne reti sociali è infatti assente l'incentivo alla condivisione e distribuzione della conoscenza, fattore cruciale per l'aggregazione fisica dei membri delle comunità, intese come aggregazioni che si formano intorno ed attraverso la manifestazione di interesse nei confronti di uno specifico tema di dialogo o discussione, che attraversa la sfera individuale, intima e personale dei suoi membri.

Il progetto *Social (Life) Shuttle* rappresenta una nuova generazione delle piattaforme di socializzazione, in cui il web diventa solamente un canale di condivisione e un serbatoio

della conoscenza generata dalla dialettica tra persone e vengono ad essere integrati i canoni classici di *blog*, *forum*, *social network* e *media* .

Una relazione sociale nata e costruita su un interesse comune stravolge l'attuale paradigma delle reti sociali virtuali, in cui il legame nasce a prescindere dalla presenza di interessi comuni o informazioni da condividere, e favorisce l'incontro tra persone aventi esperienze simili frutto di tali interessi condivisi. Ove l'esperienza riguardi anche beni o prodotti, la componente business inende offrire ai membri la possibilità di interagire con i produttori, anch'essi attori della comunità.

L'architettura di *Social (Life) Shuttle* consente di declinare la piattaforma in innumerevoli varianti, applicabili ai temi più svariati: al momento sono in fase di sperimentazione per il mondo del vino, il cibo biologico, l'arte commercializzabile e l'attività di ricerca e progettazione collaborativa.

2.3. Architettura

Tale piattaforma presenta numerose aspetti che la differenziano dalla concorrenza attuale:

- profonda integrazione degli aspetti *social* e *business*;
- nessuna distinzione tra creatori e fruitori dei contenuti (ciascun membro può condividere le proprie esperienze, segnalare eventi, pubblicare articoli critici, ...);
- l'autorevolezza di ciascun membro della comunità si rafforza o si indebolisce a seconda della qualità dei contenuti pubblicati, dei giudizi degli altri membri o di altri criteri di valutazione;
- lo sfruttamento delle tecnologie e dei dispositivi mobili per favorire l'instaurazione o la coltivazione di relazioni al di fuori dell'ambito virtuale della piattaforma (partecipazione ad eventi, raccolta e condivisione di informazioni geolocalizzate, ...).

Per quanto concerne le attività di stage, due aspetti della piattaforma assumono particolare rilevanza: i contenuti informativi e i relativi criteri di classificazione.

2.3.1. Contenuti informativi

I contenuti informativi rappresentano lo strumento essenziale per la condivisione delle esperienze e della conoscenza intorno al tema specifico della piattaforma.

L'individuazione di un insieme di classi di contenuti adatti a esprimere in una forma strutturata le informazioni è frutto di un lungo processo, che ha tenuto conto essenzialmente delle forme di espressione e di comunicazione tipiche nella vita quotidiana, che a loro volta rispecchiano l'intenzione comunicativa dell'azione compiuta e delle parole espresse da un singolo individuo.

- distinzione tra natura e formato del contenuto - natura: forme espressive nello scambio dialettico quotidiano per favorire identificazione immediata del formato appropriato
- formato: supporti del web (audio, video, testo, chat, ...) [analisi comparativa] per stabilire i modi classici e preferiti dagli utenti per condividere informazioni - non imporre vincoli alla libertà creativa degli utenti - facilitare la catalogazione

2.3.2. Criteri di classificazione

3. Stage

3.1. Piano di stage

3.1.1. Obiettivi e requisiti

L'attività di stage si colloca nell'ambito del progetto presentato nel capitolo 2 e persegue due obiettivi distinti ma correlati, focalizzandosi sulla componente *social* della piattaforma.

Criteri di classificazione

Il primo obiettivo consiste nell'estendere l'attuale sistema di classificazione (v. sezione 2.3.2) integrandovi un criterio aggiuntivo per la catalogazione dei contenuti pubblicati dagli utenti e la costruzione di un'enciclopedia della conoscenza per rendere il reperimento e la consultazione delle informazioni desiderate il più efficiente ed agevole possibile.

L'ideazione e concezione del suddetto criterio deve tener conto della natura tematica della piattaforma, riuscendo a conciliare due esigenze distinte:

- dev'essere sufficientemente astratto e flessibile per adattarsi alla molteplicità di varianti tematiche in cui la piattaforma stessa può essere declinata;
- dev'essere ottimizzato per avvantaggiarsi delle peculiarità di una piattaforma tematica, ad esempio la maggior correlazione degli argomenti trattati.

La soluzione individuata deve inoltre prescindere da assunzioni legate alla tecnologia utilizzata. Infine, alla luce di possibili evoluzioni nello sviluppo della piattaforma, si desidera che la classificazione di un contenuto informativo (assegnazione di metadati, individuazione di correlazioni, ...) possa essere - in futuro - demandata a componenti software integrate nella piattaforma.

Interfaccia grafica

Il secondo obiettivo consiste nel progettare un'interfaccia grafica per la consultazione dei contenuti informativi, che sfrutti il criterio di classificazione aggiuntivo per facilitare la ricerca ed il reperimento delle informazioni di interesse per l'utente all'interno del patrimonio enciclopedico della piattaforma. La sfida principale consiste nel progettare un'interfaccia in grado di visualizzare in maniera chiara e ordinata un ridotto o elevato

numero di contenuti, a prescindere dalla classe del dispositivo impiegato (*smartphone, tablet, notebook, ...*).

Il primo passo consiste nell'individuare le informazioni essenziali ad una rapida e precisa identificazione dei contenuti (titolo, autore, data, ...) e valutare la notazione (grafica o testuale) più adatta per esprimerle, al fine di renderle accessibili al maggior numero possibile di utenti; le informazioni aggiuntive devono essere comunque accessibili, ma solo su esplicita richiesta dell'utente. In questo ambito si inseriscono una serie di analisi e valutazioni di carattere sociologico, svolte da altri membri del team di progetto, per individuare le soluzioni più idonee a comunicare tali informazioni in modo da renderne la comprensione chiara e intuitiva a qualsiasi utente.

