Progetto database (da abbellire ovviamente)

Indice dei contenuti

Requisiti di progetto e descrizione dell’ambito applicativo

Si intende realizzare un database a supporto di un social network. Per definire il contesto su cui si va a realizzare la base di dati, tale social network è chiamato “Sententia”, che significa opinione dal latino, il cui scopo è quello di dare la possibilità a chiunque di esprimere la propria opinione su qualunque argomento essi desiderino dandone una recensione. L’idea principale che lo rende unico a tutti gli effetti è quello di non imporre limiti di alcun tipo e organizzare le recensioni in modo tale che siano trovabili e consultabili efficientemente.

Il fulcro del sistema si colloca indubbiamente nell’utente, in quanto essenzialmente tutto ruota attorno alla figura dello stesso, a ciò che decide pubblicare, ai messaggi che invia e riceve, e le sue transazioni. I loro dati sono: Email e Username che sono entrambi univoci, nome, cognome, password, e foto profilo. Una prima funzionalità riguarda la richiesta di amicizia tra utenti. L’amicizia non è necessariamente reciproca, ovvero un utente A può avere B come amico ma non necessariamente è vero per B rispetto ad A. Un’amicizia può essere annullata, ma non è necessario tenere traccia dell’intera catena di amicizie/annullamenti e neanche la data in cui è stata effettuata l’annullazione.

Le recensioni sono rappresentate nel seguente modo: esse riguardano un oggetto, che a sua volta appartiene ad una categoria, e contengono informazioni quali un codice identificativo, una foto, un titolo e un valore. Esistono due tipologie di interazioni tra utenti: quelle pubbliche, che consistono nel commentare una determinata recensione, e quelle private, gestite da un sistema di messaggistica che consente l’invio e ricezione di messaggi testuali, immagini e persino altre recensioni. I commenti riguardanti una determinata recensione sono caratterizzati da un codice identificativo in base alla recensione (per facilità di visualizzazione e ricerca dei vari commenti) e dal testo dello stesso. Si richiede inoltre la possibilità di rispondere a commenti. Un’altra forma di espressione della propria opinione è sotto forma della medaglia, che un utente può conferire (solamente una volta) ad una determinata recensione. Per quanto riguarda invece i messaggi fra utenti, in generale si necessita di un codice identificativo in base alla coppia di utenti coinvolti, inoltre occorre conoscere se siano stati letti o meno, e nel caso in cui essi siano testuali si richiede solamente il testo del messaggio stesso, in caso contrario contengono una descrizione opzionale, affiancata da un’immagine oppure una recensione.

Le pubblicazioni tra utenti necessitano di un controllo per conservare la sicurezza della piattaforma, occorre dunque introdurre una figura il cui compito è quello di scorrere la lista dei contenuti della piattaforma ed eventualmente rimuoverli se si reputa che violano le regolazioni imposte. Tale figura prende il nome di moderatore, e i dati che vengono memorizzati sono gli stessi dell’utente “utilizzatore” a meno della foto profilo. Un moderatore può fare varie cose: può effettuare un “ban” su uno o più utenti che possono essere annullati successivamente se necessario, e può “nascondere” una recensione temporaneamente. Per il fatto che le recensioni vengono considerate con più cautela, dato che può essere una fonte di monetizzazione e può richiedere svariate ore di lavoro, mentre per i commenti ci si aspetta che non siano particolarmente importanti per la piattaforma, motivo per il quale la “cancellazione” di una recensione è in molti casi soltanto temporanea, dando la possibilità al creatore di contestare la rimozione con una motivazione valida. Invece i commenti possono essere rimossi direttamente dalla base di dati, e non è necessario neanche memorizzarsi la data o il moderatore che l’ha fatto. In base al numero di violazioni che ha effettuato un determinato utente, un moderatore può decidere di “bandire” un utente, e tale operazione deve essere memorizzata assieme al moderatore e all’utente coinvolti.

Essendo il progetto una base di dati di supporto ad un social network, è importante monitorare il tempo esatto di ciascuna pubblicazione. Questo include i messaggi, i commenti e le recensioni, ma anche il ban di una recensione, le varie transazioni, le amicizie tra utenti, ecc. Questo processo di controllo dei contenuti viene affidato in parte anche agli utenti utilizzatori, in quanto si permette a questi ultimi di segnalare una o più recensioni. La segnalazione può avere una tipologia tra quelle predefinite, oppure l’utente può scrivere liberamente la sua motivazione se questa non è presente tra quelle selezionabili. Una segnalazione da parte di un utente riferita ad una particolare recensione può essere effettuata soltanto una volta.

Un altro aspetto molto importante è il discorso sulla monetizzazione. Vi sono due tipologie di utenti: gli utenti normali e gli utenti premium. Gli utenti normali sono quelli descritti fino ad ora, mentre quelli premium hanno tutte le caratteristiche degli utenti normali con l’aggiunta del fatto che possono ricevere donazioni monetarie da utenti (sia da utenti normali sia premium).

