

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*)
- Così parlò Zarathustra (*Friedrich Nietzsche*)

I libri, in particolare, differiscono 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 cogliere 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 sì/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, ...)

Una volta proposte al worker, le domande permettono di inquadrarlo dal punto di vista anagrafico, lavorativo e culturale, individuandone altresì le preferenze personali. Tali parametri, combinati alle ulteriori risposte raccolte, permettono una più specifica ed efficace analisi dei dati.

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 È stato 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.

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 visualizzabili tutte le informazioni relative al libro in analisi (inclusa la copertina), 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 selezionati, è stata modificata l'implementazione e il significato della whitelist (situata 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 italiano al fine di evitare possibili bias dal punto di vista linguistico. Il task è, infatti, stato 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 stata implementata, 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 (denominati `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 fase di testing precedente al dispiegamento del task. 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 **Jupyter Notebook** fornito, al quale sono state apportate sostanziali modifiche (disponibile nella cartella `pyHITS`).

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, è stata 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](#).

I token per l'accesso agli HITS sono i seguenti:

- ZPCUSGSGQL
- MNYRTLHFQO
- AOAACVVTRL
- TIATJWRNJX
- VTBQGMDFRI
- ZBTWBMEEUW

5 Processo di reclutamento dei worker

Con l'obiettivo di ottenere un insieme di dati maggiormente significativo, è stato deciso di aumentare il numero di worker coinvolti nel task. È stato configurato un sistema che comunica, tramite invio di una e-mail, le informazioni necessarie allo svolgimento del task. Per ogni worker reclutato, nell'ordine:

1. viene generato un ID univoco
2. un token di input è selezionato uniformemente fra i 6 disponibili
3. viene composto il link di accesso al task dispiegato su S3

Sia il link di accesso che il token vengono, quindi, inviati al worker, invitandolo a prendere parte allo svolgimento del task.

In particolare, sono stati implementati 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 forma semplificata, il download dei dati raccolti e il monitoraggio del task dispiegato.

I relativi file sono disponibili nella cartella `pyDistribution`.