UNIVERSITÀ DI PISA

Corso di Laurea in Informatica Umanistica

RELAZIONE

Raccolta di relazioni semantiche attraverso crowdsourcing

Candidato: *Nicola Biagioni*

Relatore: *Prof. Alessandro Lenci*

Correlatore: *Prof.ssa Maria Simi*

Anno Accademico 2011-2012

Indice

Introduzione

Il linguaggio è lo strumento principale di cui l'uomo dispone per creare e comunicare contenuti complessi differenti nella forma e nel significato. Per anni si è cercato di indagare e individuare il funzionamento del linguaggio, la sua struttura e il rapporto che instaura con altre facoltà cognitive dell'uomo. La linguistica, principale disciplina scientifica con il compito di studiare l'organizzazione del linguaggio, ha cercato, attraverso il dialogo con altre discipline come la filosofia, la psicologia, le scienze umane e quelle cognitive etc., un metodo generale attraverso il quale indagare e ricercare informazioni relative allo scopo.

Il progresso tecnologico e l'intensificarsi di questi rapporti tra settori scientifici differenti hanno portato alla luce aree di ricerca e di lavoro tipicamente interdisciplinari come la linguistica computazionale punto di incontro tra l'informatica e l'analisi linguistica.

In linguistica computazionale il processo di trattamento automatico del linguaggio umano viene denominato *Natural Language Processing* – NLP.

Come gli studi e le teorie linguistiche hanno fatto progressi nell'individuare le abilità del linguaggio naturale, così in particolare il lessico di una lingua si è ritagliato un ruolo di importanza sempre maggiore all'interno della ricerca.

Questo ha posto in evidenza due problematiche:

- Non esiste ancora una teoria completa e pienamente soddisfacente su come il lessico sia strutturato internamente e come le informazioni lessicali siano rappresentate.
- Visto che il lessico deve essere collegato al sistema concettuale, non c'è accordo nello stabilire in che modo le informazioni concettuali sono rappresentate.

In base a quanto detto, è però opinione unanime che le relazioni semantiche tra le parole - antonimia, sinonimia, iponimia e simili - sono estremamente rilevanti per individuare la struttura delle informazioni lessicali o concettuali.

Con questo proposito l'indagine svolta in questa tesi si propone di spiegare alcuni dei metodi impiegati nel raccogliere informazione relative alle proprietà che caratterizzano le relazioni semantiche. Viene descritto infatti il procedimento che ha portato alla creazione di un "*benchmark task oriented*" per la valutazione di modelli computazionali semantici attraverso il quale è stato raccolto un campione di relazioni semantiche da utenti nella rete tramite le tecnologie del crowdsourcing. Nello specifico viene spiegato le metodologie impiegate nel creare *Task* che richiedono di generare sinonimi, antonimi e iperonimi relativi ad una serie di stimoli e le analisi effettuare sulle risposte ottenute.

Struttura della tesi

In questo lavoro si vanno, innanzitutto, ad introdurre i concetti principali coinvolti nello studio condotto: relazioni semantiche, modelli computazionali semantici, semantica distribuzionale, crowdsourcing (Capitolo 1).

Nel capitolo 2 viene descritto il procedimento che ha portato alla creazione di *Task* attraverso la piattaforma di Amazon Mechanical Turk (AMT). Dopo una panoramica nella quale vengono spiegati i comandi di base all'interno di AMT, viene descritto il procedimento attraverso il quale sono stati estratti i dati da impiegare nell'esperimento. Viene infine presentato sia il procedimento tramite il quale sono stati creati *Task* specifici allo studio condotto, sia il processo di normalizzazione dei dati ottenuti.

Infine nel capitolo 3 sono presenti i risultati di un'analisi statistica effettuata calcolando media e deviazione standard delle risposte relative ai *Task* creati con AMT.