Il secondo passo richiede di definire le specifiche per un'interfaccia facilmente navigabile, che sia in grado di mostrare in modo ordinato e intuitivo i contenuti e le reciproche relazioni. Occorre perciò individuare opportuni criteri di raggruppamento, ordinamento e collocamento dei contenuti visualizzati per favorirne la consultazione, evitando un sovraccarico cognitivo e garantendo un livello adeguato di leggibilità.

Con il terzo ed ultimo passo si intende aggiungere la possibilità per l'utente di filtrare i contenuti visualizzati in accordo a proprietà (argomento, autore, data di pubblicazione, tipo) o metadati associati (attinenza, emozioni, giudizi, intenzioni). Per gli utenti autenticati si desidera offrire un livello aggiuntivo di personalizzazione, che consenta di filtrare automaticamente i contenuti secondo le preferenze associate al profilo (interessi, livello di esperienza).

Per individuare i requisiti essenziali si prendono innanzi tutto in considerazione alcuni casi d'uso classici:

1. l'utente naviga liberamente tra i contenuti (più recenti, più letti, più discussi, ...);
2. l'utente consulta la discussione generata da un singolo contenuto;
3. l'utente cerca le informazioni riguardanti un certo tema (contenuti affini, ...);
4. l'utente esplora gli argomenti trattati e le reciproche relazioni.

3.1.2. Pianificazione

L'attività di stage viene suddivisa in due fasi distinte per semplificarne la pianificazione:

1. l'estensione del sistema di classificazione;
2. l'analisi e la progettazione dell'interfaccia grafica.

Per ciascuna fase sono fissati gli obiettivi generali, sono individuate e organizzate su base settimanale le attività da svolgere, cercando di garantire un carico di lavoro equilibrato, e sono indicati i prodotti attesi. La durata complessiva dello stage si attesta su 8 settimane a tempo pieno, corrispondenti a 320 ore di lavoro.

ATTIVITÀ	ORE DI LAVORO
<i>Fase 1: estensione del sistema di classificazione</i>	
Analisi delle specifiche del sistema di classificazione	40
Analisi comparativa dei principali sistemi di classificazione della conoscenza	40
Progettazione del sistema di classificazione	40
Implementazione del sistema di classificazione nel modello relazionale	40
<i>Fase 2: analisi e progettazione dell'interfaccia grafica</i>	
Analisi dei requisiti dell'interfaccia grafica	40
Progettazione dell'interfaccia grafica: visualizzazione dei contenuti	40
Progettazione dell'interfaccia grafica: filtraggio dei contenuti	40
Progettazione dell'interfaccia grafica: navigazione dei contenuti	40

Tabella 3.1.: Pianificazione settimanale delle attività

3.2. Norme di stage

Ambiente di lavoro Nel corso dello stage sono stati impiegati diversi strumenti per gestire le attività di progetto e produrre la documentazione prevista.

Documentazione La documentazione è stata redatta in \LaTeX e pubblicata in formato PDF. A ciascun documento è assegnato un numero di versione $x.y$, ove x rappresenta l'ultima VERSIONE FORMALE, rivista e approvata dal referente aziendale e disponibile a terze parti interessate (membri del team di progetto, tutor interno), mentre il numero y si riferisce ad una VERSIONE PRELIMINARE per uso interno, eventualmente consultabile dal referente aziendale.

Modello relazionale Il modello relazionale del database è stato realizzato mediante lo strumento adottato dal team di progetto, ossia l'editor *MySQL Workbench*, per facilitare la condivisione e l'integrazione delle informazioni.

Casi d'uso La notazione utilizzata per identificare un caso d'uso è così definita:

$$UC.x.y$$

ove:

- UC è l'abbreviazione di *Use Case* (Caso d'uso);
- $x \in \{1, 2, \dots\}$ è il numero identificativo del diagramma cui appartiene il caso d'uso;
- $y \in \{1, 2, \dots\}$ è il numero associato al caso d'uso.

Requisiti funzionali I requisiti del sistema software sono univocamente identificati mediante la seguente notazione:

$$Rf.x.y$$

ove:

- Rf è l'abbreviazione di *requisito funzionale*;
- $x \in \{ob, de\}$ rappresenta il tipo di requisito funzionale (*ob* per obbligatorio, *de* per desiderabili);
- $y \in \{1, 2, \dots\}$ è il numero associato ad un requisito.

Tracciamento dei casi d'uso Il tracciamento delle dipendenze tra casi d'uso e requisiti software è stato realizzato mediante un foglio elettronico, ove:

- ciascuna riga rappresenta un requisito del sistema software;
- ciascuna colonna rappresenta un caso d'uso;
- ciascuna cella contiene il carattere 'X' se esiste una relazione di dipendenza tra il caso d'uso e il requisito, altrimenti è vuota.

Per ciascuna riga e colonna viene impiegata una semplice formula per asserire la completezza e la necessità della matrice dei requisiti:

$$\text{CONTA.SE}(A:Z; "X")$$

ove:

- $A:Z$ corrisponde all'intervallo di celle di una singola riga o colonna;
- X rappresenta il pattern da cercare;
- CONTA.SE è una funzione che accetta due argomenti (l'intervallo di celle ed il *pattern*) e restituisce il numero di celle appartenenti all'intervallo contenenti una o più occorrenze del *pattern* specificato.

Completezza Per ogni colonna, se la formula restituisce un valore pari a 0 (zero) sta ad indicare che il requisito utente non è soddisfatto da alcun requisito software.

