Immagine che contiene scatola

Descrizione generata automaticamente

Università degli Studi di Salerno

Immagine che contiene testo, clipart

Descrizione generata automaticamente

Dipartimento di Ingegneria dell’Informazione ed Elettrica e Matematica Applicata

Corso di Laurea in Ingegneria Informatica

**Basi di Dati 2021/2022**

**Canale I-Z**

Project Work

**Traccia N. 5 – Sentiment Analysis: la gestione delle Opinion**

Gruppo n. **08 – IZ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **WP** | **Cognome e Nome** | **Matricola** | **e-mail** | **Responsabile** |
| **1** | **Postiglione Matteo** | **0612705214** | **m.postiglione10@studenti.unisa.it** |  |
| **2** | **Relmi Emanuele** | **0612705274** | **e.relmi@studenti.unisa.com** | **X** |
| **3** | **Raffone Riccardo** | **0612705256** | **r.raffone3@studenti.unisa.it** |  |
| **4** | **Quagliuolo Francesco** | **0612705239** | **f.quagliuolo@studenti.unisa.it** |  |

**Anno accademico 2021-2022**

Sommario

[1. Descrizione della realtà di interesse 3](#_Toc71213214)

[2. Analisi delle specifiche 5](#_Toc71213215)

[2.1. Glossario dei termini 5](#_Toc71213216)

[2.2. Strutturazione dei requisiti in frasi 6](#_Toc71213217)

[2.2.1. Frasi di carattere generale **Errore. Il segnalibro non è definito.**](#_Toc71213218)

[2.2.2. Frasi relative a <…> **Errore. Il segnalibro non è definito.**](#_Toc71213219)

[2.2.3. Frasi relative a <…> **Errore. Il segnalibro non è definito.**](#_Toc71213220)

[2.2.4. Frasi relative a tipi specifici di <…> **Errore. Il segnalibro non è definito.**](#_Toc71213221)

[2.3. Identificazione delle operazioni principali 7](#_Toc71213222)

[3. Progettazione Concettuale 8](#_Toc71213223)

[3.1. Schema Concettuale 8](#_Toc71213224)

[3.1.1. Note sullo schema E-R *[opzionale]* 9](#_Toc71213225)

[3.2. Design Pattern 9](#_Toc71213226)

[3.2.1. Pattern Relazione Ternaria [Sostituire con il nome del pattern usato] 9](#_Toc71213227)

[3.2.2. Pattern XXX 9](#_Toc71213228)

[3.3. Dizionario dei Dati 9](#_Toc71213229)

[3.4. Regole Aziendali 12](#_Toc71213230)

[4. Progettazione Logica 13](#_Toc71213231)

[4.1. Ristrutturazione Schema Concettuale 13](#_Toc71213232)

[4.1.1. Analisi delle Prestazioni 13](#_Toc71213233)

[4.2. Analisi delle ridondanze 14](#_Toc71213234)

[4.2.1. Analisi della ridondanza 1: Esami sostenuti 14](#_Toc71213235)

[4.2.2. Analisi della ridondanza 2: … **Errore. Il segnalibro non è definito.**](#_Toc71213236)

[4.3. Eliminazione delle generalizzazioni 15](#_Toc71213237)

[4.3.1. Generalizzazione *Entita1* 15](#_Toc71213238)

[4.3.2. Generalizzazione *Persona* 15](#_Toc71213239)

[4.4. Partizionamento/Accorpamento Entità e Associazioni 15](#_Toc71213240)

[4.5. Scelta degli identificatori principali 16](#_Toc71213241)

[4.6. Schema ristrutturato finale 17](#_Toc71213242)

[4.7. Schema logico 18](#_Toc71213243)

[4.8. Documentazione dello schema logico 18](#_Toc71213244)

[5. Normalizzazione 20](#_Toc71213245)

[6. Script Creazione e Popolamento Database 20](#_Toc71213246)

[7. Query SQL 22](#_Toc71213247)

[7.1. Query con operatore di aggregazione e join: Titolo query 22](#_Toc71213248)

[7.2. Query nidificata complessa: Titolo query 22](#_Toc71213249)

[7.3. Query insiemistica: Titolo query 22](#_Toc71213250)

[7.4. Altre query 22](#_Toc71213251)

[7.4.1. Titolo Query 22](#_Toc71213252)

[7.4.2. Titolo Query 22](#_Toc71213253)

[8. Viste 23](#_Toc71213254)

[8.1. Vista *TitoloVista* 23](#_Toc71213255)

[8.1.1. Query con Vista: Titolo query 23](#_Toc71213256)

[9. Trigger 24](#_Toc71213257)

[9.1. Trigger inizializzazione: *TitoloTrigger* 24](#_Toc71213258)

[9.2. Trigger per vincoli aziendali 24](#_Toc71213259)

[9.2.1. Trigger1: TitoloTrigger 24](#_Toc71213260)

[9.2.2. Trigger2: TitoloTrigger **Errore. Il segnalibro non è definito.**](#_Toc71213261)

# Descrizione della realtà di interesse

Titolo: **Profilazione degli utenti per la gestione delle pubblicità**

Si vuole realizzare una base di dati per una società che fornisce servizi pubblicitari, allo scopo di profilare gli utenti che visitano siti web a carattere giornalistico e di monitorare le loro interazioni con essi, al fine di raccogliere informazioni utili per la profilazione pubblicitaria.