Una peculiarità degli utenti premium è che hanno la possibilità di definire piani di abbonamento, ciascuno con un determinato periodo e quantità di pagamento (per esempio: 2€ al mese), ai quali una molteplicità di utenti possono iscriversi. Dell’iscrizione occorre memorizzare la data e l’eventuale data di abbandono. Inoltre, gli utenti premium possono anche impostare incentivi per l’iscrizione a determinati piani per mezzo dell’esclusività temporanea di una recensione. Per ciascun piano e ciascuna recensione pubblicata dallo stesso utente si può avere un’esclusività definita come una data a partire dalla quale la visione esclusiva è permessa (solo per gli iscritti a quel piano). Una stessa recensione può fare parte di più esclusività diverse (soltanto per piani diversi, non all’interno dello stesso piano) ciascuna con una data di visualizzazione anticipata differente. Allo stesso tempo un piano può promettere l’esclusività di più recensioni con date diverse. Vi è inoltre una proprietà della recensione che ne indica la data in cui è visibile a tutti gli utenti.

Per questo progetto occorre tenere traccia di tutte le transazioni effettuate dagli utenti, in modo tale che essi stessi possano, sotto determinate condizioni, annullarla, e occorre anche tenere traccia del fatto che una transazione sia già stata annullata o meno. Le transazioni possono essere di due tipologie: le transazioni manuali e quelle automatiche. Quelle manuali possono essere effettuate liberamente da un utente qualsiasi verso un utente premium, con una quantità di denaro a libera scelta, mentre quelle automatiche sono quelle effettuate per effetto dell’iscrizione ad un piano. Esse sono identificate da un CRO, un codice utilizzato universalmente per identificare una transazione, e devono contenere il numero della carta di credito dell’utente donatore e l’identificativo dell’utente premium destinatario. Un utente può avere più carte di credito (anche nessuna) e una carta di credito può essere utilizzata da più utenti. È richiesto di imporre il vincolo che in una certa data, un determinato utente può soltanto effettuare una donazione rivolta ad un determinato utente premium, e questo va garantito sia per le transazioni manuali sia per quelle automatiche.

Senza addentrarsi troppo nel dettaglio, in quanto si andrebbe fuori dallo scopo del progetto in questione, i pagamenti vengono svolti nella modalità seguente. L’azienda sviluppatrice del social network gestisce entrate e uscite (transazioni dall’utente donatore all’azienda e quelle dall’azienda all’utente destinatario), in modo tale che essa possa incassare una certa percentuale. Per le entrate si utilizza la carta di credito, mentre per quanto riguarda le uscite vengono gestite con un bonifico tramite IBAN. L’IBAN è un’informazione necessaria solo per gli utenti premium, mentre non la si memorizza per gli utenti non-premium, ed entrambi possono fornire zero o più carte di credito (e ne devono avere almeno una se hanno intenzione di effettuare transazioni).

Essendo il progetto in questione un social network, esso necessita di continui cambiamenti e aggiunte, non solo per scopi di manutenzione ma anche per il fine di ritenere l’attenzione del pubblico, quindi è richiesta la possibilità di poter agilmente aggiungere funzionalità.

Il social network deve poter essere robusto ad eventuali sistemi di sovraccarico delle risorse, per esempio l’invio automatizzato di un numero elevato di commenti o recensioni (anche chiamato spam). Questo ha anche lo scopo di migliorare l’esperienza utente. Si richiede che un utente possa effettuare al massimo 2000 pubblicazioni (recensioni o commenti) in un’ora.

È necessario avere a disposizioni determinati dati per soddisfare alcune funzionalità del social network, per esempio il numero di medaglie e di recensione di un utente che è visualizzabile da un menù, e le varie informazioni che riguardano il guadagno degli utenti premium. In particolare, è assolutamente indispensabile poter reperire in maniera efficiente i seguenti dati: il totale incasso di un utente premium, il numero di transazioni effettuate da un particolare utente, il numero di medaglie ricevute, il numero di recensioni effettuato, e il numero di risposte ad un determinato commento. [Ricordarsi poi nella parte sui dati derivati di parlare anche della visibilità della recensione] Il numero di consultazioni di ciascun dato verrà riportato successivamente nella sezione riguardante i dati derivati.

Progetto concettuale – Schema E/R

I requisiti del progetto rilevano come fulcro la relazione tra gli utenti e le recensioni, che viene rappresentata nel seguente schema scheletro (estremamente semplificato e privo della maggior parte delle funzionalità richieste):

[ Immagine schema scheletro ]

È possibile ampliare questo schema al fine dell’implementazione di ciascuna richiesta. Innanzitutto, i commenti possono rispondere ad altri commenti, viene dunque introdotta un’auto-associazione uno a molti. Non viene gestito, in questa prima fase di progettazione concettuale, il problema che un commento può essere in risposta ad un altro commento pur non riferendosi alla stessa recensione.

[ Immagine commento ]

Per ragionare sul problema risulta efficace scomporlo in più parti:

1. La messaggistica e le amicizie tra utenti
2. I moderatori e il controllo dei contenuti
3. Gli utenti premium e le transazioni

Si può notare che queste tre componenti sono strettamente correlate agli utenti e alle recensioni. Per la natura estremamente dinamica della realtà che si va a rappresentare, si pone come obiettivo la semplicità del diagramma ER, lasciando la gestione dei vincoli più complessi ai trigger, discussi in una parte successiva di questa relazione.

La messaggistica tra utenti è gestita attraverso un’entità Messaggio legata all’entità Utente in due relazioni uno a molti (dove il lato molti è il Messaggio).