Necessità Per ogni riga, se la formula restituisce un valore pari a 0 (zero) sta ad indicare che il requisito software corrispondente è superfluo.

3.3. Criterio di classificazione

Il patrimonio di conoscenza della piattaforma è garantito essenzialmente dai contenuti pubblicati dagli utenti ed arricchito dal loro valore informativo: ciascuno di essi, a prescindere dalla forma (testo, immagini, audio, video, ...) o dalla classe (domanda, discorso, evento, recensione, ...), condivide delle informazioni inerenti uno o più elementi del dominio tematico della piattaforma.



Figura 3.2.: Contenuti informativi e conoscenza

Contenuti informativi Allo state attuale, la piattaforma si limita ad essere un serbatoio di CONTENUTI INFORMATIVI disaggregati, priva degli strumenti per classificare e catalogare il sapere custodito conferendovi una struttura ordinata, una sorta di indice enciclopedico in grado di facilitarne la ricerca, il reperimento e la consultazione.

Ciascun contenuto rappresenta - dal punto di vista conoscitivo - una collezione di frammenti di informazioni, ciascuno dei quali contribuisce ad arricchire la conoscenza relativa a qualche ENTITÀ, oggetto di discussione all'interno del dominio della piattaforma.

Dominio conoscitivo L'insieme di entità definite - in un certo istante - all'interno della piattaforma ne costituisce il DOMINIO DELLA CONOSCENZA (di seguito per brevità DOMINIO), in maniera analoga a quanto accade con i lemmi di un'enciclopedia. Ciascun frammento di informazione presente in un contenuto è concettualmente associabile e riferibile ad una o più entità del dominio.



Figura 3.3.: Valore informativo di un contenuto

L'obiettivo primario del nuovo criterio di classificazione consiste dunque nel modellare tale dominio e le relazioni esistenti tra le entità ed i contenuti informativi.



Figura 3.4.: Dominio di conoscenza della piattaforma

3.3.1. Entità

Le ENTITÀ $d_i \in D$ del dominio rappresentano elementi concreti (luoghi, persone, eventi, ...) o astratti (concetti, ...) a cui afferiscono i contenuti. Ciascuna di esse costituisce metaforicamente un lemma dell'enciclopedia del sapere disponibile presso la piattaforma e - pur avendo un preciso valore semantico - dev'essere identificata sul piano sintattico mediante un'ETICHETTA $e_j \in E$.

Ambiguità sintattica Gli utenti possono tuttavia riferirsi ad un'entità d_i con termini o espressioni differenti ($e_{i,j} \in E_i$): tale ambiguità non può essere ignorata o trascurata e impone di considerare la relazione tra l'entità e le etichette con cui può essere riferita di tipo uno-a-molti. Il criterio di classificazione dev'essere quindi in grado di esprimere il fatto che tali etichette rappresentino sinonimi di una stessa entità.



Figura 3.5.: Ambiguità sintattica di un'entità

Etichette duplicate Allo stesso tempo occorre impedire la proliferazione di etichette duplicate, ossia equivalenti sul piano semantico ma sintatticamente differenti. Un fenomeno simile avrebbe inevitabili ripercussioni sull'efficacia del criterio di classificazione e sull'efficienza della ricerca: reperire tutte e sole le informazioni inerenti una certa entità richiederebbe infatti di individuare tutte le etichette con cui possa essere riferita e cercare riscontri per ciascuna di esse nei contenuti pubblicati.

Per incrementare l'efficienza di catalogazione e ricerca dei contenuti risulta conveniente che l'entità sia identificata univocamente nei contenuti informativi, a prescindere dalla specifica etichetta utilizzata.

Sintassi e semantica Il fattore essenziale consiste nel mantenere separata la componente semantica (il dominio delle entità) da quella sintattica (il dizionario delle etichette): solo così è possibile stabilire un'associazione naturale e diretta tra i contenuti e le entità, a prescindere dalle etichette.



Figura 3.6.: Sintassi e semantica di un'entità

Ciò non toglie la necessità di individuare un'etichetta, che permetta di esprimere e comunicare l'entità cui si fa riferimento: per soddisfare tale condizione si individua - per ciascuna entità - un'ETICHETTA PRIMARIA $e_{i,0}$, che la identifica univocamente nell'ambito

della piattaforma, mentre le restanti (ETICHETTE SECONDARIE) ne vengono considerate sinonimi.

Relazioni Riprendendo la metafora enciclopedica, ciascun lemma contiene spesso riferimenti ad altre voci, che trattano temi specifici o attinenti: ci si aspetta che tali relazioni siano replicabili anche nel dominio della piattaforma sotto forma di legami multi-a-molti tra le entità.



Figura 3.7.: Relazioni tra entità

A questo punto il dominio può essere interpretato come un grafo orientato ove:

- ciascun nodo rappresenta un'entità, identificata dalla relativa etichetta primaria;
- ciascun arco uscente identifica un'entità riferita;
- ciascun arco uscente identifica un'entità referente.

3.3.2. Etichette

Un'ETICHETTA $e_j \in E$ rappresenta la forma sintattica - una stringa di lunghezza variabile - mediante la quale gli utenti ed il sistema identificano un'entità del dominio. L'insieme di etichette definite in un certo istante costituisce il DIZIONARIO della piattaforma.

Accezioni Analogamente ad un lemma enciclopedico, un'etichetta può risultare semanticamente ambigua ed essere caratterizzata da svariate ACCEZIONI $a_{j,k} \in A_j$, ciascuna delle quali assume un significato e si riferisce ad un'entità distinta. Da questa prospettiva, la natura tematica della piattaforma dovrebbe contribuire a limitare il numero medio di accezioni per ciascuna etichetta.