In particolare, si vogliono rappresentare:

* i diversi **utenti**, caratterizzati da un *indirizzo IP*. Se gli utenti sono **registrati**, vogliamo memorizzare anche il *nome*, il *cognome*, l’*età*, l’*e-mail*, lo *username*, il *sesso* e il *numero di commenti*;
* il **sito**, a cui gli utenti accedono, è caratterizzato da un *URL* e dal *nome del sito* scelto, e inoltre memorizza informazioni inerenti all’utente tramite i cookie;
* gli **articoli** vengono identificati da un *link* e un *titolo*, inoltre contengono informazioni riguardo all’ *autore*, al *corpo* e al *timestamp di creazione*, sono classificati tramite dei *tag* e hanno un *tempo di lettura stimato*;
* i **cookies**, che sono in grado di tenere traccia dell’utente, sono caratterizzati da un **ID** univoco, da un *nome*, una *descrizione* e dalla *data di creazione*. I cookies possono essere specializzati in: **cookies di sessione**, che memorizzano il *sistema operativo*, *preferenze sulla lingua*, *risoluzione dello schermo*, il *browser* e il *dispositivo* utilizzati dall’utente, e **cookies persistenti** che memorizzano *preferenze sul tema* e *geolocalizzazione* e sono caratterizzati da una *data di scadenza*;
* i **commenti** rilasciati dagli utenti sono identificati dalla *data di scrittura*, da un *ranking* e dal *testo* del commento.

## Analisi della realtà di interesse

L'obiettivo del progetto è realizzare un database in grado di profilare gli utenti che accedono al sito specifico, al fine di promuovere materiale pubblicitario mirato, basandosi sulle preferenze tracciate tramite la navigazione.

L’utente che visita il sito, effettuando la registrazione ad esso, è identificato dall’indirizzo IP (IPv4) composto esattamente da 16 cifre. L'indirizzo IP è un identificatore univoco associato al dispositivo utilizzato dall'utente per accedere al sito web. Una volta collegato al sito, quest’ultimo gli assegna un cookie che ha un identificativo univoco. È inoltre possibile, per l’utente, leggere più articoli da diversi siti, ma ha anche la possibilità di leggere l’articolo tramite un link diretto senza accedere al sito.

I cookies per la profilazione dell’utente sono divisi in cookies di sessione e cookies persistenti.   
Riguardo ai primi, essi sono un tipo di cookie, la cui durata è limitata alla sessione di navigazione al termine della quale vengono eliminati. I cookies persistenti, invece, sono utilizzati dai siti web per memorizzare le informazioni sul dispositivo dell'utente per un periodo di tempo più lungo rispetto ai cookies accennati precedentemente, venendo salvati sul disco rigido del dispositivo dell'utente, rimanendo immagazzinati anche dopo la chiusura del browser web.

Un utente che vuole registrarsi deve accettare obbligatoriamente i cookies persistenti, pena l’annullamento della registrazione. In caso di non registrazione, all’utente vengono assegnati solo i cookies di sessione.

Nel caso proposto, poniamo attenzione sulle interazioni che l’utente ha con gli articoli, andando a memorizzare uno o più eventuali commenti rilasciati.

Si specifica che un utente può rilasciare al più un commento solo se ha effettuato il login al sito.

Si precisa che nel database viene tralasciata l’analisi delle singole frasi che compongono il contenuto dei commenti rimanendo ad un livello più elevato (*document level*).

Nel caso in analisi, il visitatore del sito potrebbe non aver creato un account e potrebbe anche aver scelto di non accettare i cookies. Ciò significa che abbiamo solo alcune informazioni sull’utente, in quanto alcuni cookie di sessione vengono utilizzati per personalizzare l'esperienza di navigazione.  
Tuttavia, questo tipo di cookie è automaticamente eliminato quando l'utente chiude il browser o termina la sessione di navigazione. Pertanto, non possiamo essere completamente certi dell'identità dell'utente che sceglie di non accettare i cookie, ma possiamo effettuarne una stima utilizzando le informazioni limitate disponibili.

# Analisi delle specifiche

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP0** | Analisi delle specifiche | Intero Gruppo |

## Glossario dei termini

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Termine** | **Descrizione** | **Sinonimi** | **Collegamenti** |
| **1** | Utente | Persona che utilizza un dispositivo per  navigare sul sito e ricercare gli articoli | Visitatore | Sito, Articolo |
| **2** | Sito | Pagina web visitabile da un utente | - | Articolo, Utente, Cookie |
| **3** | Articolo | Articolo giornalistico leggibile da un utente | - | Sito, Commento |
| **4** | Commento | Commento effettuato da un utente inerente all’articolo o l'argomento da  esso trattato | Opinione, Interazione | Utente registrato, Articolo |
| **5** | Cookie | Piccolo file di testo che viene inviato da un sito web al browser dell'utente e memorizzato sul dispositivo dell'utente | - | Sito |
| **6** | URL | L'URL (Uniform Resource Locator) è una stringa di caratteri che identifica univocamente un indirizzo di risorsa su Internet | Link | Sito, Articolo |
| **7** | Indirizzo IP | Un indirizzo IP (Internet Protocol address) è una serie univoca di numeri che identifica un dispositivo connesso a una rete informatica che utilizza il protocollo Internet (IP). | - | Utente |

Tabella 1. Glossario dei Termini

## Strutturazione dei requisiti in frasi

2.2.1. **Frasi di carattere generale**

Si vuole realizzare una base di dati per una società che fornisce servizi pubblicitari, allo scopo di profilare gli utenti che visitano siti web a carattere giornalistico e di monitorare le loro interazioni con essi, al fine di raccogliere informazioni utili per la profilazione pubblicitaria.