[ Immagine messaggio ]

Il messaggio contiene un codice che identifica lo stesso a partire dalla coppia di utenti “mittente-destinatario”, e le tipologie differenti vengono modellate con l’uso di una gerarchia, come si può vedere dalla seguente figura.

[ Immagine gerarchia messaggio ]

Si tratta di una gerarchia totale ed esclusiva perché un messaggio può essere solamente uno e uno solo di quei 3 tipi. Ci si rende conto del fatto che la descrizione opzionale è comune alle due entità MImmagine e MRecensione, si potrebbe dunque riscrivere come un’altra gerarchia che ha come entità padre “MNonTesto” o similmente denominata, tuttavia per questioni di semplicità di rappresentazione si sceglie di evitare di modificare il diagramma in questo modo. In ogni caso le decisioni di traduzione in uno schema logico verranno intraprese nella fase successiva.

Le amicizie vengono modellate attraverso un’auto-associazione molti a molti.

I moderatori e gli utenti normali condividono una grande porzione di dati, è possibile introdurre anche in questo caso una gerarchia totale ed esclusiva, dove l’entità padre viene chiamata “Persona”.

[ Immagine gerarchia persona ]

Un moderatore può bandire un determinato utente, e questo si traduce come una semplice associazione uno a molti.

Il moderatore può “disattivare”, ovvero rendere invisibile temporaneamente, le recensioni. Per progettare questo, si può ricorrere all’uso di un’entità “rimozione” che ha come chiave composta la data di effettuazione e il riferimento alla recensione, e contiene un’eventuale data di annullamento. In questo modo una rimozione può essere effettuata più volte in date diverse, ma questo deve essere permesso soltanto se vi sia una data di annullamento non nulla (che indica che sia effettivamente stata annullata), un vincolo che verrà gestito nelle fasi successive. Un utente può effettuare un report, il quale è identificato dall’utente e dalla recensione che sono coinvolti. In questo modo una recensione può essere disattivata e riattivata più volte.

Di seguito è riportato il diagramma E/R aggiornato.

[ Immagine ER aggiornato ]

Per quanto riguarda il sistema di monetizzazione di questo progetto, innanzitutto si può notare che gli utenti premium non sono altro che utente con alcune caratteristiche aggiuntive, in altre parole una specializzazione di utente. Questo si rappresenta all’interno dello schema ER sotto forma di una gerarchia non totale ed esclusiva come segue.

[ Immagine gerarchia Utente ]

La gerarchia è parziale perché un utente può non essere premium (gli utenti normali), e una discussione sull’esclusività in una gerarchia con una sola entità figlia è per certi aspetti irrilevante.

Il piano è in una relazione uno a molti con l’entità UPremium, dove la partecipazione di UPremium è facoltativa, e l’esclusività di una relazione viene gestita con una semplice associazione molti a molti tra Piano e Recensione contenente un attributo DAnticipata, ovvero la data oltre la quale gli utenti iscritti a quel piano sono abilitati alla visualizzazione della recensione. In questa fase non viene gestito il fatto che un piano può abilitare l’esclusività della recensione pubblicata da un altro utente, persino non premium.

Le iscrizioni vengono gestite con altrettanta semplicità, con un’entità Iscrizione che è identificata dall’utente che la effettua, il piano, e la data. Essa contiene un’eventuale data di abbandono. Similmente al ragionamento fatto sull’entità Rimozione, occorre assicurarsi successivamente che non sia possibile iscriversi due volte ad uno stesso piano senza averlo prima abbandonato. Dato che bisogna memorizzare alcune caratteristiche sulla carta di credito associata ad un particolare utente, si crea un’entità apposita “CartaCredito” in relazione molti a molti con l’entità Utente. Essa è identificata dal numero e contiene informazioni quali la data di scadenza e l’ente emittente. Lo schema E/R fin’ora considerato è il seguente:

[ Immagine E/R aggiornato ]

Le due tipologie di transazione sono per certi aspetti differenti, in quanto la transazione manuale deve contenere la somma di denaro donata, la data, l’utente premium destinatario della donazione, la carta di credito utilizzata e l’utente donatore, mentre la transazione automatica si riferisce ad un piano che contiene già l’utente premium e la somma di denaro. Può essere ragionevole pensare che un piano possa essere modificato nel tempo da un utente premium e persino cancellato. Si può notare, tuttavia, una problematica legata a questa casistica; se di una transazione automatica si tiene solamente traccia del piano che l’ha attivata, allora nel momento in cui quest’ultimo viene eliminato i dati riguardanti la somma pagata e l’utente premium verranno totalmente persi, rendendo conseguentemente lo storico di transazioni inutilizzabile. A questo punto si possono intraprendere due strade, di cui si espongono i rispettivi lati positivi e negativi. Come prima cosa, è possibile salvare all’interno della transazione automatica anche i dati della transazione in aggiunta del piano che l’ha attivata. Questo comporta una certa ridondanza dei dati che implica un piccolo spreco di memoria e i costi associati al mantenimento della consistenza attraverso i trigger, tuttavia si può liberamente modificare la tabella dei piani senza avere alcuna ripercussione significativa sullo storico. D’altra parte, si può impedire la modifica permanente di un piano, sostituendola con l’aggiunta di un nuovo piano con i cambiamenti apportati. Adottando questa strategia si memorizzano quindi tutti i piani che sono mai esistiti. Dato che in questo caso si vuole memorizzare uno storico, e i dati riguardanti i piani passati possono rivelarsi utili, si opta per la seconda scelta.