L'etichetta e_j identifica dunque entità del dominio distinte, a seconda dell'accezione $a_{j,k}$ considerata: la distinzione tra etichetta primaria e secondaria si trasferisce dunque alle accezioni, che si distinguono in CHIAVE o SINONIMICA.



Figura 3.8.: Accezioni di un'etichetta

Accezione chiave

L'accezione chiave $a_{j,0}$ indica che la relativa etichetta e_j identifica univocamente l'entità corrispondente: ciò significa che - in ogni istante - a ciascuna entità dev'essere associata una e una sola etichetta primaria. Qualora si rimuova un'accezione chiave di un'etichetta, occorre individuare una nuova etichetta primaria per l'entità riferita.

Accezione sinonimica

L'accezione sinonimica $(a_{j,1}, \dots, a_{j,|A_j|})$ indica che l'etichetta associata e_j rappresenta un sinonimo dell'entità corrispondente. La relazione uno-a-molti tra entità ed etichette si traduce dunque in:

- una relazione uno-a-molti tra le etichette e le accezioni:
- una relazione uno-a-molti tra le entità e le accezioni:

Sinonimi Sebbene le entità siano identificate univocamente dalle corrispondenti etichette primarie, è assai utile includere e conservare nel dizionario anche i sinonimi (etichette secondarie) note o utilizzate dagli utenti al fine di garantire maggiore copertura sintattica, aumentando così la probabilità che i termini o le espressioni utilizzati successivamente per riferire una certa entità siano già presenti nel dizionario e quindi immediatamente riconoscibili.

Ricerca L'utente alla ricerca di informazioni riguardanti un particolare tema cerca in genere le etichette che presentino maggior attinenza. Per evitare la proliferazione di etichette semanticamente identiche e sintatticamente simili, si fissa uno standard per il formato, che normi la capitalizzazione delle lettere, la gestione degli spazi, ... al fine di uniformare la struttura sintattica delle etichette.

L'osservanza e l'adesione a tali regole da parte delle etichette inserite viene accertata automaticamente, provvedendo - ove necessario - ad apportare le opportune correzioni per

adeguarle allo standard, e rappresenta un requisito essenziale per garantire la consistenza del dizionario della piattaforma.

3.3.3. Contenuti

L'obiettivo primario del criterio di classificazione consiste nel tenere traccia delle entità riferite all'interno di ciascun contenuto per rendere facilmente reperibili le informazioni che le riguardano.

Valore informativo Ciascun contenuto reca con sé informazioni riguardanti una o più entità: ne consegue che la relazione tra entità e contenuti sia di tipo molti-a-molti ($n:m$). Preservare la distinzione tra entità (semantica) ed etichette (sintassi) consente di identificare in maniera chiara e univoca nell'intera piattaforma ciò a cui si fa riferimento nei vari contenuti, a prescindere dai termini o dalle espressioni linguistiche impiegati per indicarli.

Catalogazione L'assegnazione di un'etichetta ad un contenuto è un processo che traduce un ingresso sintattico dell'utente (l'etichetta) in un'uscita semantica (l'entità), che viene effettivamente e concretamente assegnata al contenuto.

Sebbene si sia scelto di riferire - nei contenuti pubblicati - ciascuna entità sempre e solo con l'etichetta primaria, per ragioni di semplicità ed efficienza (l'etichetta primaria è un'informazione strettamente correlata all'entità), nulla impedirebbe di assegnare ai contenuti qualsiasi etichetta scelta dagli utenti. Ciò comporterebbe tuttavia la necessità di associare al contenuto non solo l'entità, ma anche - tra le possibili - l'etichetta scelta per indicarla.

3.3.4. Utente

Le scelte progettuali descritte nelle sezioni precedenti sono state valutate tenendo presenti alcuni casi d'uso inerenti l'esplorazione dei contenuti pubblicati nella piattaforma da parte dell'utente, arrivando a delinearne alcuni di particolare interesse o rilevanza. La metafora dell'enciclopedia, che accosta il nuovo criterio di classificazione ad un indice enciclopedico, esprime in maniera assai accurata e favorisce la comprensione delle dinamiche e degli obiettivi con cui gli utenti consultano il sapere della piattaforma.

Ricerca per temi La ricerca per temi consiste essenzialmente in un'esplorazione dei contenuti informativi a partire dalle entità, che sono assimilabili agli argomenti di discussione, in maniera analoga alla consultazione di un'enciclopedia, quando si procede da un lemma all'altro, facendosi guidare dai riferimenti tra le voci sino a trovare quella di interesse. Nella piattaforma ciò si esprime in un'esplorazione libera del grafo orientato delle entità, ciascuna identificata dalla relativa etichetta primaria.

Ricerca per etichette La classica strategia di ricerca prevede l’inserimento di alcune parole chiave, che agli occhi dell’utente identificano l’informazione cui è interessato e hanno maggiore probabilità di essere associate o comparire nei contenuti.

Nella piattaforma, una volta riconosciuta un’etichetta in un termine inserito, occorre identificare il valore semantico (l’entità) attribuitole dall’utente, inevitabilmente associato ad un’accezione dell’etichetta stessa, così da giungere ad identificare univocamente l’entità cercata.

Ove la ricerca si soffermi a considerare l’etichetta, i risultati possono risultare parziali, includendo i soli contenuti in cui l’entità sia riferita dalla specifica etichetta, o addirittura non pertinenti, includendo contenuti ove l’etichetta è presente ma in un’accezione differente.