2.2.2. **Frasi relative a Utente**

Si vogliono rappresentare i diversi **utenti**, caratterizzati da un *indirizzo IP*. Se gli utenti sono **registrati**, vogliamo memorizzare anche il *nome*, il *cognome*, l’*età*, l’*e-mail*, lo *username*, il *sesso* e il *numero di commenti*.

2.2.3. **Frasi relative a Sito**

Il **sito**, a cui gli utenti accedono, è caratterizzato da un *URL* e dal *nome del sito* scelto, e inoltre memorizza informazioni inerenti all’utente tramite i cookie.

2.2.4. **Frasi relative ad Articolo**

Gli **articoli** vengono identificati da un *link* e un *titolo*, inoltre contengono informazioni riguardo all’ *autore*, al *corpo* e al *timestamp di creazione*, sono classificati tramite dei *tag* e hanno un *tempo di lettura stimato*.

2.2.5. **Frasi relative ai Cookies**

I **cookies**, che sono in grado di tenere traccia dell’utente, sono caratterizzati da un **ID** univoco, da un *nome*, una *descrizione* e dalla *data di creazione*. I cookies possono essere specializzati in: **cookies di sessione**, che memorizzano il *sistema operativo*, *preferenze sulla lingua*, *risoluzione dello schermo*, il *browser* e il *dispositivo* utilizzati dall’utente, e **cookies persistenti** che memorizzano *preferenze sul tema* e *geolocalizzazione* e sono caratterizzati da una *data di scadenza.*

2.2.6. **Frasi relative a Commento**

I **commenti** rilasciati dagli utenti sono identificati dalla *data di scrittura*, da un *ranking* e dal *testo* del commento.

## Identificazione delle operazioni principali

**Operazione 1:** registrazione di un nuovo utente su uno specifico sito (operazione effettuata 150 volte al giorno)

**Operazione 2**: inserimento di un commento specifico (operazione da effettuare 10 volte al giorno)

**Operazione 3**: controllo degli utenti che hanno accettato i cookie (operazione effettuata 10 volte al giorno)

**Operazione 4**: stampa il numero di commenti scritto da un utente registrato (operazione da effettuare 5 volte ogni giorno)

**Operazione 5**: controllo dell’articolo più letto dagli utenti (operazione effettuata 1 volta a settimana)

**Operazione 6**: controllo del sistema operativo utilizzato dagli utenti cookieless (operazione effettuata 1 volta ogni 2 settimane)

**Operazione 7**: conteggio dei tag che appaiono più frequentemente negli articoli letti (operazione da effettuare mensilmente)

# Progettazione Concettuale

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP1** | Progettazione Concettuale | Postiglione Matteo |

## Schema Concettuale

Immagine che contiene disegno, schizzo, diagramma, Line art

Descrizione generata automaticamente

Figura 1. Schema E-R

### Note sullo schema E-R

Per lo sviluppo dello schema concettuale si è fatto uso della strategia top-down partendo da uno schema iniziale che poi abbiamo raffinato mediante l’aggiunta degli attributi e le relative cardinalità.

### Design Pattern

### Pattern “Parte-di”

Nel contesto della progettazione del database, l'entità **Sito** rappresenta un sito web specifico, mentre l'entità **Articolo** rappresenta un articolo di giornale pubblicato su quel sito. La relazione *parte-di* indica che ogni articolo è una parte integrante del sito su cui viene pubblicato.

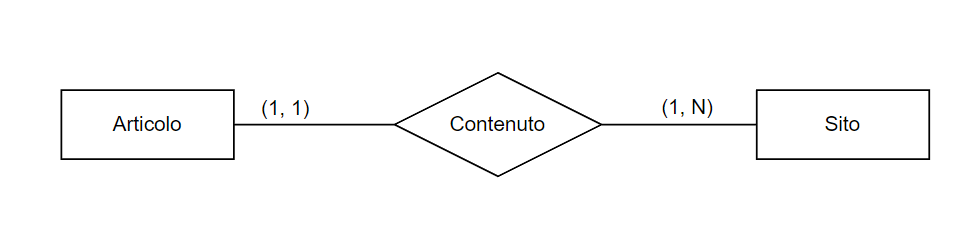
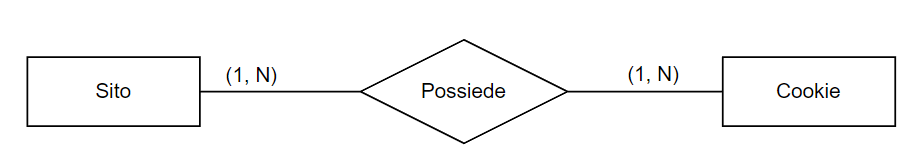


Figura 1. Schema successivo all’applicazione del Pattern PARTE DI

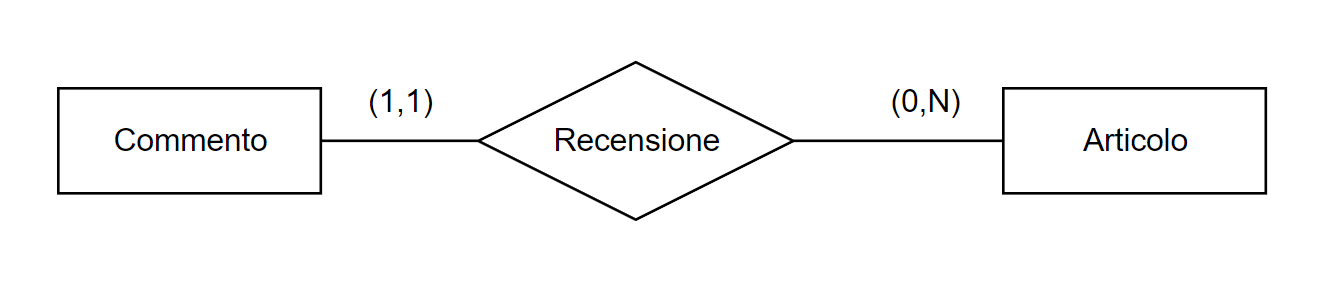
### Pattern “Parte-di”

La relazione *parte-di* in questo caso indica che uno o più **cookies** appartengono ad uno o più **siti**.