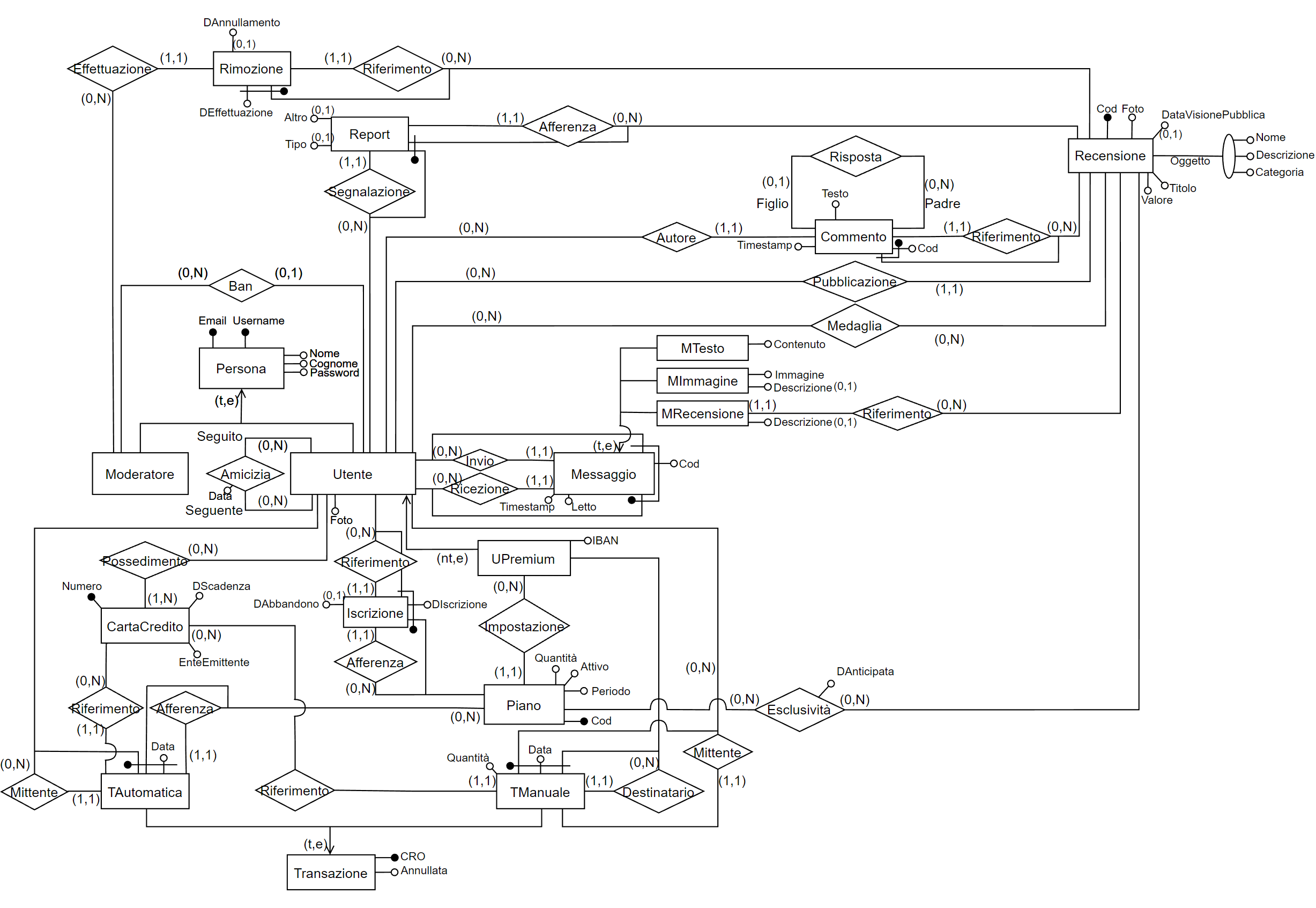
Da qua la progettazione si dirama ancora una volta, in quanto vi sono più modi per gestire la dicotomia tra piani attivi e inattivi. Una possibilità è quella di tenere due entità separate, e al posto di eliminare o modificare i dati già esistenti si crea una copia e la si inserisce nell’entità aggiuntiva “PianoInattivo”. Questo presenta il vantaggio considerevole della semplicità implementativa in certe condizioni, in quanto basta operare sull’entità “PianoAttivo” (o semplicemente “Piano” senza dover controllare ogni volta un attributo booleano e in generale impattare il funzionamento ordinario del database, il che potrebbe essere soggetto a errori. Tuttavia, a differenza della scelta di una semplice aggiunta di un attributo booleano, le relazioni presenti tra un piano (che sia attivo oppure no) e le altre entità potrebbero essere difficili da gestire. Come si vedrà nelle fasi successive, si opta per la presenza di un attributo booleano con l’obiettivo di ovviare il problema della complicazione delle interrogazioni facendo uso di una vista.