Un ulteriore aspetto da considerare riguarda la dimensione del dizionario E e del dominio D : sapendo che ciascuna accezione rappresenta una coppia $\langle \text{etichetta}, \text{entità} \rangle$ univoca, possiamo esprimere il numero medio di etichette associate a ciascuna entità come:

$$\alpha = \frac{\sum |A_j|}{|D|} \quad (3.1)$$

Assegnando ai contenuti un’etichetta qualsiasi si aumenterebbe di una costante moltiplicativa α la complessità dell’operazione ricerca, dovendola ripetere per ciascuna delle α etichette anziché per la sola entità corrispondente.¹

La rilevanza di un contenuto informativo rispetto ai criteri di ricerca è infine determinata dal grado di corrispondenza rispetto alle entità cercate: maggiore è il numero di riscontri, maggiore sarà approssimativamente la rilevanza attribuita nel contesto della ricerca.

Se E_s è l’insieme delle etichette cercate e E_c l’insieme delle etichette assegnate a ciascun contenuto si possono distinguere tre casi principali:

Corrispondenza completa: $E_s \subseteq E_c$

Al contenuto risultano assegnate tutte le entità cercate (massima attinenza).

Corrispondenza parziale: $E_s \cap E_c \neq \emptyset$

Al contenuto risulta assegnata parte delle entità cercate (media attinenza).

Nessuna corrispondenza: $E_s \cap E_c = \emptyset$

Al contenuto non risulta assegnata alcuna entità cercata (attinenza nulla).

Ricerca per affinità Il terzo approccio di ricerca consiste nell’identificare i contenuti attinenti ad uno dato, ossia aventi il maggior numero di entità in comune. Valgono

¹Si assuma per semplicità che la ricerca verifichi - per ciascun contenuto - quali termini cercati (entità o etichette) siano presenti, uno per volta, e che la complessità computazionale sia equivalente in entrambi i casi.

per questo scenario le medesime considerazioni fatte nella sezione precedente, previa sostituzione di E_s con l'insieme delle etichette assegnate al contenuto corrente.

3.3.5. Modello relazionale

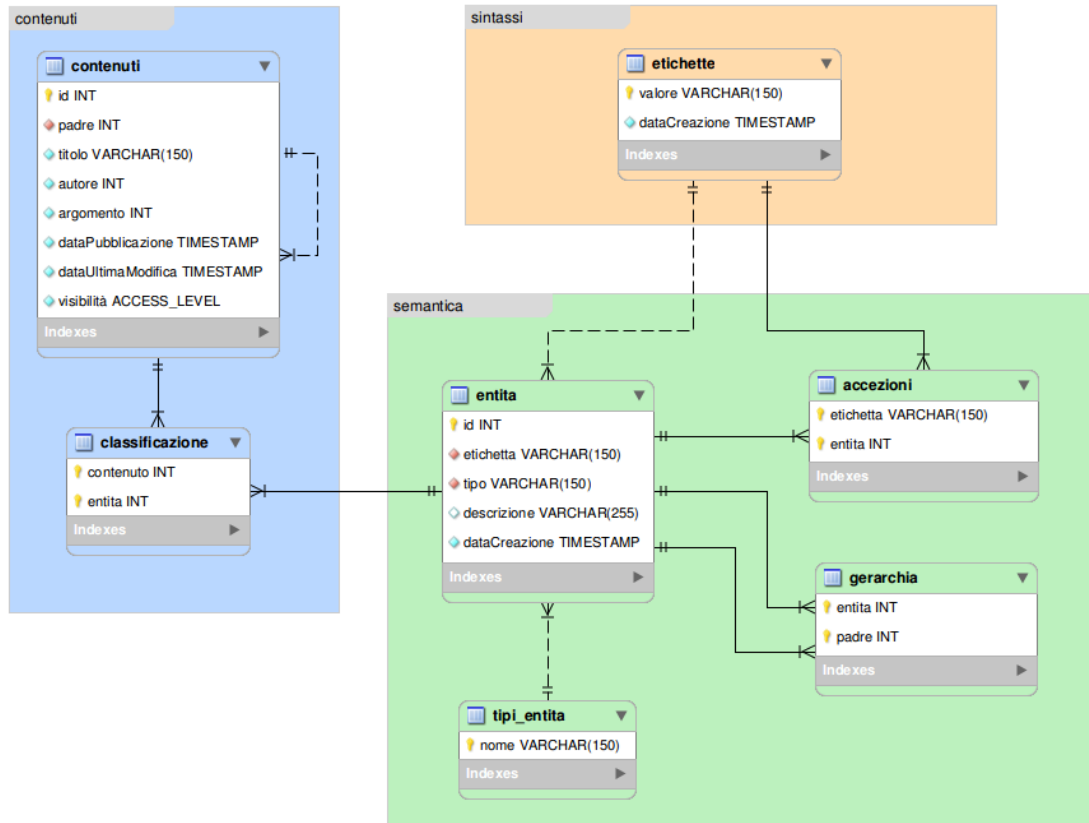


Figura 3.9.: Modello relazionale del criterio di classificazione

Al termine della fase progettuale, si rende necessario aggiungere al modello relazionale della piattaforma le opportune modifiche per integrare le informazioni aggiuntive legate al nuovo criterio di classificazione.

Rispetto all'immagine 3.3.5, la sola tabella **contenuti** risulta importata dal modello relazionale della piattaforma per evidenziare alcune relazioni fondamentali. Le rimanenti sono organizzate in tre *layer* (**contenuti**, **semantica** e **sintassi**) per chiarirne il ruolo all'interno del sistema di classificazione.

Entità Le entità vengono rappresentate mediante la tabella **entita** e si caratterizzano per due vincoli referenziali:

tipo Chiave esterna verso la tabella **tipi_entita**, indica se si tratti di luogo, evento, persona, concetto astratto,

etichetta Chiave esterna verso la tabella **etichette**, rappresenta l'etichetta primaria associata all'entità.

La relazione di tipo molti-a-molti tra le entità si traduce nella tabella **gerarchia**, che contiene due chiavi esterne verso la medesima tabella **entita** a rappresentare rispettivamente le entità riferite e quelle referenti.