Figura 2. Schema successivo all’applicazione del Pattern PARTE DI

### Pattern “Parte-di”

La relazione *parte-di* in questa casistica indica che uno o più **commenti** appartengono ad un solo **articolo**.



*Figura 3. Schema successivo all’applicazione del Pattern PARTE DI*

## Dizionario dei Dati

|  |  |  |  |
| --- | --- | --- | --- |
| **Entità** | **Descrizione** | **Attributi** | **Identificatore** |
| Utente | Persona che utilizza un dispositivo per  navigare sul sito e ricercare gli articoli | IP | IP |
| Utente registrato | Utente che naviga sul sito e/o legge articoli accettando i cookies persistenti | Nome, Cognome, Età, Sesso, E-mail, Username, Numero di commenti | Username |
| Utente cookieless | Utente che naviga sul sito e/o legge articoli profilata solo dai cookies di sessione | - | - |
| Sito | Pagina web visitabile da un utente | URL, Nome | URL |
| Articolo | Articolo giornalistico leggibile da un utente | Link, Titolo, Autore, Keywords, Corpo, Tag, Tempo di lettura stimato, | Link, Titolo |
| Cookie | Piccolo file di testo che viene inviato da un sito web al browser dell'utente e memorizzato sul dispositivo dell'utente | ID, Data di creazione, Nome, Descrizione | ID |
| Cookie di sessione | Cookies utilizzati per profilare gli utenti che declinano i cookie persistenti | Sistema Operativo, Lingua, Schermo, Browser, Dispositivo | - |
| Cookie persistente | Cookies atti a profilare gli utenti registrati, immagazzinandone le informazioni personali | Tema, Geolocalizzazione, Data scadenza | - |
| Commento | Commento effettuato da un utente inerente all’articolo o l'argomento da  esso trattato | Autore, Corpo, Ranking, Data Scrittura | Autore |

Tabella 2. Dizionario dei dati – Entità

|  |  |  |  |
| --- | --- | --- | --- |
| **Relazioni** | **Descrizione** | **Entità Coinvolte** | **Attributi** |
| Accesso | Associa un utente al sito che visita | Utente (1, N), Sito (1, N). | - |
| Possiede | Associa i cookies al sito | Sito (1, N), Cookies (1, N) | - |
| Contenuto | Associa gli articoli al sito visitato | Sito (1, N), Articolo (1,1) | - |
| Assegnazione | Associa l’utente cookieless al cookie di sessione | Utente cookieless (1,1),  Cookie di sessione (1,1) | - |
| Profilazione | Associa l’utente registrato al cookie permanente | Utente Registrato (1, N),  Cookie Persistente (1, N) | - |
| Scrittura | Associa il commento all’utente registrato | Commento (1,1),  Utente Registrato (0, N) | - |
| Recensione | Associa il commento all’articolo | Commento (0,1), Articolo (0, N) | - |
| Lettura | Associa ogni utente agli articoli | Utente (1, N), Articolo (1, N) | Tempo di Lettura Effettivo |

Tabella 3. Dizionario dei dati - Relazioni

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP4** | Regole Aziendali | Quagliuolo Francesco |

## Regole Aziendali

|  |
| --- |
| **Regole di Vincolo** |
| **(RV1)** Un utente non registrato non può rilasciare un commento all’articolo letto.  **(RV2)** Un utente che vuole registrarsi deve accettare i cookies persistenti  **(RV3)** Un utente non può rilasciare più di un commento ad un articolo |

Tabella 4. Regole di vincolo

|  |
| --- |
| **Regole di derivazione** |
| **(RD1)** Il tempo di lettura stimato si ottiene facendo una media dei tempi effettivi impiegati dagli utenti.  **(RD2)** Il numero dei commenti scritti da un Utente Registrato è calcolabile sfruttando l’accesso alla relazione Scrittura collegata all’entità Commento, contandone le occorrenze. |

Tabella 5. Regole di derivazione

# Progettazione Logica

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP2** | Progettazione Logica | Relmi Emanuele |

## Ristrutturazione Schema Concettuale

### Analisi delle Prestazioni

#### Tavola dei volumi

|  |  |  |
| --- | --- | --- |
| **Concetto** | **Tipo** | **Volume** |
| Utente | E | 1000 |
| Sito | E | 5 |
| Articolo | E | 5000 |
| Commento | E | 8000 |
| Utente registrato | E | 800 |
| Utente cookieless | E | 200 |
| Cookie | E | 25 |
| Cookie di sessione | E | 5 |
| Cookie persistente | E | 20 |
| --- | - | --- |
| Accesso | R | 4000 |
| Contenuto | R | 1000 |
| Recensione | R | 4000 |
| Scrittura | R | 600 |
| Lettura | R | 5000 |
| Assegnazione | R | 200 |
| Profilazione | R | 3500 |