Capitolo I
Relazioni semantiche e modelli computazionali

1.1 Le relazioni semantiche

Ogni lingua possiede un lessico ovvero un insieme di parole, ognuna delle quali viene individuata da una forma e un significato.
Questa distinzione che caratterizza una parola è un passaggio fondamentale nel comprendere le proprietà relative alle relazioni semantiche.
In questo caso nonostante la forma sia differente, il significato è il medesimo e si parla quindi di sinonimia.
Le relazioni semantiche, in generale, vengono definite come le associazioni che esistono sia tra i significati delle parole (relazioni semantiche a livello di parola), sia tra i significati delle frasi (relazioni semantiche a livello di frase) .
Nel corso degli anni in particolare la linguistica teorica ha avuto un ruolo fondamentale nel classificare e nel definire in maniera dettagliata e completa le principali caratteristiche che le individuano.
In generale le principali relazioni a livello di parola sono:

- *Sinonimia*: (A è equivalente a B). Esiste tra due o più parole che hanno lo stesso significato e appartengono alla stessa parte del discorso. Le parole che fanno parte di questo tipo di relazione semantica sono detti sinonimi.
  Il termine *duplicato* è sinonimo di *copia*.
- *Antonimia*: (A è l'opposto di B). Indica la relazione esistente tra due lessemi di significato opposto. Le parole che fanno parte di questa relazione si dicono antonimi e normalmente fanno parte entrambi della stessa categoria grammaticale. Il termine *brutto* è l'antonimo di *bello*.
- *Omonimia*: (Due concetti, A e B, sono espressi dallo stesso simbolo). E' la relazione che esiste tra due o più parole le quali fanno parte della stessa categoria grammaticale, vengono scritte e pronunciate nello stesso modo, ma esprimono significati diversi.

Ad esempio *vite* (oggetto appuntito) e vite (pianta). Si distingue dalla polisemia in quanto i diversi significati di un lessema sono uniti in un'unica forma solo per caso. Nella polisemia invece i diversi significati della parola sono correlati etimologicamente e semanticamente. Ad esempio *capitale* (città) e *capitale* (somma di denaro) sono due termini omonimi.

- *Iperonimia*: (A è un tipo di B). Indica una specifica relazione semantica tra due termini, uno dei quali è detto "iperonimo" e include nel proprio significato quello dell'altro termine, detto "iponimo". Si tratta di un rapporto gerarchico tra le parole. Nel caso di "macchina" B corrisponde a "veicolo", così come mobile è iperonimo di sedia, armadio ecc. L'inverso dell'iperonimia è l'iponimia
- *Meronimia*: (A è una parte di B). Un meronimo indica un costituente o un membro di qualcosa. Ad esempio *dito* è un meronimo di *mano*, così come *ruota* è meronimo di *automobile*.
- *Polisemia*: in semantica indica la proprietà che una parola ha di esprimere più significati. Il significato del termine, nell'accezione comune, si è esteso anche ad altri segni: non più solo alla parola, ma anche all'immagine, al suono, ecc.
  Il termine *volume* può indicare sia l'intensità di un suono, sia un libro, sia la misura di un corpo solido.

Queste relazioni hanno catturato l'interesse di un numero elevato di ricercatori che operano nello studio della semantica lessicale e nelle scienze cognitive.
Così non solo linguisti, psicologi e filosofi, ma anche informatici , ingegneri, logici, neuro-fisiologi etc., si trovano sempre più spesso a lavorare su temi comuni di ricerca. Questo, da un lato, ha prodotto un numero elevato di ricerche interessanti, relative alle relazioni semantiche, analizzate secondo prospettive teoriche e metodologiche differenti. (Murphy, 2003)

1.2 Modelli computazionali semantici

Nel settore della linguistica computazionale. Lo studio delle proprietà relative alle relazioni semantiche tra le parole ha consentito di organizzare numerosi dati lessicali in maniera differente. Sono state create ontologie e lessici computazionali specificatamente finalizzati a rappresentare il contenuto semantico delle parole attraverso le loro relazioni semantiche, come in *WordNet.*

Nell'ambito del *Natural Language Processing* sono stati elaborati modelli per l'analisi del significato chiamati "*Distributional Semantic Model*" - *DSM* o anche *semantic spaces*, *vector-space models* ecc. Questi sono modelli computazionali che costruiscono rappresentazioni semantiche contestuali a partire dai dati di un *corpus*.