In seguito a queste riflessioni, il diagramma E/R risultante è il seguente:

[ Immagine E/R aggiornato ]

Dato che nella transazione viene memorizzata sia la carta di credito sia l’utente donatore, come conseguenza negativa con questo modello E/R è possibile immagazzinare una transazione da parte di un utente che non è in possesso della carta a cui si riferisce la transazione stessa, si tratta di un altro vincolo che verrà gestito in seguito tramite i trigger.

Si presenta dunque lo schema finale:



Progetto logico – Schema relazionale

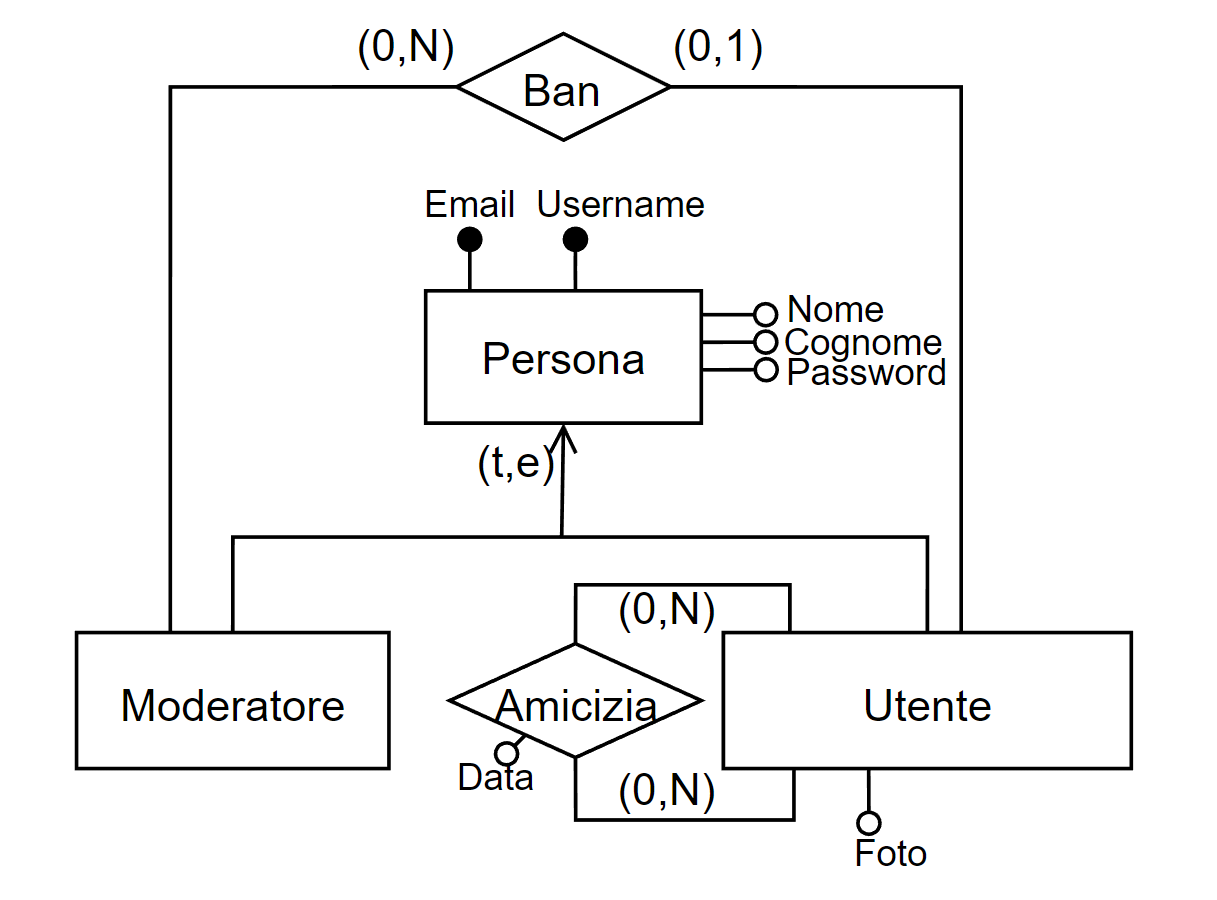
La progettazione logica permette di definire alcuni dettagli implementativi per poter operare nella pratica sui concetti stabiliti durante la fase di progetto concettuale, e viene decomposta nelle seguenti fasi:

* Eliminazione delle gerarchie ISA
* Selezione delle chiavi primarie, eliminazione delle identificazioni esterne
* Trasformazione degli attributi composti o multipli
* Traduzione di entità e associazioni in schemi di relazioni
* Verifica di normalizzazione