Tabella 6. Tavola dei volumi

* *Ipotesi 1*: in media un utente scrive 10 commenti à 10 \* 800 = 8000 commenti;
* *Ipotesi 2*: in media un sito ha 4 cookies persistenti e 1 cookie di sessione à 5 cookies per sito à 5 \* 5 = 25 cookies totali
* *Ipotesi 3*: in media un utente legge 5 articoli al giorno à 1000 \* 5 = 5000 articoli totali

#### Tavola delle operazioni

|  |  |  |
| --- | --- | --- |
| **Operazione** | **Tipo** | **Frequenza** |
| **Operazione 1**: registrazione di un nuovo utente su uno specifico sito | I | 150 al giorno |
| **Operazione 2**: inserimento di un commento specifico | I | 10 al giorno |
| **Operazione 3**: controllo degli utenti che hanno accettato i cookie | I | 10 al giorno |
| **Operazione 4**: stampa il numero di commenti scritto da un utente registrato | I | 5 al giorno |
| **Operazione 5**: controllo dell’articolo più letto dagli utenti | B | 1 a settimana |
| **Operazione 6**: controllo del sistema operativo utilizzato dagli utenti cookieless | B | 1 ogni 2 settimane |
| **Operazione 7**: conteggio dei tag che appaiono più frequentemente negli articoli letti | B | 1 al mese |
| **Operazione 8**: stampare gli articoli letti di ogni utente | I | 2 al giorno |

Tabella 7. Tavola delle operazioni

## Analisi delle ridondanze

* **Ridondanza 1**: attributo *Numero di Commenti* dell’entità Utente Registrato.TIPO: Attributo Derivabile da conteggio di occorrenze.

L’attributo *Numero di Commenti* può essere derivato contando le occorrenze dell’associazione *Scrittura*.

* **Ridondanza 2**: attributo *Articolo Letto* delle entità Utente Registrato e Utente Cookieless.  
  TIPO: 3) Attributo derivabile da conteggio di occorrenze.

L’*articolo letto* è possibile derivarlo contando le occorrenze della relazione *Lettura Registrata* e *Lettura Anonima*.

### Analisi della ridondanza 1: Esami sostenuti

* **Operazione 4: Stampa il numero di commenti scritto da un utente registrato**

***Con Ridondanza***

|  |  |  |  |
| --- | --- | --- | --- |
| **CONCETTO** | **COSTRUTTO** | **ACCESSI** | **TIPO** |
| Utente Registrato | E | 1 | L |

Costo = Costo complessivo = 1L = 1 accesso

***Senza Ridondanza***

|  |  |  |  |
| --- | --- | --- | --- |
| **CONCETTO** | **COSTRUTTO** | **ACCESSI** | **TIPO** |
| Utente Registrato | E | 1 | L |
| Scrittura | R | 10 | L |

Costo = 1L + 10L = 11 accessi

Costo complessivo = 2 \* 11 = 22 accessi

* **Operazione 8:** **Stampa il numero di articoli letti di ogni utente**

***Con Ridondanza***

|  |  |  |  |
| --- | --- | --- | --- |
| **CONCETTO** | **COSTRUTTO** | **ACCESSI** | **TIPO** |
| Utente | E | 1 | L |
| Lettura | R | 1 | L |

Costo = 1L + 1L = 2 accessi

***Senza Ridondanza***

|  |  |  |  |
| --- | --- | --- | --- |
| **CONCETTO** | **COSTRUTTO** | **ACCESSI** | **TIPO** |
| Utente | E | 1000 | L |
| Lettura | R | 5000 | L |

Costo = 1000L + 5000L = 6000 accessi

#### Valutazione della ridondanza 1

Dopo aver analizzato le operazioni che coinvolgono la ridondanza si osserva che, con il carico considerato:

* In presenza di ridondanza il costo delle operazioni è di circa 3 accessi giornalieri
* L’occupazione di memoria è di circa 16000 byte
* In assenza di ridondanza il costo delle operazioni è di circa 6000 accessi giornalieri

Pertanto, si decide di mantenere la ridondanza in quanto riduce il numero di accessi.

## Eliminazione delle generalizzazioni

Nello schema concettuale sono presenti due gerarchie: quella relativa a Utente e quella relativa a Cookie.

### Generalizzazione *Utente*

Per gli utenti si può notare che le entità figlie, *Utente Registrato* ed *Utente* *Cookieless*, hanno attributi specifici che li distinguono.

Essendo la generalizzazione totale procediamo ad accorpare l’entità padre all’interno delle figlie.

### Generalizzazione *Cookie*

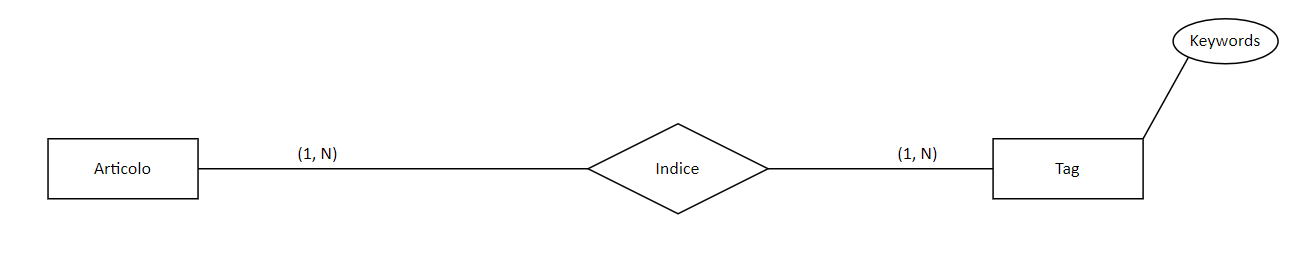
Per i cookie notiamo che le entità figlie, *Cookie Di Sessione* e *Cookie Permanenti*, hanno attributi specifici che le distinguono. Anche in questo caso, essendo la generalizzazione totale, procediamo ad accorpare l’entità padre all’interno delle figlie.

