



# UNIVERSITÀ DI PISA

## DIPARTIMENTO DI INFORMATICA

### BASI DI DATI (A)

### ANNO ACCADEMICO 2020/2021

## Gestione progetti ANSA

Tommaso Cavalieri (mat. 597707)  
DATA SCIENCE & BUSINESS INFORMATICS

29 gennaio 2021

### 1 Descrizione del dominio

#### SOCIETÀ

Il software da progettare è pensato per permettere alle società che producono informazioni giornalistiche, come ad esempio l'ANSA, di distribuire in tempo reale i comunicati stampa alle redazioni dei giornali che si iscrivono al servizio tramite un'architettura di tipo client-server. Risulta dunque evidente come le SOCIETÀ coinvolte nell'utilizzo di questo software siano due: le agenzie d'informazione o di stampa, che offrono il servizio di diffusione delle notizie, e le redazioni dei giornali, che solitamente le ricevono e le pubblicano, ma in alcuni casi si prestano anch'esse a diffonderle inviandole al server. I due tipi di società appena menzionate sono dunque due sottoclassi della classe qui descritta e ne formano una partizione. Per ognuna di esse sono disponibili, oltre al nome, la sede legale ed un rating, il cui valore viene calcolato sulla base dell'affidabilità, precisione e celerità con cui la società in questione solitamente diffonde le notizie.

#### REDAZIONI (UTENTI/CLIENTI)

Le sottoclasse REDAZIONI rappresenta quelli che di fatto sono gli utenti del servizio offerto, ovvero la distribuzione di notizie. Esse infatti si possono iscrivere ai canali tematici che preferiscono seguire, ottenendo così l'autorizzazione a ricevere le notizie riguardanti tali argomenti fintanto che non effettueranno un'eventuale cancellazione. In alcuni casi è possibile che le redazioni abbiano ulteriori autorizzazioni come inviare notizie al server o effettuare operazioni amministrative quali la creazione di nuovi canali, la disattivazione di quelli esistenti o la possibilità di trasferire tali autorizzazioni ad altri utenti ancora. In quest'interpretazione

dunque, in cui non è il privato cittadino ad essere considerato come cliente, bensì il giornale per cui lavora, è stato ritenuto sensato garantire a tutti i dipendenti della medesima redazione le medesime autorizzazioni e perciò sono state definite a questo livello. Ogni redazione ha un caporedattore, un formato di visualizzazione delle notizie di default ed un metodo di pagamento (mensile, trimestrale, semestrale o annuale) in base al quale varia l'importo complessivo da versare. Quest'ultimo dipende anche dal tipo di utenza selezionata, ECONOMY o PREMIUM: esse sono due sottoclassi di questa classe e ne formano una partizione.

## **AGENZIE D'INFORMAZIONE/STAMPA**

Questa sottoclasse rappresenta le SOCIETÀ che utilizzano il server per erogare il servizio da loro offerto, ovvero la diffusione di notizie. A differenza delle REDAZIONI però, esse non sfruttano il software anche con il fine di ricevere notizie e dunque non necessitano distinzioni sulla tipologia di utenza, in quanto risulta omogenea per tutte siccome non hanno la possibilità di iscriversi ai canali tematici. In queste agenzie non vi è un caporedattore, ma è comunque individuato un cosiddetto responsabile, che può essere ad esempio il presidente o l'amministratore delegato della società (dipende da com'è strutturata).

## **ECONOMY**

Le redazioni che sottoscrivono questo tipo di utenza si garantiscono la possibilità di iscriversi solamente ai CANALI TEMATICI di tipo ECONOMY ad avranno dunque una scelta più ristretta.

## **PREMIUM**

Le redazioni che sottoscrivono questo tipo di abbonamento invece si garantiscono la possibilità di iscriversi sia ai CANALI TEMATICI di tipo ECONOMY che a quelli di tipo PREMIUM. Le utenze PREMIUM inoltre danno, a differenza di quelle ECONOMY, la possibilità di ricevere offerte speciali e sconti sul servizio.

