

Progetto di Social Computing

Crowdsourcing con Crowd Frame parte 1

Parata Loris (144338)
Arzon Francesco (142439)
Dal Fabbro Lorenzo (142300)

Indice

1	Crowd Frame	2
1.1	Introduzione	2
1.2	Progettazione del task	2
1.2.1	Obiettivo del task	2
1.2.2	Creazione del task	2
1.2.3	Le istruzioni inerenti al task	2
1.2.4	Questionario	3
1.3	Il task	3
1.3.1	Le dimensioni	3
1.3.2	Variazione del file HTML	4
1.4	Struttura delle HITS	5
1.4.1	Generazione dei token	5
1.5	Modifiche e impostazioni del task	5

Capitolo 1

Crowd Frame

1.1 Introduzione

Questo secondo progetto di Social Computing si suddivide in due parti, la fase di progettazione di un task di crowdsourcing e la fase di raccolta dati e analisi dei dati ottenuti da un piccolo gruppo di workers. In questa relazione si documenta la prima parte dell'esperimento di crowdsourcing, la progettazione del task richiesto dalla traccia.

1.2 Progettazione del task

1.2.1 Obiettivo del task

La traccia richiede di creare sei HIT differenti riguardanti tre libri, per ognuno dei quali si devono individuare tre edizioni differenti. Ogni HIT contiene tre edizioni, una per ogni libro di riferimento, per le quali ogni worker deve rispondere a 6 dimensioni differenti, quattro imposte dalla traccia e due scelte dal gruppo di progetto. Per verificare un'eventuale preferenza di alcune tipologie di formato dei libri, abbiamo deciso di selezionare delle edizioni differenti, ma selezionandone una in formato Paperback, una in formato Hardcover ed una in formato Kindle/Ebook.

Lingua del task

Abbiamo deciso di sviluppare l'intero progetto in lingua inglese, in modo da rendere il progetto il più realistico e simile ai task presenti sulla piattaforma di crowdsourcing di MTurk. Di conseguenza abbiamo scelto dei libri le cui edizioni sono scritte in lingua inglese.

1.2.2 Creazione del task

La creazione del task è stata effettuata mediante l'utilizzo dell'interfaccia grafica del framework **Crowd Frame** per la semplicità con cui si presenta la GUI. Anche se le modifiche successivamente sono state effettuate a livello di codice, vista la mancanza della possibilità di salvataggio della configurazione precedente.

1.2.3 Le istruzioni inerenti al task

Le istruzioni accolgono il worker e chiariscono l'obiettivo del task. Chiarendo ogni eventuale dubbio sull'utilizzo dei dati raccolti e lo avvisa che non è strettamente necessario conoscere uno specifico libro o una sua specifica edizione per svolgere il task. Al worker viene indicato che è possibile rispondere alle domande anche con le sole informazioni sommarie mostrate durante il task, dando un'opinione personale riguardo al libro proposto.


1.2.4 Questionario

Successivamente il worker viene sottoposto ad un insieme di questionari, composto da sette questionari standard ed una domanda CRT. Questi quesiti ci permettono di conoscere in parte il background letterario del worker e le sue abitudini nell'acquisto di un libro. Inoltre queste informazioni ci permettono anche di stabilire in parte la qualità delle risposte che verranno date alle dimensioni proposte. Mentre la domanda CRT ci permette di valutare se l'utente risponde in maniera "istintiva".

- How much time do you usually spend reading?
- How many books do you usually read during the year?
- What's your favorite literary genre?
- Do you prefer books in the original language or translated?
- What format do you buy most of your books in?
- Where do you usually find yourself reading?
- Based on what you choose a book to buy?
- A book and a pen cost \$1.10 in total. The book cost \$1.00 more than the pen. How much does the pen cost?

1.3 Il task

Please study the following Book



Title: Fahrenheit 451 (Paperback)
Author: Ray Bradbury
Number of pages: 160
Editor: Flamingo
Price: 7.68\$
Year of publication: 2008

Instructions

Please, do the following steps:

- Answer all of the following questions regarding the featured edition.
- At question 5 indicate the level of agreement with which you agree on the underlying statement by choosing from the 5 possible options.
- At question 6 The response must have one of the following forms: 'I have read this book' or 'I have not read this book' or 'I have read this edition [https://www.amazon.com/Three-Musketeers-Tim-Curry / ...]'

Would you buy this book?