## Partizionamento/Accorpamento Entità e Associazioni

Un possibile accorpamento nel nostro schema può essere effettuato tra le entità **Utente Cookieless** e **Cookie Di Sessione**. Si può osservare che l’operazione 6 fa riferimento al *Sistema Operativo* utilizzato dagli *Utenti Cookieless*. Si potrebbe pensare quindi di accorpare le due entità.

Tuttavia, anche se andassimo ad accorparle, volendo conoscere solamente il *Sistema Operativo*, ci basterebbe accedere una sola volta all’entità *Cookie Di Sessione* senza dover passare per l’utente. Decidiamo quindi di non effettuare tale accorpamento.

Per poter eliminare l’attributo multivalore *Tag* associato all’entità **Articolo**, introduciamo una nuova entità **Tag** legata tramite un’associazione molti a molti con l’entità *Articolo*, che viene privata del relativo attributo. Di seguito mostriamo l’attributo modificato.



## Scelta degli identificatori principali

Solo l’entità **Utente** **Registrato** presenta due identificatori: l’*indirizzo* *IP* e lo *username*. Tra i due nel nostro caso è preferibile scegliere il secondo.

Infatti, un *indirizzo* *IP*, essendo composto da 16 cifre di tipo INT (4 byte), andrebbe ad occupare 64 byte di memoria, in opposizione allo username, il quale essendo un ***VARCHAR*** (1 byte) composto da massimo 20 caratteri, occupa circa 20 byte.

Inoltre, poiché l’**Utente Registrato** può scrivere un commento ed essendo quest’ultimo collegato al primo, risulta essere più identificativo l’utilizzo dello *username* al fine della profilazione.