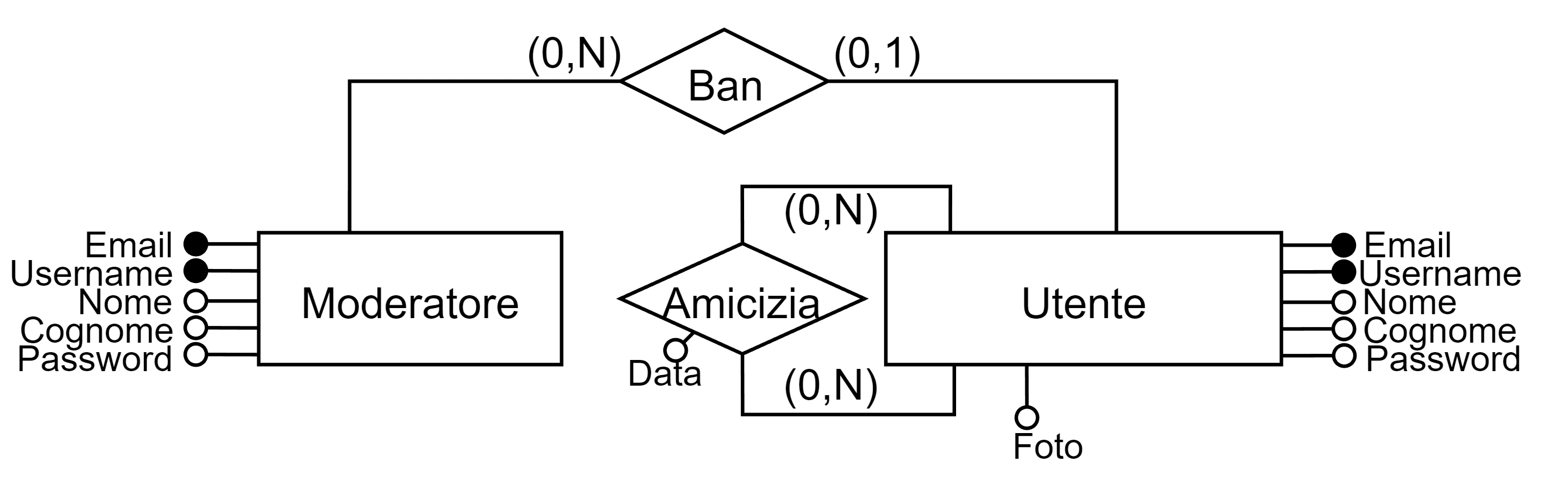
Si procede nell’ordine descritto.

Eliminazione delle gerarchia ISA

Nello schema E/R sono presenti 4 gerarchie ISA, si propone di focalizzare l’attenzione sulla seguente gerarchia per prima:

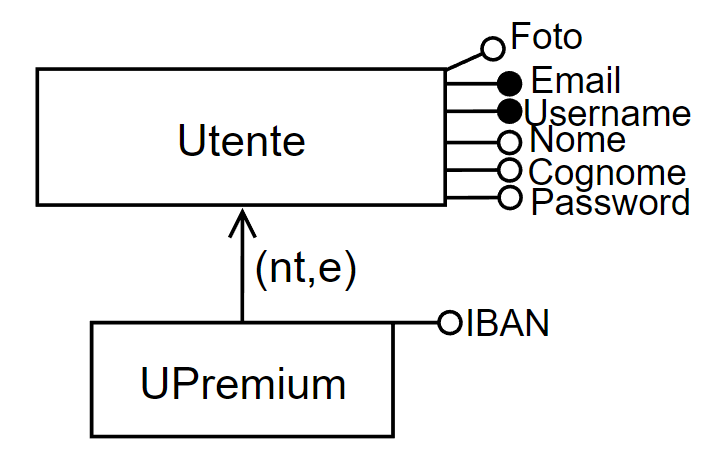


Nell’immagine sono evidenziate alcune associazioni la cui presenza influisce sulla decisione. Per effettuare una decisione appropriata sull’eliminazione, occorre tenere in considerazione che vi è un’ulteriore gerarchia sottostante, con Utente come entità padre. Proprio per questo motivo, e per il fatto che l’entità padre non si interfaccia con alcuna entità nella base di dati, si ritiene che la scelta migliore in questo caso sia un collasso verso il basso, come segue:

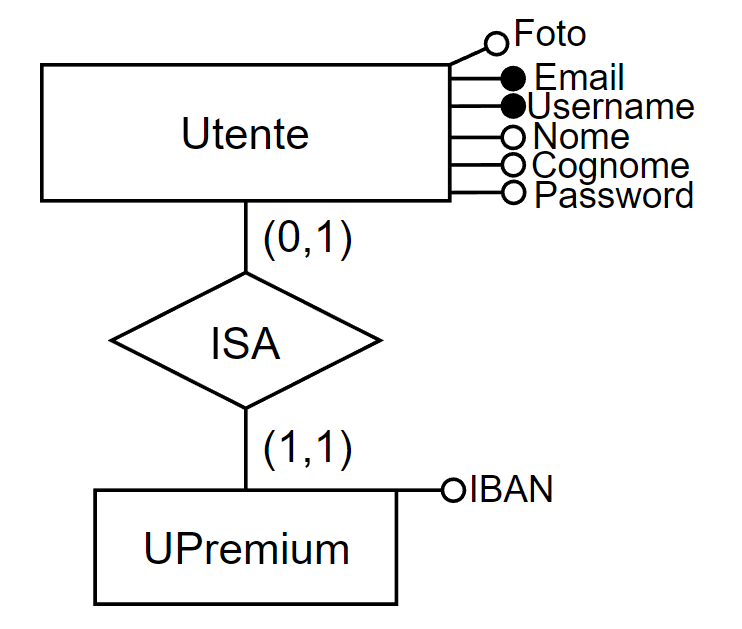


Questo richiederà, a livello implementativo, la creazione di alcuni trigger che si occuperanno di garantire che i valori di “Email” e “Username” siano univoci all’interno di entrambe le tabelle, dato che la gerarchia è esclusiva.