Select the option that indicates whether you would ever purchase this edition.

a. ☐ Yes (Yes, I would buy this book) ☐ No (No, I would not buy this book)

Does the price seem adequate to you?


Select the option that indicates whether you think the price is appropriate for this edition.

a. ☐ Yes (I find that the price is adequate) ☐ No (I find that the price is not adequate)

Indicate how adequate it seems to you

(Little = 1, Much = 5)
Selected value:

a. ☐



Title: Fahrenheit 451 (Paperback)
Author: Ray Bradbury
Number of pages: 160
Editor: Flamingo
Price: 7.68\$
Year of publication: 2008

What is your impression of this edition?

Please indicate your impressions of this edition based on your knowledge.

a.

Select the option that indicates how much you agree with the following statement.

'I would recommend this edition to another person'

☐ Completely agree (I would recommend it to anyone)
☐ Generally agree (I would recommend it)

a. ☐ I don't know (I don't have a clear opinion on this book)
☐ Generally disagree (I would not recommend it)
☐ Completely disagree (I would not recommend it to anyone)

If you've read any edition of this book, indicate a url to identify it

Indicate whether you have read this specific edition or not. In case you have read this book, but a different edition indicates which one by indicating the URL of the specific edition.

a.

(*) you have to fill each field to proceed

(a) Layout del task parte 1

(b) Layout del task parte 1

Figura 1.1: Layout del task

1.3.1 Le dimensioni

Ad ogni worker verranno presentate tre schermate con il layout della **Figura 1.1**

Oltre alle quattro dimensioni imposte dalla traccia:

- **Dimensione 1:** Would you buy this book?
- **Dimensione 2:** Does the price seem adequate to you?
- **Dimensione 3:** Indicate how adequate it seems to you
- **Dimensione 4:** What is your impression of this edition?

Abbiamo deciso di aggiungere :

- **Dimensione 5:** Select the option that indicates how much you agree with the following statement:
"I would recommend this edition to another person"
- **Dimensione 6:** If you've read any edition of this book, indicate a URL to identify it.

Dimensione 5

La dimensione 6 è stata introdotta per verificarne la coerenza con la risposta della dimensione 1, in modo da avere un minimo di controllo ,visto il mancato utilizzo delle gold question.

L'utilizzo delle gold questions non è stata richiesta dalla traccia, ed essendo non ancora implementato nella GUI il controllo delle gold questions, era necessario modificare il codice del framework. La verifica di coerenza verrà effettuata nella fase di analisi. Il worker ha a disposizione 5 possibili opzioni, che vanno dal pienamente d'accordo al completamente in disaccordo, con la possibilità di prendere una posizione neutrale relativa all'affermazione sottoposta a quesito.

Dimensione 6

La dimensione 5 è stata introdotta per verificare se il worker abbia letto una qualsiasi edizione del libro ed eventualmente individuarne delle versioni alternative preferite dai workers. Nella situazione in cui un worker non abbia mai letto quel libro si possono valutare diversamente la qualità delle risposte. La risposta deve essere strutturata secondo le modalità descritte nelle istruzioni del task ed un relativo controllo che richiede una risposta di almeno 4 parole. Valore individuato per scrivere la frase minima "I have read it", come scritto nelle opzioni possibili nelle istruzioni.

1.3.2 Variazione del file HTML

Per poter ottenere quel layout è stato necessario effettuare delle modifiche al file `skeleton.component.html` tra le righe 187 e 211 (**Figura 1.2 a**):