## Schema ristrutturato finale

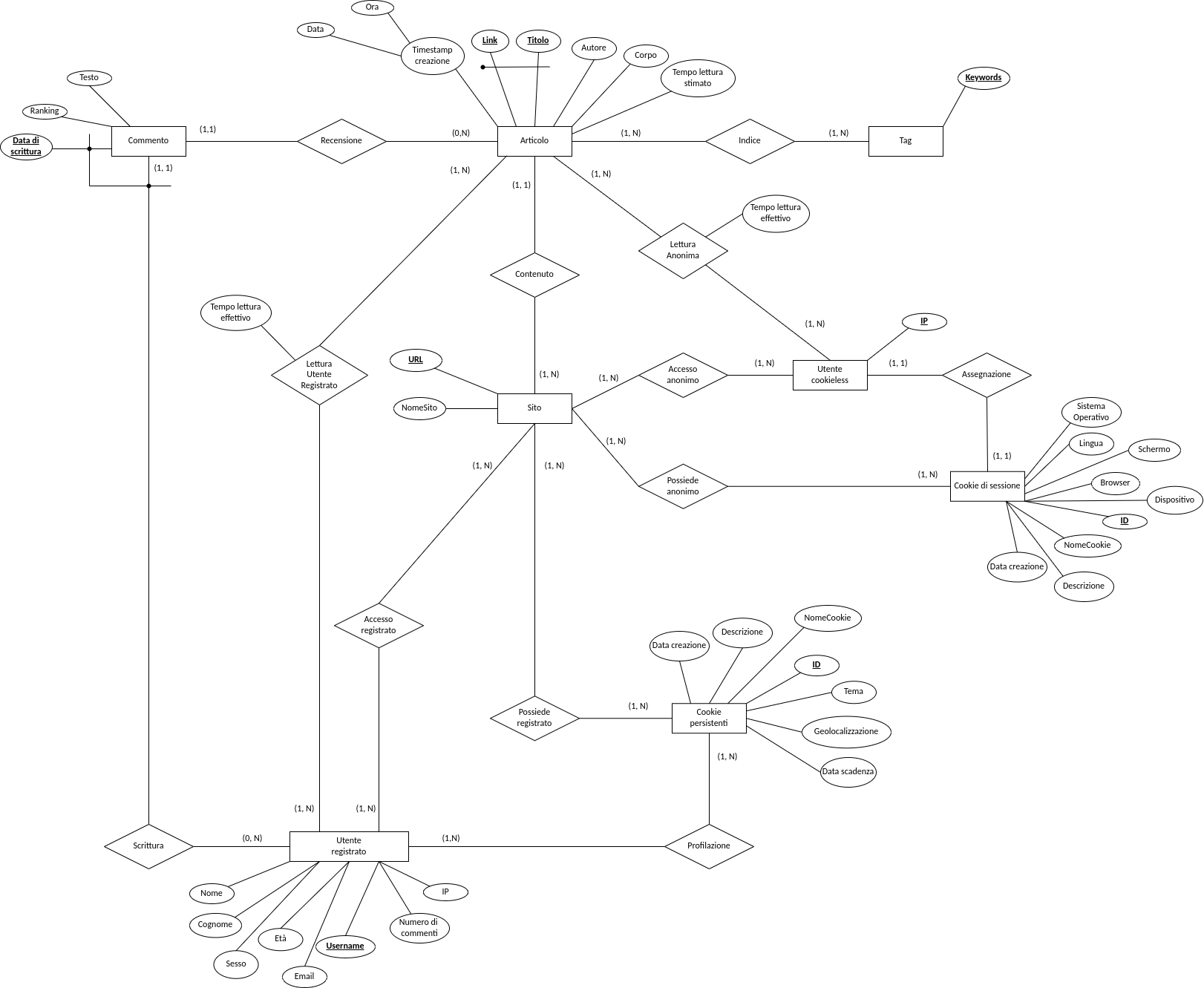


Figura 3. Schema ER Ristrutturato

## Schema logico

**Utente Registrato** (Username, Nome, Cognome, Età, Sesso, E-mail, IP, Numero Commenti)

**Profilazione** (Utente Registrato, Cookie Persistenti)

**Accesso Registrato** (Utente Registrato, Sito)

**Lettura Registrata** (Utente Registrato, Link Articolo, Titolo Articolo, Tempo Lettura Effettivo)

**Utente Cookieless** (IP, ID Cookie)

**Lettura Anonima** (Utente Cookieless, Link Articolo, Titolo Articolo, Tempo Lettura Effettivo)

**Accesso Anonimo** (Utente Cookieless, Sito)

**Cookie Persistenti** (ID, Nome Cookie, Descrizione, Geolocalizzazione, Tema, Data Creazione, Data Scadenza)

**Possiede Registrato** (Cookie Persistente, Sito)

**Cookie Sessione** (ID, Nome Cookie, Descrizione, Sistema Operativo, Lingua, Schermo, Browser, Data Creazione)

**Possiede Anonimo** (Cookie Sessione, Sito)

**Sito** (URL, Nome Sito)

**Articolo** (Link, Titolo, Corpo, Autore, Data Creazione, Ora Creazione, Tempo Lettura Stimato, URL Sito)

**Commento** (Data Scrittura, Username Utente, Ranking, Testo, Link Articolo, Titolo Articolo)

**Tag** (Keywords)

**Indice** (Link Articolo, Titolo Articolo, Tag)

## Documentazione dello schema logico

(*Opzionale*) Inserire la documentazione dello schema logico. Ad esempio, inserendo una rappresentazione grafica dei vincoli di integrità referenziale e/o una descrizione (paragrafo 8.4.7 del libro, figure 8.27-8.28,8.29)

ESEMPIO

Immagine che contiene tavolo

Descrizione generata automaticamente

# Normalizzazione

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP3** | Normalizzazione | Raffone Riccardo |

|  |  |  |
| --- | --- | --- |
| ***Relazione*** | ***FN*** | ***Tabelle*** |
| Utente Registrato | 1-2-3 | R1(**Username**, Nome, Cognome, Età, Sesso, E-mail, Articolo Letto, Numero Commenti, IP) |
| Sito | 1-2-3 | R1(**URL**, Nome Sito) |
| Articolo | 1-2-3 | R1(**Titolo**, **Link**, Autore, Corpo)  T2(**Autore**, Data Creazione, Ora Creazione, Tempo Lettura Stimata) |
| Tag | 1-2-3  Boyce & Codd | R1(**Keywords**) |
| Commento | 1-2-3 | R1(**Data Scrittura**, **Username**, Ranking, Testo) |
| Utente Cookieless | 1-2-3 | R1(**IP**, Articolo Letto) |
| Cookie di Sessione | 1-2-3  Boyce & Codd | R1(**ID**, Nome Cookie, Descrizione, Data Creazione, SO, Lingua, Schermo, Browser, Dispositivo) |
| Cookie Persistenti | 1-2-3  Boyce & Codd | R2(**ID**, Nome Cookie, Descrizione, Data Creazione, Data Scadenza, Tema, Geolocalizzazione) |

Per quanto concerne la normalizzazione dello schema relativo alla progettazione logica precedentemente effettuata, abbiamo avuto modo di denotare come, ad eccezione della relazione **Articolo**, l’intero schema fosse già adattato in *terza forma normale* (*3FN*).

## Normalizzazione di “Articolo”

Circa **Articolo**, abbiamo innanzitutto verificato la presenza della *prima* e *seconda forma normale* (*1FN* e *2FN*).

|  |  |  |
| --- | --- | --- |
| ***Relazione*** | ***FN*** | ***Tabelle*** |
| Articolo | 1 | R1(**Titolo**, **Link**, Autore, Corpo, Data Creazione, Ora Creazione, Tempo Lettura Stimata)  (A1, A2) à A3, A4, A5, A6, A7 |
| Articolo | 2 | R1(**Titolo**, **Link**, Autore, Corpo)  R2(Autore, Data Creazione, Ora Creazione, Tempo Lettura Stimata)  (A1, A2) à A3, A4  A3 à A5, A6, A7 |

Appurato ciò, abbiamo impostato l’attributo *Autore* come chiave di una seconda relazione, andando a decomporre **Articolo** in due relazioni, senza perdita di informazioni o eventuali anomalie, rispettando i requisiti della *3FN*.

|  |  |  |
| --- | --- | --- |
| ***Relazione*** | ***FN*** | ***Tabelle*** |
| Articolo | 1-2-3 | R1(**Titolo**, **Link**, Autore, Corpo)  T2(**Autore**, Data Creazione, Ora Creazione, Tempo Lettura Stimata) |

# Script Creazione e Popolamento Database

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP2** | SQL: Script creazione e popolamento | Relmi Emanuele |

*Inserire lo script di creazione del database e lo script per il popolamento con dati di esempio (può essere anche un solo script che contiene sia creazione che popolamento) del database progettato.*

*Gli script devono essere incollati qui (come testo o come immagine) e poi inseriti nella consegna finale come file di testo .sql*

CREATE DATABASE “Cookies Profilation”