1.3 Semantica distribuzionale

Questi metodi o modelli semantici fanno parte della Semantica Distribuzionale secondo la quale la distribuzione statistica delle parole nei contesti linguistici, gioca un ruolo chiave nel comprenderne il comportamento semantico.
In generale il lessico viene concepito come uno spazio metrico i cui elementi, ovvero le parole, sono separati da distanze che dipendono dal loro grado di similarità semantica. Quest'ultima viene misurata attraverso le distribuzioni statistiche di co-occorrenza nei testi, assumendo come principio epistemologico fondamentale la cosiddetta ipotesi distribuzionale, secondo la quale due parole sono tanto più semanticamente simili, quanto più tendono a ricorrere in contesti linguistici simili (Miller and Charles, 1991) .
Questa ipotesi ha guadagnato posizioni soprattutto negli ultimi anni grazie in particolar modo alla disponibilità di corpora testuali di grandi dimensioni e di tecniche statistiche e informatiche più sofisticate per estrarre gli schemi distribuzionali dei lessemi.

La nozione di spazio semantico si basa su una semplice analogia con lo spazio geometrico. Come ciascun punto dello spazio è definito da un vettore di n numeri che rappresentano le sue coordinate rispetto a n assi cartesiani (le dimensioni dello spazio), così il contenuto semantico di una parola è rappresentato dalla sua posizione in uno spazio definito da un sistema di coordinate, determinato dai contesti linguistici in cui la parola può ricorrere. Formalmente, uno spazio semantico di parole è definito dalla quadrupla < T, B, M, S >.

T è l'insieme delle parole target che formano gli elementi che popolano lo spazio e di cui questo fornisce una rappresentazione semantica. B è la base che definisce le dimensioni dello spazio e contiene i contesti linguistici rispetto ai quali viene valutata la similarità distribuzionale delle parole target. M è una matrice di co-occorrenza che fornisce una rappresentazione vettoriale di ogni parola in T.

L'ipotesi distribuzionale è stata così tradotta in modelli computazionali per la costruzione di spazi semantico-lessicali, che sono stati applicati nella simulazione di diversi aspetti della competenza semantica.

Quindi alla base di tutto risiede l'idea che due parole che tendono a combinarsi con elementi linguistici simili vengono a collocarsi anche in punti dello spazio semantico più vicini rispetto a quelli occupati da parole che invece si distribuiscono in maniera diversa nel testo. Questa assunzione è tipicamente formalizzata rappresentando ogni parola come un vettore a *n dimensioni*, ciascuna delle quali registra il numero di volte in cui la parola compare in un certo contesto definito dalla base B.

Ogni parola target corrisponde, dunque, a una riga della matrice M, le cui colonne corrispondono invece agli elementi in B. Nel caso più semplice, il valore di una cella della matrice è equivalente alla frequenza di co-occorrenza della parola in un dato contesto: nell'esempio riportato nella Tabella 1, la parola presidente ricorre 7 volte nel contesto di repubblica, nel quale invece non compaiono mai né torta né panino.

	Dire	Mangiare	Aprire	Pensare	Repubblica	Gustoso
Ministro	6	2	5	4	1	0
Presidente	10	3	2	5	7	0
Torta	0	4	2	0	0	3
Panino	0	7	0	0	0	1

Tabella 1 – matrice di co-occorrenza tra parole.

L'ultimo elemento che definisce la struttura dello spazio semantico è la metrica S che misura la distanza tra i suoi punti nello spazio. Per determinare la posizione di due parole, è necessario comparare i loro vettori rispetto a tutte le dimensioni che li costituiscono.
Maggiore è il numero di dimensioni nelle quali due vettori presentano valori simili, maggiore è la loro vicinanza nello spazio e – data l'ipotesi distribuzionale – la similarità semantica delle parole corrispondenti (Lenci, 2009).

1.4 Metodi standard nel valutare DSM

In *NLP* nel valutare un sistema o un algoritmo si distinguono due procedimenti base:

- *Intrinsic evaluations*, ovvero testare un sistema in sé.
- *Extrinsic evaluations*, misurare la performance del sistema attraverso l'utilizzo di alcuni *task* o applicazioni.