Etichette Le etichette sono rappresentate mediante la tabella **etichette** e sono identificate univocamente dalla stringa associata, che rappresenta l'attributo **etichette.valore**. Tale scelta risponde alla naturale identificazione dell'etichetta nella sequenza di caratteri corrispondente, costituisce una garanzia contro la presenza di duplicati e consente di recuperare il valore dell'etichetta primaria di un'entità senza dover effettuare un'operazione di *join* tra le tabelle **entita** ed **etichette**.

Accezioni Le accezioni rappresentano un legame univoco tra le etichette e le entità e possono essere di tipo chiave o sinonimico. Esse non presentano attributi propri significativi, ma prevedono tre vincoli referenziali:

1. a ciascuna entità è associata una e una sola ACCEZIONE CHIAVE (relazione uno-a-uno), che identifica l'etichetta primaria;
2. a ciascuna entità sono associate $0 \dots n$ ACCEZIONI SINONIMICHE (relazione uno-a-molti), che rappresentano i sinonimi dell'etichetta primaria;
3. ciascuna etichetta possiede $0 \dots n$ ACCEZIONI (relazione uno-a-molti).

Ne consegue che sia la superclasse (accezioni) sia le sottoclassi (**accezioni_chiave** e **accezioni_sinonimiche**) presentano dei vincoli referenziali; in particolare, la presenza delle prime due relazioni costringe a distinguere - dal punto di vista dell'entità - tra accezioni chiave e sinonimiche.

Per modellare tale scenario vengono presi in considerazione tre possibili approcci:

Tabella unica

La tabella unica ben si adatta a gestire l'assenza di attributi propri per le sottoclassi e ad esprimere i vincoli referenziali che coinvolgono la superclasse, ma non è in grado di esprimere e adeguatamente rappresentare quelli coinvolgenti le sottoclassi.



Figura 3.10.: Modello ad oggetti delle accezioni

Partizionamento orizzontale

Il partizionamento orizzontale riesce a modellare i vincoli referenziali delle sottoclassi, ma non quello della superclasse, e genera due classi aventi i medesimi attributi.

Partizionamento verticale

Il partizionamento verticale consente di modellare correttamente tutti e tre i vincoli referenziali, relativi sia alla superclasse sia alle sottoclassi. Tuttavia si rende più complesso modificare il tipo di un'accezione e si introduce l'esigenza di un'operazione *join* per recuperare la lista completa delle accezioni, pur non possedendo le sottoclassi attributi propri.

A seguito di alcune osservazioni si decide di adottare la soluzione della tabella unica:

- la distinzione tra etichette chiave e sinomimiche ha rilevanza essenzialmente dal punto di vista della classe **entita**;
- il vincolo referenziale tra **etichette** ed **accezioni** suggerisce che la distinzione di cui al punto precedente sia irrilevante dal punto di vista delle etichette, ragion per cui risulta utile mantenere tutte le accezioni nella medesima tabella.²
- le sottoclassi non hanno attributi propri, per cui il partizionamento verticale e orizzontale sono da ritenersi soluzioni inadeguate o carenti.

Il soddisfacimento delle condizioni richieste viene raggiunto eliminando qualsiasi riferimento al tipo dell'accezione nella classe **accezioni** e modellando la relazione uno-a-uno tra le entita e le relative etichette primarie mediante un vincolo referenziale di chiave esterna nella classe **entita**, ossia **entita.etichetta**, che identifica la corrispondente etichetta primaria nella tabella **etichette**. Così facendo si riescono ad esprimere tutti i vincoli referenziali senza dover definire le sottoclassi.

²Si consideri ad esempio il caso d'uso della ricerca di un'entità a partire da un'etichetta, ove occorre recuperare la lista completa delle relative accezioni.

Contenuti Il vincolo referenziale tra i contenuti e le entità rappresenta l'essenza del criterio di classificazione, consentendo di classificare e catalogare le informazioni disponibili nella piattaforma. La relazione molti-a-molti tra **contenuti** ed **entità** viene modellata aggiungendo la nuova classe **classificazione**, che tiene precisamente traccia delle entità associate a ciascun contenuto.

3.4. Fase 2: interfaccia grafica

La seconda fase dell'attività di stage consiste nell'analisi e nella progettazione di un'interfaccia grafica per la consultazione dei risultati di ricerche relative ai contenuti informativi pubblicati dagli utenti nella piattaforma. Le principali qualità attese sono:

Intuitiva

L'interfaccia grafica deve risultare agevole e facilmente utilizzabile da qualsiasi categoria di utenti, a prescindere dal livello di esperienza e dalla familiarità con piattaforme web esistenti (*chat*, *forum*, *social network*, ...). Ciò richiede una particolare attenzione alle esigenze ed aspettative dell'utente generico della piattaforma, evidenziate da indagini di carattere sociologico svolte da altri membri del team di progetto.

Multipiattaforma

L'interfaccia grafica dev'essere fruibile dal maggior numero possibile di dispositivi (*computer*, *tablet*, *smartphone*), tenendo ben presente le differenti modalità di interazione.

Ordinata

L'interfaccia grafica dev'essere in grado di rappresentare in maniera ordinata ed efficace le informazioni, a prescindere dal numero di contenuti caricati: quelle essenziali devono essere immediatamente disponibili e facilmente identificabili, mentre quelle aggiuntive o accessorie devono essere comodamente accessibili.

Tale interfaccia dev'essere in grado di:

- rappresentare - in forma grafica o testuale - le proprietà fondamentali di ciascun contenuto (...), i metadati di classificazione (emozioni, giudizi, intenzioni, ...) e i legami reciproci;
- raffinare i risultati di ricerca secondo alcune proprietà (autore, data di pubblicazione, ...) o parametri di classificazione (argomento, emozione, etichetta, giudizio, intenzione) dei contenuti;
- mostrare la discussione generata da un contenuto informativo, incluso tra i risultati di ricerca iniziali.