WITH  
    OWNER = postgres  
    ENCODING = 'UTF8'  
    LC\_COLLATE = 'en\_GB.UTF-8'  
    LC\_CTYPE = 'en\_GB.UTF-8'  
    TABLESPACE = pg\_default  
    CONNECTION LIMIT = -1  
    IS\_TEMPLATE = False;

# Query SQL

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP3** | SQL: Query | Raffone Riccardo |

*Devono essere presenti almeno 3 query non banali:*

* *1 query non banale con operatore di aggregazione e almeno una join a tre tabelle (sezione 7.1)*
* *1 query non banale nidificata con interpretazione complessa (Sezione 7.2)*
* *1 query non banale insiemistica (sezione 7.3)*

*Altre query inserite come sottosezioni della sezione 7.4*

## Query con operatore di aggregazione e join: “Estrarre il numero di utenti registrati che accedono ad un sito con uno specifico indirizzo IP”

La query richiesta permette di ottenere il numero totale di utenti registrati che accedono ad un determinato sito, in base all’*IP* dell’utente e al *Nome Sito*, sfruttando le relazioni **Utente Registrato**, **Sito** e **Accesso Registrato**.

|  |
| --- |
| SELECT COUNT (ALL ar."Utente Registrato") AS "Tot Utenti/Sito"  FROM "Accesso Registrato" ar  JOIN "Utente Registrato" ur ON ar."Utente Registrato" = ur."Username"  JOIN "Sito" s ON ar."Sito" = s."URL"  WHERE s."Nome Sito" = 'La Repubblica' AND ur."IP" = '192.168.0.11'; |

## Query nidificata complessa: “Stampare i commenti con un ranking superiore alla media”

*La query pocanzi citata si occupa di stampare tutti i commenti che rispettano la condizione imposta, la quale prevede che il ranking dei commenti da restituire debba essere superiore al ranking medio per singolo articolo.*

SELECT C.\*

FROM "Commento" C

WHERE C."Ranking" > (

SELECT AVG(C1."Ranking")

FROM "Commento" C1

WHERE C1."Link Articolo" = C."Link Articolo" AND C1."Titolo Articolo" = C."Titolo Articolo"

)

ORDER BY C."Link Articolo";

## Query insiemistica: “Estrarre quanti utenti di sesso femminile hanno visitato un determinato sito con un'età superiore ai 16 anni”

*Per la query insiemistica, abbiamo deciso di effettuare un’operazione che va ad estrarre gli utenti registrati di sesso femminile che hanno visitato un sito specifico, nel caso in esame si è scelto “La Gazzetta dello Sport”, e che hanno un’età superiore ai 16 anni.*

*SELECT UR."Username"*

*FROM "Utente Registrato" UR*

*GROUP BY UR."Username"*

*HAVING UR."Età" > 16 AND UR."Sesso" = 'F'*

*INTERSECT*

*SELECT AR."Utente Registrato"*

*FROM "Accesso Registrato" AR*

*WHERE AR."Sito" = 'https://www.gazzetta.it'*

## Eventuali Altre query

### “Estrarre i cookie di sessione che denotano l'utilizzo dello stesso Sistema Operativo”

*La query indicata si occupa di individuare gli utenti cookieless, che utilizzano lo stesso Sistema Operativo, andando ad esaminare i cookie di sessione affibbiati agli stessi.*

*SELECT \**

*FROM "Cookie Sessione" CS1*

*WHERE EXISTS (*

*SELECT \**

*FROM "Cookie Sessione" CS2*

*WHERE CS2."Sistema Operativo" = CS1."Sistema Operativo" AND CS2."ID" <> CS1."ID"*

*);*

# Viste

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP4** | Viste | Quagliuolo Francesco |

## *Vista Lettura Registrata*

La vista ideata seleziona dalla relazione ***Lettura Registrata***, il *titolo degli articoli* e, sfruttando la correlazione tra gli ***Utenti Registrati*** e gli ***Articoli*** letti, calcola il numero totale di letture per ogni articolo.

DROP VIEW IF EXISTS "Vista Lettura Registrata";

CREATE VIEW "Vista Lettura Registrata" ("Titolo Articolo", "NumUtenti") AS

SELECT "Titolo Articolo", COUNT (\*)

FROM "Lettura Registrata"

GROUP BY "Titolo Articolo";

### Query con Vista: “Estrarre il titolo dell'articolo più letto”

SELECT "Titolo"

FROM "Articolo"

WHERE "Titolo" IN (

SELECT "Titolo Articolo"

FROM "Vista Lettura Registrata"

WHERE "NumUtenti" = (

SELECT MAX ("NumUtenti")

FROM "Vista Lettura Registrata"

)

);

# Trigger

## Trigger inizializzazione: *Titolo Trigger*

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP1** | Trigger inizializzazione/popolamento database | Postiglione Matteo |

*Inserire una breve descrizione del trigger. Il trigger dovrebbe scattare durante la fase di popolamento/inizializzazione del database*

*Incollare lo script del trigger.*

## Trigger per vincoli aziendali

|  |  |  |
| --- | --- | --- |
| ***Workpackage*** | ***Task*** | ***Responsabile*** |
| **WP4** | Trigger per vincoli aziendali | Quagliuolo Francesco |

*La sezione deve contenere almeno due trigger. È possibile aggiungerne altri se necessario*

### Trigger1: Titolo Trigger

*Inserire una breve descrizione del trigger*

*Incollare lo script del trigger.*

*Se presente un altro TRIGGER, inserire una sottosezione 9.2.2 … e ripetere la struttura. EVITARE TRIGGER BANALI O CHE SIANO TROPPO SIMILI A QUELLI GIA’ REALIZZATI*