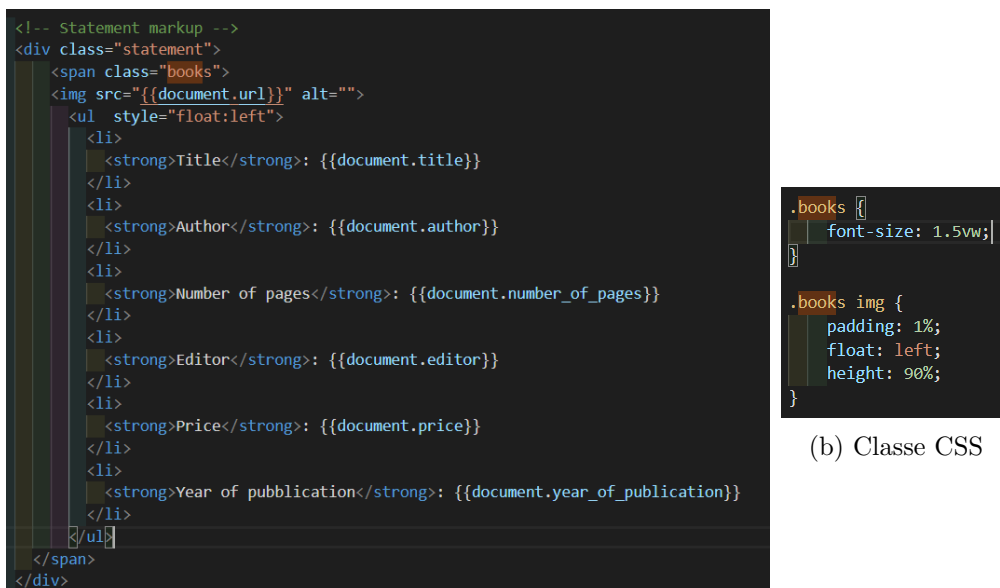


Figura 1.2: Codice HTML e CSS

In modo da visualizzare a sinistra l'immagine ed a destra gli attributi che descrivono il libro. Le informazioni riguardanti il libro rimarranno sempre visibili al worker durante lo svolgimento del task, in modo da facilitarne lo svolgimento. Inoltre è stata aggiunta una classe CSS nel file **skeleton.component.scss**, come si può vedere in **Figura 1.2 b**, in modo da poter rendere un minimo responsive l'immagine della copertina del libro ed i relativi attributi.

1.4 Struttura delle HITS

1.4.1 Generazione dei token

Ogni HIT è accessibile ad un worker inserendo un Token che lo identifica univocamente e rilasciando nel caso di esito positivo un relativo token che permette all'utente di esser pagato dall'eventuale piattaforma di crowdsourcing. I token e parte delle HIT sono stati generati mediante il file **hits.tokens.generator.ipynb**, ricevendo in input file csv contenente gli attributi che andranno a identificare ogni edizione del rispettivo libro. Adattando il codice iniziale a questo task, che ha previsto la rimozione del codice relativo alle gold question e l'utilizzo degli attributi necessari per descrivere ogni libro. Infine verranno generati 9 oggetti contenente:

- **token_input**: token relativo alla hit che verrà assegnato al worker per accedervi.
- **token_output**: token relativo alla hit che verrà assegnato al worker nel caso in cui svolge correttamente il task.
- **unit_id**: per identificare la hit contenente tutti i documenti sottoposti ad esame al worker.

E di seguito anche tutti i documenti che verranno valutati dai workers, con i relativi attributi, .

- **title**: indica il titolo del libro ed il formato del libro
- **author**: indica l'autore/gli autori del libro o eventuali traduttori.
- **number of pages**: indica il numero di pagine di cui è composto il libro
- **editor**: indica l'editore che ha pubblicato il libro
- **url**: contiene l'URL che fa riferimento all'immagine della copertina del libro
- **year of publication**: anno di pubblicazione dell'edizione di riferimento

Di questi 9 oggetti ne verranno utilizzati 6, corrispondenti ai relativi HIT necessari. Manualmente sono stati composti i raggruppamenti dei documenti (libri) di ogni HIT. Sono stati creati 3 gruppi contenente una sola tipologia di libro (hardcover, ebook, paperback) per poter verificare eventuali preferenze nella fase di analisi. I restanti 3 sono stati creati in maniera casuale mantenendo un ordine casuale.

I token generati con la relativa corrispondenza sono memorizzati nel file **data\build\mturk\token.csv**

1.5 Modifiche e impostazioni del task

Nella sezione relative all'impostazione del task sono stati inseriti il nome del task e della relativa batch. Abbiamo stabilito che un worker ha a disposizione un solo tentativo ed un tempo minimo necessario a leggere un questionario o documento di **5 secondi**. Sarebbe stato ottimale poter distinguere queste due tempistiche. E' stato anche impostato un messaggio relativo alla casistica in cui un worker abbandona la sessione durante lo svolgimento e tenta di accedervi nuovamente.

Per accedervi al task è possibile effettuarlo accedendo alla sandbox dei worker di MTurk ed inserire la parola books nella barra di ricerca. Oppure direttamente accedendo tramite questo url: <https://parata-crowd-deploy-eu.s3.eu-central-1.amazonaws.com/Libri-task-1/Batch-libri-1/index.html>