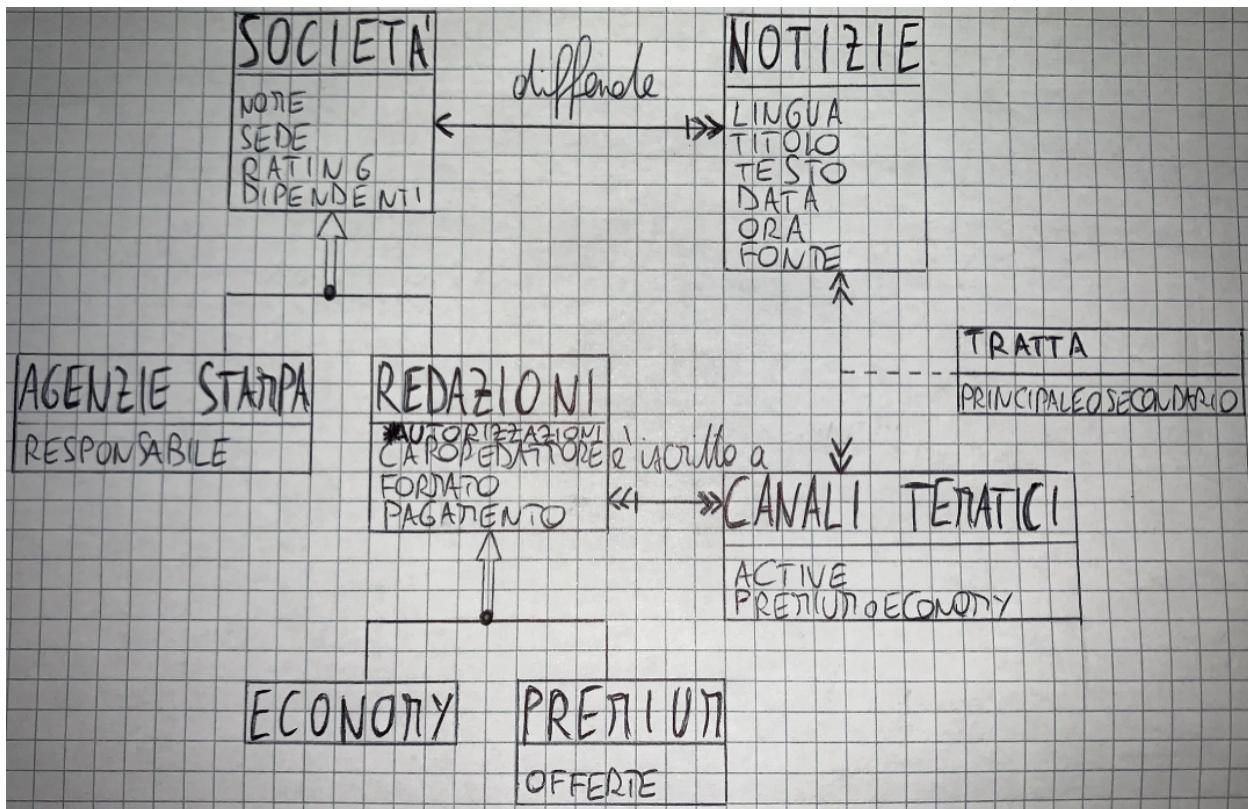
## **NOTIZIE**

Sono la componente principale del servizio offerto, che si concentra appunto sulla loro distribuzione. Esse possono essere in varie lingue ed ognuna presenta un titolo, un testo, una data e ora di pubblicazione, la fonte da cui è stata ottenuta e uno o più argomenti inerenti ad essa. Tra questi è inoltre possibile fare una distinzione tra quelli che sono gli argomenti principali e quelli che invece risultano essere secondari. Ogni notizia viene diffusa da una ed una sola società (non sono previste collaborazioni).

## **CANALI TEMATICI**

Rappresentano i diversi temi/argomenti trattati dalle notizie diffuse dal server. Tra di essi si possono trovare sia canali ormai chiusi che canali tutt'ora attivi e come già anticipato si distinguono fra canali di tipo PREMIUM ed ECONOMY, che corrispondono ai due tipi di iscrizioni sottoscrivibili dagli utenti. Non vi è invece limite massimo al numero di canali che una redazione decide di seguire, ma vi è invece un limite minimo banale siccome una redazione risulterà iscritta al servizio di distribuzione di news nel momento in cui si iscrive ad almeno un canale tematico. Si assume inoltre che tutti i canali abbiano al loro interno almeno una notizia, in quanto è stato ipotizzato che per aprirne uno nuovo sia necessario pubblicare almeno una notizia che lo tratti.