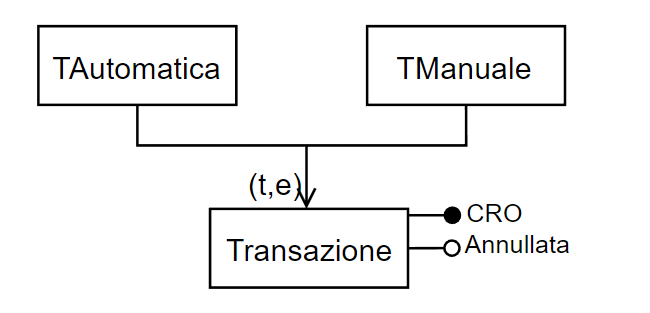
Si prende ora in considerazione la gerarchia immediatamente sottostante (nell’immagine è già stata applicata l’eliminazione della gerarchia già considerata):



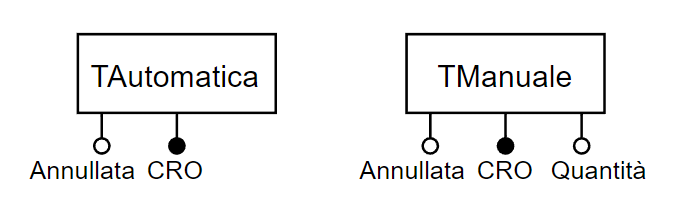
Vi sono numerose interazioni/associazioni che legano sia l’entità padre che l’entità figlia a tutte le altre, inoltre si può notare che vi sono numerose associazioni che legano le due entità Utente e UPremium, per esempio tutti gli utenti possono donare ma solamente gli utenti premium sono in grado di ricevere denaro. Si ritiene quindi che un mantenimento delle entità sia benefico in questo caso, dato che un collasso verso il basso non è possibile a causa della parzialità della gerarchia, e un collasso verso l’alto comporterebbe numerose complicazioni per la gestione delle varie associazioni. La sua traduzione è quindi la seguente:



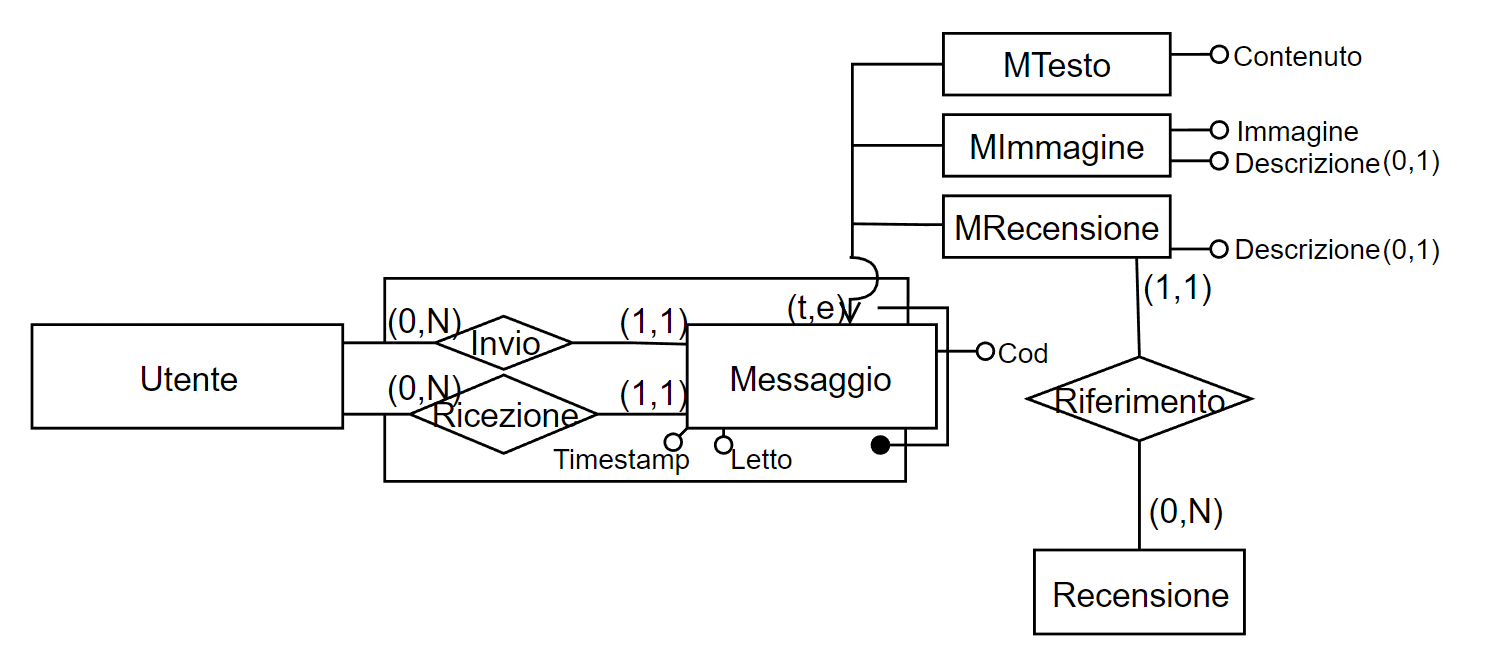
Per quanto riguarda la gerarchia che riguarda le transazioni:



