

Secondo progetto di Social Computing 2020-2021

Francesco Bombassei De Bona (144665)
Andrea Cantarutti (141808)
Alessandro Zanatta (143154)

09 gennaio 2021

1 Introduzione

Il seguente elaborato espone il processo di modifica e dispiegamento di un task di Crowdsourcing, seguendo il procedimento e i metodi previsti dal framework Crowd Frame. In particolare, il lavoro è stato suddiviso nelle seguenti fasi:

- Selezione dei libri, dei questionari e delle dimensioni
- Modifiche apportate al codice
- Creazione degli HITS e dispiegamento del task
- Processo di reclutamento dei worker

2 Selezione dei libri, dei questionari e delle dimensioni

2.1 Selezione dei libri

Sulla base della consegna ricevuta, sono stati selezionati i seguenti tre libri:

- Assassinio sull'Orient Express (*Agatha Christie*)
- Le cronache di Narnia - Il leone, la strega e l'armadio (*C.S. Lewis*)
- Thus Spoke Zarathustra (*Friedrich Nietzsche*)

I libri, in particolare, differiscono in sulla base del genere letterario e della popolarità. Per ognuno di essi sono state selezionate tre differenti edizioni, delle quali una in italiano, una in inglese e una in formato eBook.

L'obiettivo preposto è, infatti, quello di capire le preferenze in fatto di lingua e supporto di lettura dei worker.

2.2 Scelta del questionario

Abbiamo deciso di introdurre le seguenti domande:

- Quanti anni hai? (classi di età, come 18-25, 26-35, ...)
- Qual è il tuo impiego attuale? (radio button con le più comuni posizioni lavorative)
- Qual è il tuo genere letterario preferito? (elenco dei principali generi letterari)
- Possiedi un eBook reader (Kindle, KoBo, ...) o utilizzi un'applicazione per la lettura di libri digitali? (risposta booleana si/no)
- Nelle tue giornate, quanta importanza ha la lettura? (scala Likert a quattro valori, da "Nulla" a "Fondamentale")
- Quanti libri leggi mediamente in un anno? (classi con numero di libri letti, come 0, 1-2, ...)

2.3 Scelte delle dimensioni a piacere

Abbiamo implementato, oltre alle quattro dimensioni richieste dalla consegna, le seguenti dimensioni a piacere:

- Hai letto questo libro? (scala nominale non dicotomica)
 - Sì, esattamente questa edizione
 - Sì, ma una edizione differente
 - No
- Sfoglieresti il libro vedendone la copertina? (scala nominale dicotomica)
 - Sì
 - No

3 Modifiche effettuate al codice sorgente di Crowd Frame

Admin login Si è implementato un più robusto sistema di autenticazione che prevede l'utilizzo di password di lunghezza arbitraria con hashing a 512 bit (SHA512). Non è stato implementato l'utilizzo di salt, tuttavia è sufficiente utilizzare password con sufficiente entropia per evitare qualsiasi tipo di attacco.

File modificati:

- `framework/src/app/components/loader/loader.component.ts`
- `framework/data/build/admin.json`

Aggiunta degli attributi degli HITS Si sono modificati, al fine di rendere disponibile la copertina del libro assieme alle informazioni richieste, entrambi i file sotto indicati.

Files modificati:

- `framework/data/build/document.ts`
- `framework/src/app/components/skeleton/skeleton.component.html`

Modifica del funzionamento della whitelist Al fine di rendere disponibile l'accesso *unicamente* ai worker preventivamente inseriti, è stata modificata l'implementazione e il significato della whitelist (nel file `framework/data/build/task/workers.json`).

File modificato: `framework/src/app/components/skeleton/skeleton.component.ts`

Utilizzo della lingua italiana Si è deciso di sostituire il linguaggio inglese a quello italiana al fine di evitare possibili bias dal punto di vista linguistico per i worker. Il task è stato infatti svolto da persone madrelingua italiana non tutte dotate di una sufficiente competenza nella lingua inglese. Si è prestata particolare attenzione ad una traduzione efficace ed espressiva, mantenendo un ottimo grado di fedeltà rispetto al testo inglese originario.

File modificati:

- `framework/src/app/components/skeleton/skeleton.component.ts`
- `framework/src/app/components/instructions/instructions.component.html`
- `framework/src/app/components/skeleton/skeleton.component.ts`
- `framework/src/app/components/instructions/instructions-dialog.component.html`

Modifiche alla formattazione delle HITS Sono stati variati degli aspetti minori nel layout delle HITS. In particolare, è stato implementato, nel box di presentazione del libro, una visualizzazione responsive, al fine di rendere accessibile il task anche da dispositivi mobili. Sono inoltre stati rimossi dei superflui elenchi puntati presenti nelle dimensioni (chiamati `dimensions-counter` nel codice).

File modificati:

- `framework/src/app/components/skeleton/skeleton.component.html`
- `framework/src/app/components/skeleton/skeleton.component.ts`

Correzione del bug relativo ai timestamp Si è notata la presenza di un errore relativo al calcolo dei timestamp durante la raccolta dei dati. L'errore è stato corretto manualmente.

File modificato: `framework/src/app/components/skeleton/skeleton.component.ts`

4 Creazione degli HITS e dispiegamento del task

Le HITS sono state create utilizzando il notebook Jupyter fornito, al quale sono state apportate lievi modifiche (fornito in allegato).

La configurazione del task è stata creata utilizzando l'apposito generatore fornito da Crowd Frame.

Ottenuti tutti i file di configurazione necessari per il dispiegamento del task, è stato effettuata una build utilizzando il codice sorgente opportunamente modificato come indicato al [paragrafo precedente](#). La build è successivamente stata dispiegata utilizzando gli script forniti sui bucket Amazon S3, preventivamente configurati.

Il task dispiegato è [qui disponibile](#).

5 Processo di reclutamento dei worker

Al fine di ottenere un insieme di dati maggiormente significativo, è stato deciso di aumentare il numero di worker coinvolti nel task. Per permettere un reclutamento più efficiente ed efficace, è stato configurato un sistema che comunica, tramite invio di mail, le informazioni necessarie allo svolgimento del task, ovvero il token di input da utilizzare (scelto uniformemente fra i sei token disponibili) e il link completo di worker ID univoco.

In particolare, sono stati implementati degli script che si occupano dell'invio della mail (previo inserimento dell'indirizzo destinatario), dell'aggiornamento della whitelist e della sincronizzazione della stessa con il bucket S3.

Ulteriori script permettono, in maniera semplificata, il download dei dati raccolti fornendo funzioni atte a monitorare il task dispiegato.

I file relativi sono disponibili nella cartella `pyDistribution`.