## 2 Schema concettuale

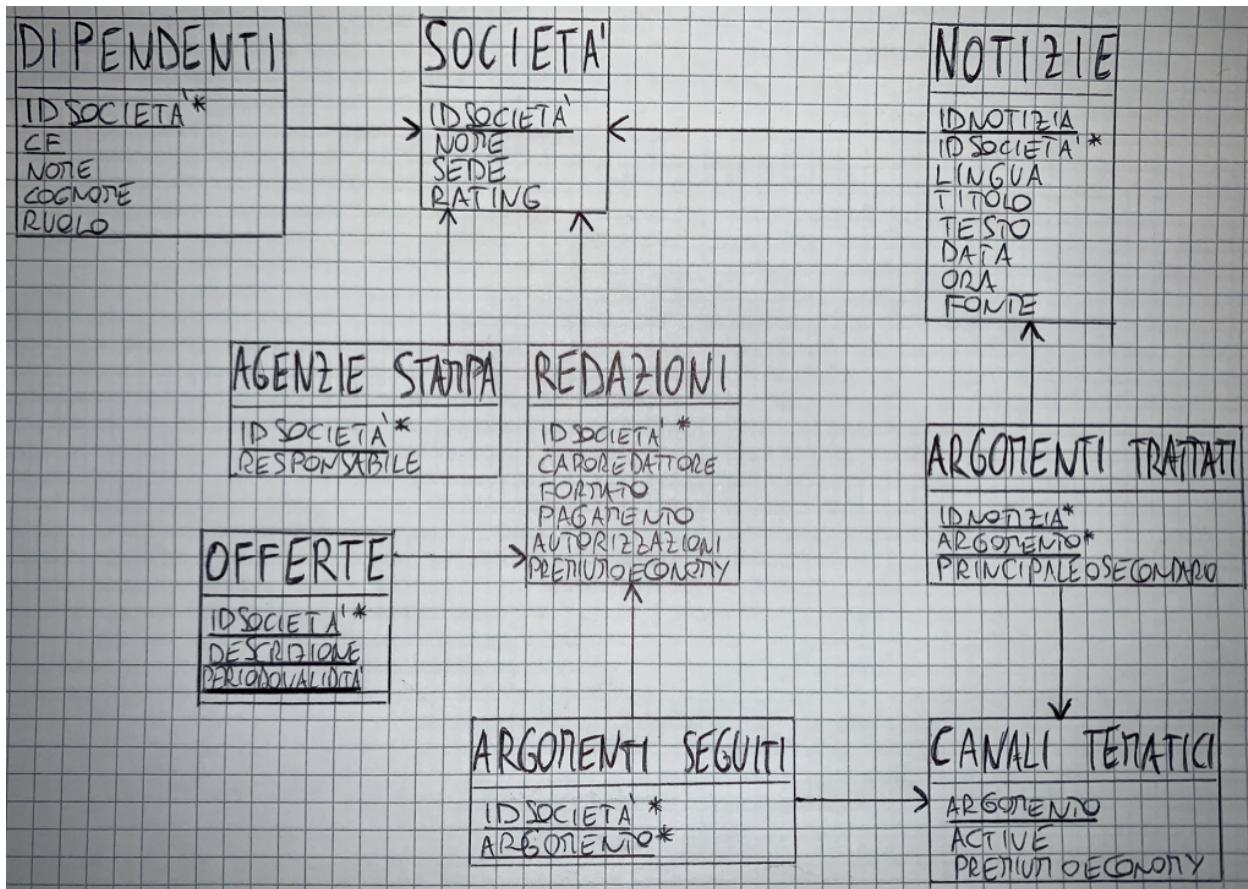


## 3 Schema logico relazionale

### 3.1 Rappresentazione testuale

- SOCIETÀ(IDsocietà, Nome, Sede, Rating)
- DIPENDENTI(IDsocietà\*, CF, Nome, Cognome, Ruolo)
- AGENZIE STAMPA(IDsocietà\*, Responsabile)
- REDAZIONI(IDsocietà\*, Caporedattore, Formato, Pagamento, Autorizzazioni, PremiumOEconomy)
- OFFERTE(IDsocietà\*, Descrizione, PeriodoValidità)
- ARGOMENTI SEGUITI(IDsocietà\*, Argomento\*)
- CANALI TEMATICI(Argomento, Active, PremiumOEconomy)
- ARGOMENTI TRATTATI(IDnotizia\*, Argomento\*, PrincipaleOSecondario)
- NOTIZIE(IDnotizia, IDsocietà\*, Lingua, Titolo, Testo, Data, Ora, Fonte)

### 3.2 Rappresentazione grafica



### 3.3 Dipendenze Funzionali

Oltre alle dipendenze funzionali banali (nelle quali abbiamo la chiave primaria delle tabelle a sinistra della freccia e tutti gli altri attributi della medesima tabella sulla destra), omesse per motivi di sintesi, sono state individuate le seguenti dipendenze funzionali:

- Nella Tabella DIPENDENTI:  
 $CF \rightarrow Nome, Cognome, Ruolo$
- Nella Tabella SOCIETÀ:  
 $Nome \rightarrow IDsocietà$   
 $Sede \rightarrow IDsocietà$
- Nella Tabella NOTIZIE:  
 $Titolo \rightarrow IDnotizia$   
 $Testo \rightarrow IDnotizia$   
 $ID_società, Data, Ora \rightarrow IDnotizia$

Esclusa la prima (quella nella Tabella DIPENDENTI), tutte le altre dipendenze funzionali sono in Forma Normale di Boyce-Codd poiché presentano tutte la chiave primaria della tabella a destra della freccia e di conseguenza gli attributi a sinistra della freccia formano una superchiave. Si noti che, a differenza di quanto accade per le persone, non sono contemplati casi di omonimia tra società in quanto si assume che non possano esserne fondate due con lo stesso identico nome.