Per semplificare la fase di analisi e progettazione il modello concettuale del prodotto software atteso è stato suddiviso in quattro componenti (di seguito anche **LIVELLI**):

Ricerca

Insieme dei parametri e criteri di ricerca (parole chiave, ambito di ricerca, ...).

Filtri

Insieme dei filtri di raffinamento della ricerca.

Contenuti

Insieme dei contenuti informativi corrispondenti ai criteri di ricerca.

Navigazione

Interfaccia di navigazione dei risultati di ricerca.

3.4.1. Casi d'uso

Attori L'individuazione, la classificazione e l'analisi dei casi d'uso inizia dal riconoscimento di due attori principali, ossia l'**UTENTE AUTENTICATO** ed il visitatore (di seguito semplicemente **UTENTE**). L'utente autenticato rappresenta un tipo specializzato del visitatore, dal momento che alcune informazioni di profilo, come gli interessi dichiarati o il livello di esperienza, vengono sfruttate per filtrare automaticamente i risultati di ricerca: a lui solo, dunque, è offerta la possibilità di abilitare o meno tali filtri personalizzati.

Obiettivi Le finalità d'interazione dell'utente con l'interfaccia grafica sono in buona parte riconducibili alla scomposizione concettuale del sistema, illustrata nella sezione precedente:

1. impostare i parametri iniziali di ricerca o modificarli a posteriori;
2. filtrare i risultati di ricerca in accordo a criteri di classificazione o proprietà dei contenuti;
3. individuare i contenuti di interesse e consultare le informazioni associate;
4. impostare i filtri personalizzati (solo per utenti autenticati).

UC.1 - Ricerca di contenuti informativi

La ricerca di contenuti informativi prevede l'inserimento dei termini di ricerca (*UC.1.1*), separati da virgola, e la restrizione dell'ambito alle sole entità associate o alle informazioni presenti in un contenuto (titolo, corpo, ...) (*UC.1.2*).

Durante la ricerca, qualora vengano individuate delle corrispondenze tra i termini cercati e le etichette del dizionario e queste presentino accezioni multiple, l'utente è

chiamato a selezionare quella corrispondente all'entità rispetto alla quale intende svolgere la ricerca (*UC.1.3*).

Al termine della ricerca, l'utente può consultare la lista delle entità individuate a partire dai criteri di ricerca specificati (*UC.1.4*) e intervenire sui risultati di ricerca eliminando una di esse (*UC.1.5*) o sostituendola con un'altra, avente un legame diretto (*UC.1.8*, *UC.1.9*). Per sostituire un'entità con un'altra l'attore dev'essere in grado di visualizzare - per ciascuna - quelle che la riferiscono (*UC.1.6*) o da lei riferite (*UC.1.7*).

UC.2 - Raffinamento dei criteri di ricerca

Per facilitare l'individuazione dei contenuti di interesse tra i risultati della ricerca, l'utente dev'essere messo in condizione di filtrarli ...

Parametri Il primo passo consiste nell'individuare le proprietà dei contenuti informativi aventi particolare rilevanza e definite su un insieme U di valori adeguato ad essere associato ad un filtro: il solo requisito è la possibilità di partizionare l'insieme U in due sottoinsiemi, che rappresentano rispettivamente i valori autorizzati (U_a) e bloccati (U_b) in un certo istante.

La configurazione predefinita di ciascun filtro prevede che tutti i valori siano ammissibili, ossia $U_a = U$ e $U_b = \emptyset$. L'azione dell'utente consiste essenzialmente nell'alterare tale partizionamento dell'insieme U in tre modi principali:

- bloccando un valore u ($u \in U_b$);
- autorizzando un valore ($u \in U_a$);
- azzerando il filtro, ossia ripristinando la configurazione iniziale e autorizzando tutti i valori ($U = U_a$).

Considerata una generica proprietà,

Classi di filtri Il secondo passo

- quali criteri di classificazione? quali proprietà ?
- classificazione

UC.2 - Consultazione dei risultati di ricerca - entità rappresentate mediante etichetta primaria

UC.2 - Gestione dei filtri utente

3.4.2. Requisiti

3.4.3. Progettazione

4. Valutazioni finali

4.1. Consuntivo

4.2. Competenze professionali

- multidisciplinarietà: coinvolgimento di figure professionali provenienti da numerosi e variegati settori professionali (psicologia, sociologia, marketing, economia, informatica, ingegneria, ...)

4.3. Stage e università

A. Glossario

B

Bitbucket - <https://bitbucket.org/>

Piattaforma web per la gestione delle attività di progetto con supporto a strumenti di controllo di versione distribuito.

C

CamelCase

Convenzione per la scrittura di espressioni composte unendo le parole tra loro e mantenendo ciascuna iniziale in maiuscolo.

Chilometri zero

Filosofia di consumo basata sull'acquisto di beni agroalimentari direttamente dal produttore, evitando la filiera, tutelando l'ambiente e valorizzando la produzione locale del territorio.

G

gedit - <http://projects.gnome.org/gedit/>

Editor di testo ufficiale dell'ambiente desktop GNOME.

Gruppi di acquisto

Insieme (stabile o provvisorio) di consumatori che acquista mediante ordine collettivo un consistente numero di beni direttamente dal produttore, spesso per conseguire prezzi vantaggiosi o ammortizzare eventuali spese accessorie.

L

LaTeXila - <http://projects.gnome.org/latexila/>

Editor LaTeX integrato per l'ambiente desktop GNOME.

LibreOffice Calc - <http://www.libreoffice.org/>

Applicazione per fogli di calcolo della suite di produttività *LibreOffice*.

M

Mercurial - <http://mercurial.selenic.com/>

Strumento multi piattaforma, gratuito ed open source per il controllo di versione distribuito.