Ad esempio l'*intrinsic evaluations* di un "*dependency parser*" misurerà l'accuratezza di quest'ultimo nell'identificare specifiche relazioni sintattiche, mentre l'*extrinsic evaluations* si sofferma sull'impatto del *parser* nello svolgere alcuni *task* come "*question answering*" o "*machine traslation*".
Gli approcci recenti impiegati nel valutare le performance di un *DSM* si basano sul secondo dei due metodi e vengono definiti anche "*Task oriented*" poiché utilizzano task semantici specifici come individuare sinonimi, riconoscere analogie, classificare parafrasi, etc.
Misurare le prestazioni di un modello semantico sulla base di questi task, rappresenta un valido sistema per testare la loro abilità nel catturare i significati lessicali. Fino ad ora però, i "*task-oriented benchmarks*", impiegati nella semantica distribuzionale, non sono stati creati appositamente con l'intento di valutare i DSM.
Ad esempio un benchmark standard come il "*TOEFL synonym detection task*" è stato utilizzato con l'intento di testare la competenza di alcuni studenti nell'inglese come seconda lingua e non con l'obbiettivo di comprendere la struttura delle loro rappresentazioni semantiche.

Per questo motivo i benchmark esistenti devono essere modificati per ottenere una visione completa sotto ogni punto di vista delle capacità di un DSM nell'individuare aspetti della semantica lessicale ed avere la possibilità di eseguire test migliori sulla conoscenza acquisita dai modelli.

Esistono tre tipi specifici di benchmarks adottati attualmente nelle valutazioni di un DSM e ognuno di essi propone sfide differenti da sottoporre ai modelli
(Lenci, Baroni, 2011).

- "*Synonym detection task*".
- "*Semantic similarity rating set*".
- "*Concept categorization task*".

Della prima delle tre categorie, già visto in precedenza, il TOEFL task è probabilmente il benchmark più utilizzato nel campo della semantica distribuzionale e fu introdotto in linguistica computazionale da Landauer e Dumais nel 1997.
Il test prevede 80 domande a scelta multipla composta da 4 termini uno dei quali sinonimo di una parola (nome, verbo, aggettivo o avverbio) posta all'inizio della domanda. Il compito del sistema è quello di individuare il sinonimo corretto tra le quattro possibili opzioni.
La sinonimia è la sola relazione semantica presente all'interno del TOEFL task.

Il "*WordSim 353 data set*" è un esempio ampiamente utilizzato di "*Semantic similarity rating set*". Viene chiesto al sistema di valutare una serie di 353 coppie di parole sulla base di una scala di similitudine e calcolare una media delle valutazioni per ogni coppia analizzata.

Recentemente un metodo differente che ha guadagnato consensi nel valutare *DSMs,* è il "*concept categorization task*" il quale prevede che un modello raggruppi un set di nomi che esprimono concetti di base, all'interno di categorie standard.

Un esempio è identificato dal dat set di Almuhareb-Poesio (Almuhareb, 2006) - AP - un set di 402 concetti i quali devono essere raggruppati in 21 classi ognuna delle quali deve contenere tra i 13 e i 21 nomi.
Nonostante questo genere sia interessante in quanto simula uno degli aspetti base della cognizione umana, è limitato, come nel caso del TOEFL task, ad un solo tipo di relazione semantica - "*discovering coordinates*". Secondo questa relazione Y è un termine coordinato di un altro termine X se X e Y hanno un iperonimo in comune.

L'obiettivo del nostro progetto è stato quello di creare un "*benchmark task oriented*" per la valutazione di *DSM*, raccogliendo relazioni semantiche lessicali da soggetti attraverso il metodo di *crowdsourcing.* Sono stati così progettati task che richiedono ai soggetti di generare sinonimi, antonimi e iperonimi di una serie di parole stimolo.