## 4 Interrogazioni in SQL

### 4.1 Uso di proiezione, join e restrizione

Si restituiscano descrizione e periodo di validità delle offerte fatte a redazioni che hanno l'autorizzazione ad inviare notizie e fare operazioni amministrative.

```
SELECT O.Descrizione, O.PeriodoValidità  
FROM Offerte O JOIN Redazioni R USING IDsocietà  
WHERE R.Autorizzazioni = 'Inviare ed amministrare'
```

### 4.2 Uso di group by con having, where e sort

Restituire la data ed il numero di notizie in inglese diffuse nelle giornate in cui ne sono state pubblicate almeno due (sempre in inglese), ordinate per numero di notizie.

```
SELECT Data, COUNT(*) as NumNotizie  
FROM Notizie  
WHERE Lingua = 'Inglese'  
GROUP BY Data  
HAVING NumNotizie >= 2  
ORDER BY NumNotizie
```

### 4.3 Uso di join, group by con having e where

Si cerchino nome, sede e numero di notizie pubblicate dall'inizio del lockdown ad oggi dalle società con un punteggio di rating superiore a 3 ed almeno 6 pubblicazioni effettuate nel periodo considerato.

```
SELECT S.Nome, S.Sede, COUNT(*) as NumPubblicazioni  
FROM Società S JOIN Notizie N USING IDsocietà  
WHERE N.Data > '10032020' AND S.Rating > 3  
GROUP BY S.Nome, S.Sede  
HAVING NumPubblicazioni > 5
```

### 4.4 Uso di select annidata con quantificazione esistenziale

Restituire nome, sede e rating delle società che hanno pubblicato almeno una notizia sportiva nell'ultimo mese (come data odierna si prenda in considerazione quella indicata all'inizio del progetto, 29 gennaio 2021)

```
SELECT S.Nome, S.Sede, S.Rating  
FROM Società S  
WHERE EXISTS(SELECT *  
    .     FROM Notizie N JOIN Argomenti Trattati AT USING IDnotizia  
    .     WHERE S.IDsocietà = N.IDsocietà AND AN.Argomento = 'Sport' AND N.Data > '29122020')
```