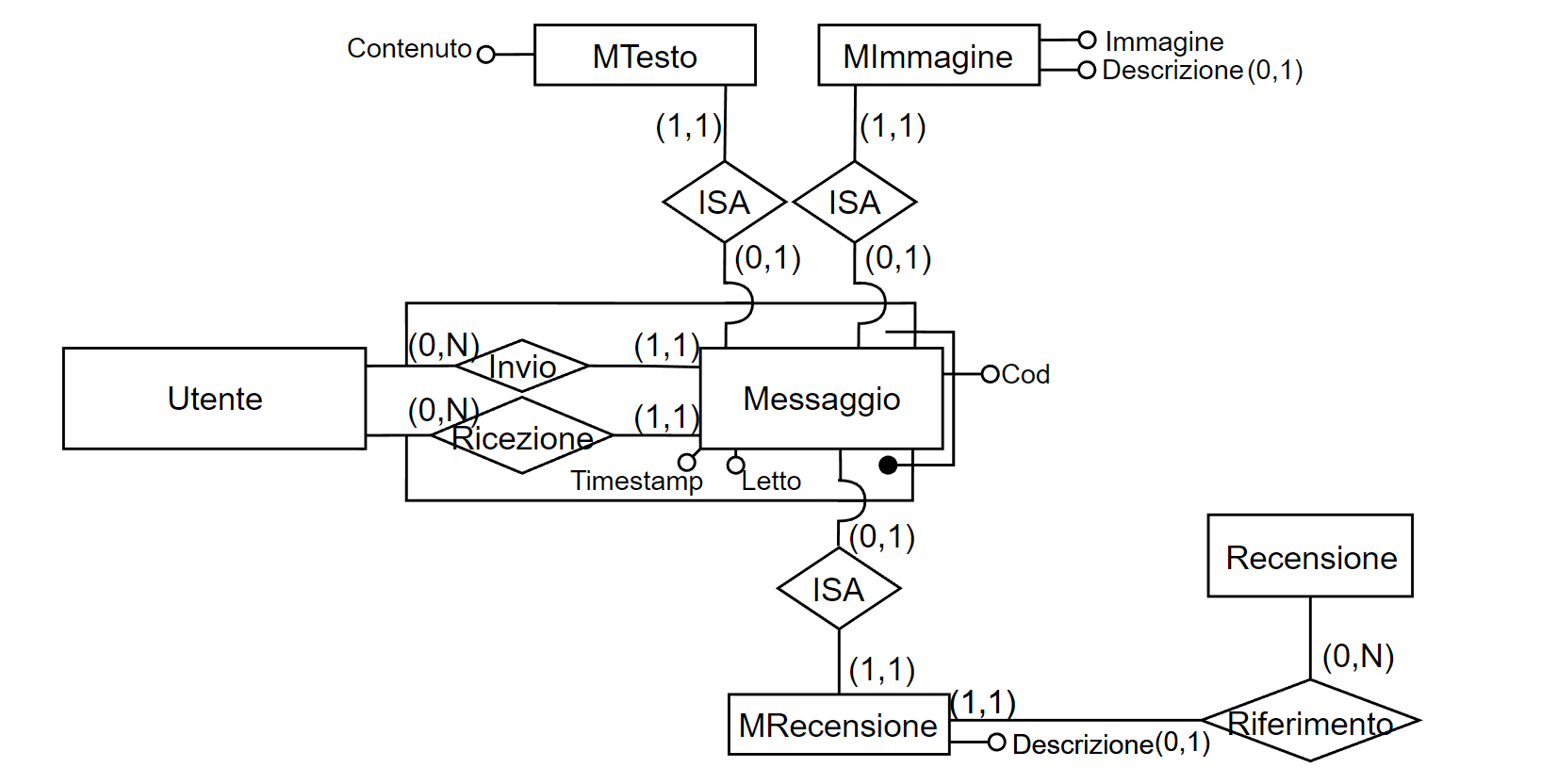
Essa può essere tradotta in più modi, tuttavia per il fatto che l’entità padre contiene solamente due attributi e nessuna associazione che la lega ad altre entità, un semplice collasso verso il basso è sufficiente.



Infine occorre gestire l’eliminazione della gerarchia di Messaggio. Si tratta di una gerarchia totale ed esclusiva, rappresentata nell’immagine seguente:



Innanzitutto occorre specificare che ci si aspetta che la maggior parte dei messaggi siano testuali, perciò non converrebbe effettuare un collasso verso l’alto in quanto si avrebbe un numero di NULL eccessivamente elevato. Il collasso verso il basso è altrettanto […] in questo caso, perché complicherebbe la traduzione delle due associazioni tra Utente e Messaggio, in aggiunta ai lati negativi intrinseci di questa modalità, ovvero il dover mantenere l’unicità della chiave su più tabelle separate attraverso i trigger. Il mantenimento delle entità si considera ottimale in questo caso, risultando nella seguente traduzione:



Selezione delle chiavi primarie, eliminazione delle identificazioni esterne

Si comincia con l’eliminazione delle identificazioni esterne, per poi passare alla selezione delle chiavi primarie.

L’entità rimozione possiede una chiave composta

Operazioni previste dalla base di dati – Descrizione e relativo codice SQL