MySQL Workbench - <http://www.mysql.it/products/workbench/>

Applicazione multi piattaforma, gratuita ed open source per la progettazione, lo sviluppo e l'amministrazione di database MySQL.

P

PDF (Portable Document Format)

Formato di file per la rappresentazione di documenti in maniera indipendente dalla piattaforma hardware e software.

Pencil - <http://pencil.evolus.vn/>

Applicazione multi piattaforma, gratuita ed open source per la realizzazione di prototipi di interfacce grafiche.

ProjectLibre - <http://sourceforge.net/projects/projectlibre/>

Applicazione multi piattaforma, gratuita ed open source per il *project management*, che consente di realizzare diagrammi di Gantt e di PERT, gestire le risorse allocate e le attività pianificate,

R

Rete sociale

Insieme di persone, aventi interessi in comune e inclini a collaborare e condividere idee o informazioni, e di relazioni di tipo esperienziale definite tra tali soggetti.

U

Ubuntu - <http://www.ubuntu.com/>

Distribuzione Linux gratuita derivata da Debian.

UML (Unified Modelling Language) - <http://www.uml.org/>

Standard internazionale per un linguaggio di modellazione, che definisce un insieme di notazioni grafiche per la rappresentazione visiva di sistemi.

UMLet - <http://www.umlet.com/>

Applicazione multi piattaforma, gratuita ed open source per la realizzazione di diagrammi UML.

B. Criterio di classificazione

Legenda

Entità

D

Dominio delle entità.

d_i

Entità del dominio ($d_i \in D$).

$1 \leq i \leq |D|$, $i \in \mathbb{N}$.

Etichette

E

Dizionario delle etichette.

E_i

Insieme delle etichette relative all'entità d_i ($E_i \subset E$).

e_j

Etichetta del dizionario ($e_j \in E$).

$1 \leq j \leq |E|$, $j \in \mathbb{N}$.

$e_{i,j}$

Etichetta del dizionario relativa all'entità d_i ($e_{i,j} \in E$).

$e_{i,0}$

Etichetta primaria dell'entità d_i ($e_{i,0} \in E$).

Accezioni

A_j

Insieme delle accezioni di un'etichetta e_j .

$a_{j,k}$

Accezione di un'etichetta e_j ($a_{j,k} \in A_j$).

$1 \leq k \leq |A_j|$, $j \in \mathbb{N}$.

Codice SQL

Etichette

Ricerca di un'etichetta Restituisce le stringhe corrispondenti alla sequenza di caratteri @termine inserita dall'utente, eventualmente adeguata allo standard di formato previsto.

```
SELECT *
  FROM etichette
 WHERE valore LIKE '@termine%'
```

Ricerca delle accezioni di un'etichetta Restituisce la lista delle accezioni di un'etichetta @etichetta.

```
SELECT *
  FROM accezioni
 WHERE etichetta='@etichetta'
```

Numero di assegnazioni di un'entità Restituisce il numero di contenuti cui sia stata assegnata l'entità @entità.

```
SELECT COUNT(*) AS num
  FROM classificazione
 WHERE entita='@entita'
```

Inserimento di un'etichetta Inserisce l'etichetta @stringa...

```
INSERT INTO etichette
  VALUES (@stringa,current_timestamp)
```

... e le associa almeno un'accezione, riferita all'entità @entità.

```
INSERT INTO accezioni(etichetta,entita)
  VALUES (@etichetta,@entita)
```

Entità

Calcolo di affinità tra due entità L'affinità tra due entità è espressa dal numero di contenuti in cui compaiano entrambe.

```
SELECT COUNT(*) AS contenuti
  FROM classificazione
 WHERE entita='@entita1' OR entita='@entita2'
GROUP BY contenuto
  HAVING COUNT(*)>=2
```

Ricerca delle entità referenti Restituisce la lista delle entità che riferiscono @entita.

```
SELECT padre
  FROM gerarchia
 WHERE figlio=@entita'
```

Ricerca delle entità riferite Restituisce la lista delle entità riferite da @entita.

```
SELECT figlio
  FROM gerarchia
 WHERE padre=@entita'
```

Ricerca delle etichette Restituisce la lista delle etichette con cui è riferibile l'entità @entita.

```
SELECT et.valore AS etichetta
  FROM accezioni AS acc JOIN etichette AS et
    ON (acc.etichetta=et.valore)
 WHERE acc.entita=@entita'
```

Contenuti

Assegnazione di un'etichetta Assegna l'entità @entita al contenuto @contenuto.

```
INSERT INTO classificazione(contenuto,entita)
  VALUES (@contenuto,@entita)
```

Ricerca delle etichette assegnate Restituisce la lista delle entità assegnate ad un contenuto.

```
SELECT entita
  FROM classificazione
 WHERE contenuto=@contenuto'
```

Ricerca di contenuti generici Restituisce i contenuti cui sia stata assegnata l'entità @entità.

```
SELECT contenuti
  FROM classificazione
 WHERE entita=@entita'
```

Ricerca di contenuti specifici Restituisce i contenuti cui siano state assegnate le entità @entita1, @entita2,

```
SELECT contenuto
  FROM classificazione
 WHERE entita='@entita1' [OR entita='@entita2' ...]
GROUP BY contenuto
HAVING COUNT(*)>=@num_etichette
```

C. Interfaccia grafica

Legenda

$$U = U_a \cup U_b$$

Insieme dei possibili valori di una proprietà.

$$U_a \subseteq U$$

Insieme dei valori autorizzati di una proprietà ($U_a \cap U_b \neq \emptyset$).

$$U_b \subseteq U$$

Insieme dei valori bloccati di una proprietà. ($U_a \cap U_b \neq \emptyset$)

Casi d'uso

Requisiti

Architettura