## 4.5 Uso di select annidata con quantificazione universale

Si restituisca l'elenco dei canali tematici di tipo Economy i cui seguaci sono tutti di tipo Premium ed il numero di utenti ad essi iscritti.

```
SELECT CT.Argomento, COUNT(*) as NumIscritti
FROM Canali Tematici CT JOIN Argomenti Seguiti AS USING Argomento
WHERE CT.PremiumOEconomy = 'Economy' AND NOT EXISTS(SELECT *
    . . .
    FROM Redazioni R JOIN Argomenti Seguiti AS USING IDsocietà
    WHERE AS.Argomento = CT.Argomento AND R.PremiumOEconomy = 'Economy')
GROUP BY CT.Argomento
```

## 4.6 Uso di subquery di confronto quantificato usando una subquery

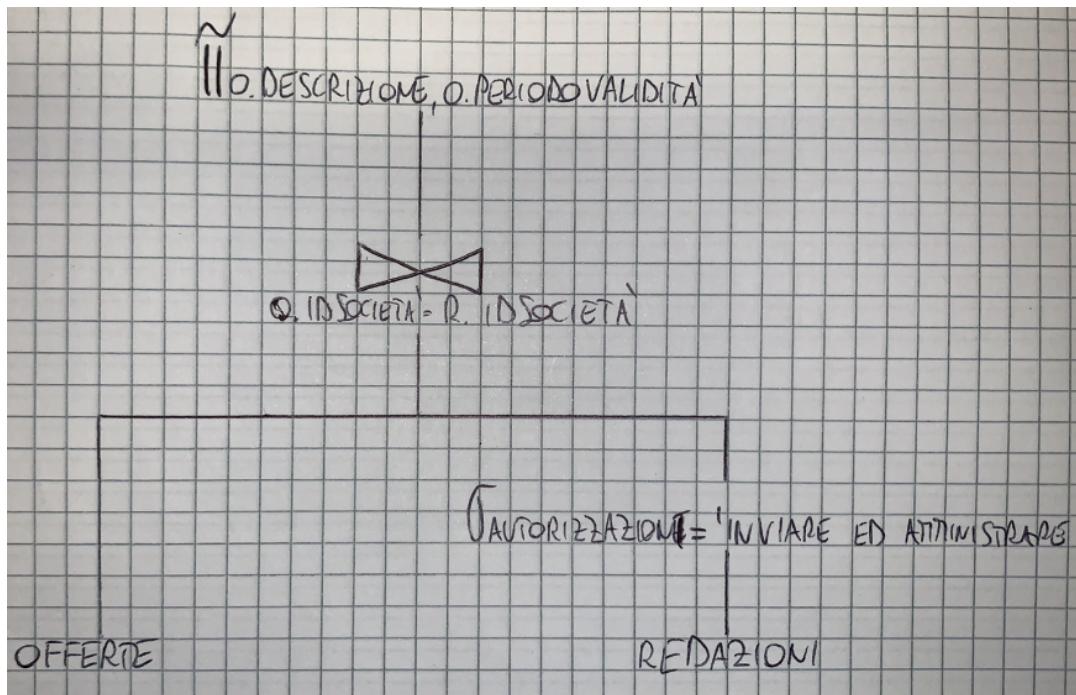
Trovare per ogni canale tematico non più attivo il titolo e la data dell'ultima notizia di quell'argomento ad essere stata pubblicata prima della sua disattivazione.

```
SELECT CT.Argomento, N.Data, N.Titolo
FROM Notizie N JOIN Argomenti Trattati AT ON N.IDnotizia = AT.IDnotizia JOIN Canali Tematici CT
ON AT.Argomento = CT.Argomento
WHERE CT.Active = FALSE AND N.Data >= ALL(SELECT N.Data
    . . .
    FROM Notizie N JOIN Argomenti Trattati AT USING IDnotizia
    WHERE AT.Argomento = CT.Argomento)
```

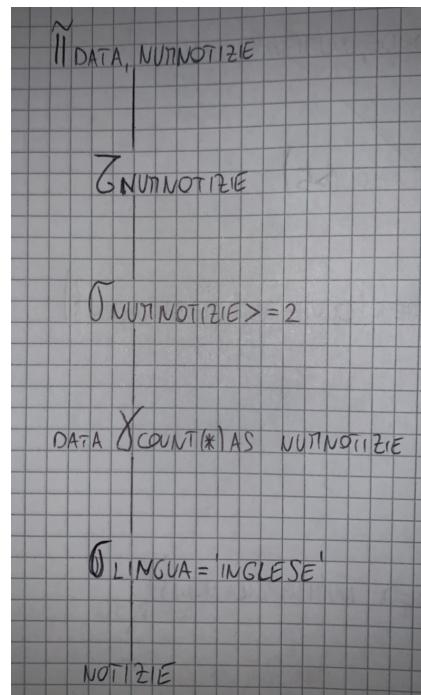
## 5 Piani di accesso

### 5.1 Piani logici

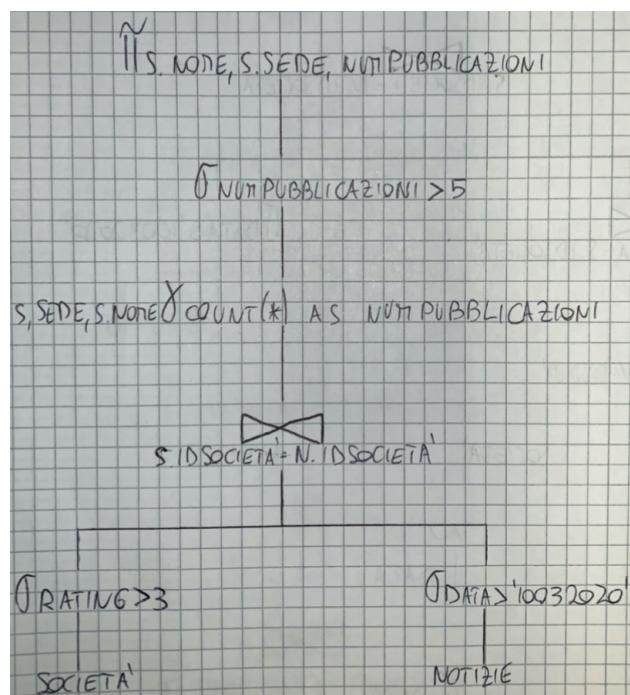
Query A



### Query B

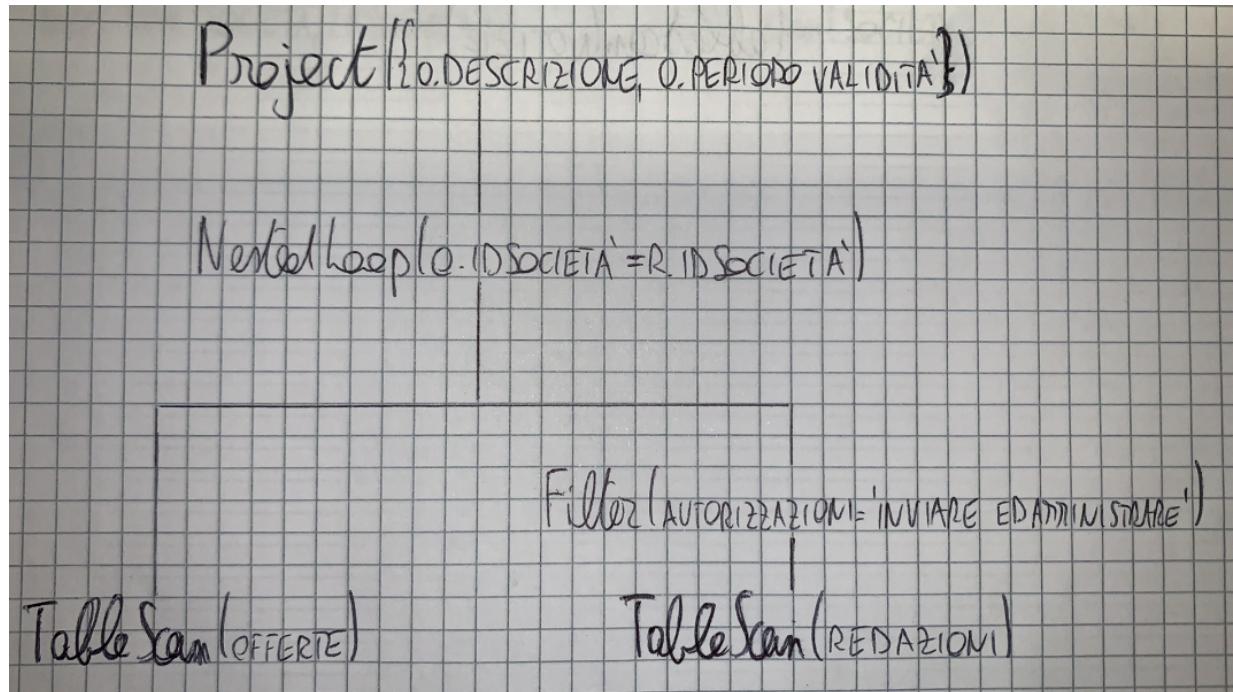


### Query C

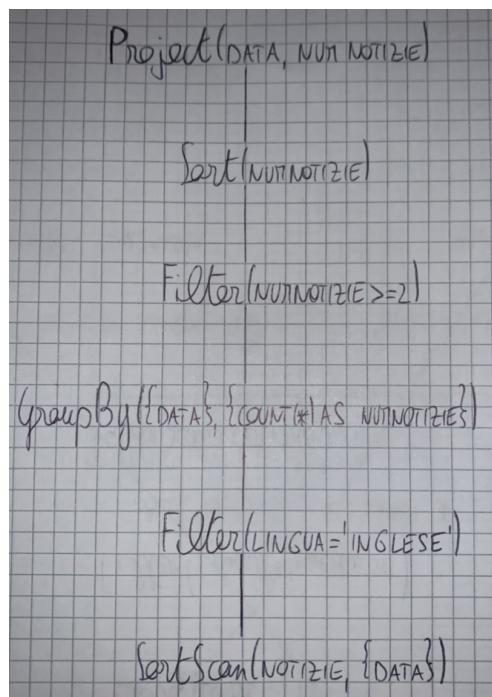


## 5.2 Piani fisici

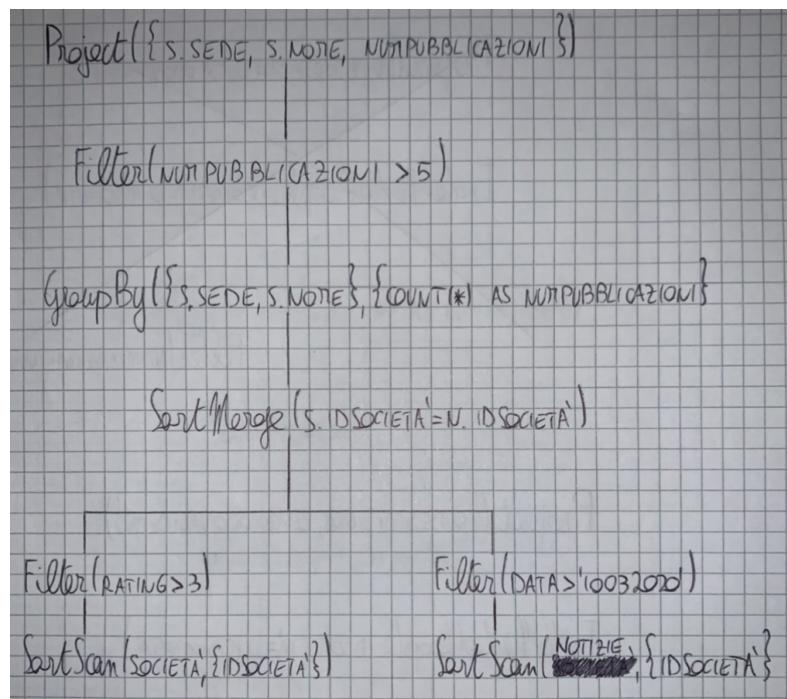
Query A



Query B



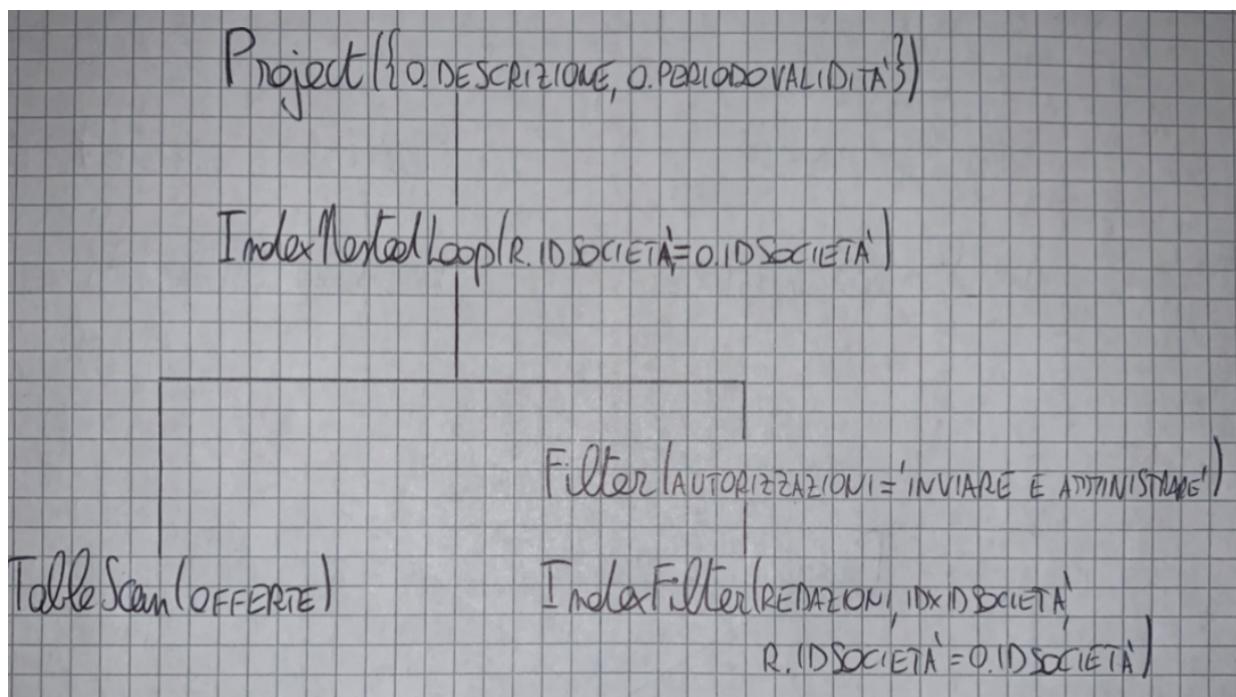
### Query C



In questo caso si è potuto evitare di inserire un Sort prima del GroupBy poiché sono stati sfruttati gli operatori SortScan e SortMerge. Infatti, nonostante essi abbiano ordinato gli elementi su IDsocietà, come anticipato precedentemente la corrispondenza è biunivoca sia tra IDsocietà e Nome che tra IDsocietà e Sede e dunque i record risultano già ordinati anche per Nome e Sede.

### 5.3 Piani fisici con indici

#### Query A



### Query B

```
Project([DATA, NUMNOTIZIE])  
Sort([NUMNOTIZIE])  
Filter([NUMNOTIZIE >= 2])  
GroupBy([DATA], {COUNT(*) AS NUMNOTIZIES})  
Sort([DATA])  
IndexFilter(NOTIZIE, IDXLINGUA, LINGUA='INGLESE')
```

### Query C

```
Project({S.NOTE, S.SEDIE, NUMPUBBLICAZIONI})  
Filter([NUMPUBBLICAZIONI > 5])  
GroupBy({S.NOTE, S.SEDIE}, {COUNT(*) AS NUMPUBBLICAZIONI})  
Sort([S.NOTE, S.SEDIE])  
IndexMatchedLoop([ID SOCIETA=S.ID SOCIETA])  
Filter([DATA > '10/03/2020'])  
IndexFilter(SOCIETA, IDX RATING,  
RATING > 3)  
IndexFilter(NOTIZIE, IDX ID SOCIETA,  
N.ID SOCIETA = S.ID SOCIETA)
```