1.5 Tecnologie di Crowdsourcing

Il termine *crowdsourcing* è stato coniato nel 2006 da Jeff Howe il quale lo ha utilizzato in un articolo su *Wired*.
Secondo la definizione data dallo stesso Howe, fare crowdsourcing significa appaltare un compito ad un vasto ed indefinito gruppo di persone (crowd significa folla), tramite una chiamata aperta a cui chiunque può rispondere.
Il termine crowdsourcing definisce, quindi, un modello di business basato sul lavoro distribuito attraverso il web. Un'azienda o un'istituzione richiede lo sviluppo di un progetto, di un servizio o di un prodotto ad un insieme distribuito di persone non già organizzate in una comunità virtuale .
Il crowdsourcing ha avuto la sua genesi nel movimento dei Software Open Source (Linux), ma ormai è utilizzato in diversi ambiti, come il marketing commerciale (Zooppa) e marketing research, il settore amministrativo (Co-Create London) e settore creativo/culturale. Un esempio di crowdsourcing volontario potrebbe essere quello di Wikipedia la quale affida agli utenti sparsi per la rete il compito, anche se non retribuito, di modificare e aggiungere contenuti mancanti all'interno dell'enciclopedia stessa.

Howe individua quattro tipologie di crowdsourcing:

- Crowd-wisdom, ovvero l'intelligenza collettiva: consiste nel mettere a frutto la conoscenza dei gruppi, in quanto superiore alla conoscenza dei singoli;
- Crowd-creation: utilizza non solo la conoscenza ma anche l'energia creativa di persone comuni per lo svolgimento di attività;
- Crowd-voting: adopera le scelte e i giudizi delle persone comuni per organizzare le informazioni (l'esempio più noto è Google);
- Crowd-funding: permette ai gruppi di raccogliere auto-finanziamenti.

Inoltre sottolinea come spesso i progetti più fortunati derivino dalla combinazione di questi quattro approcci.
Anche Henry Jenkins, direttore del " Comparative Media Studies Program " presso il "Massachusetts Institute of Technology" riguardo il crowdsourcing individua quattro diverse modalità di cultura partecipativa:

- Affiliation: creazione di comunità, formali ed informali, accentrate intorno a diverse forme di media (per esempio Facebook);
- Expression: produzione collettiva di contenuti e nuove forme creative;
- Collaborative problem-solving: lavoro di gruppo allo scopo di portare a termine obiettivi e sviluppare la conoscenza (per esempio Wikipedia);
- Circulation: dare valore al flusso dei media, come nel caso dei blog o dei podcasting.

In realtà nonostante le categorie sopra elencate, non sono ancora ben chiari quei confini che identificano ciò che è definibile come crowdsourcing.
Alcuni ritengono che questo neologismo si possa configurare semplicemente con le azioni volte a completare progetti altrui ma la definizione non è ben chiara a molti.

1.6 Crowdsourcing in linguistica computazionale

Le tecnologie di crowdsourcing sono state impiegate principalmente negli studi che riguardano l'elaborazione del linguaggio.
Nello specifico la prima ad utilizzare questo nuovo metodo fu nel 2007 un'impresa, Powerset, con sede a San Francisco in California. Questi utilizzarono la piattaforma di Amazon Mechanical Turk con l'intento di creare "training data for semantic indexing and relevancy judgments for its natural language search system ".
Per più di un anno la società Americana è stata l'unico grande "Requester" all'interno di AMT. Da allora il progresso e le innovazioni che hanno caratterizzato le tecnologie di crowdsourcing hanno riguardato anche campi di ricerca differenti come la linguistica computazionale.

La linguistica computazionale ha fatto uso di queste tecniche collaborative per portare avanti gli studi riguardanti il "Natural Language Processing", applicando le nuove metodologie anche per quanto riguarda lavori di annotazione, traduzione e trascrizione ecc (Munro, Tily, 2011).

Capitolo II
Creazione di un benchmark task-oriented per la valutazione di DSM

2.1 Introduzione all'esperimento

Nel creare i task relativi alle relazioni semantiche ed avere come campione un elevato numero di utenti è stata utilizzata una piattaforma, specializzata nel creare applicazioni che richiedono l'intervento umano:
Amazon Mechanical Turk – AMT.
Il nome si ispira al "turco meccanico", automa creato da Wolfgang von Kempelen nel 1769 che teoricamente avrebbe dovuto simulare il gioco degli scacchi. In realtà si trattava di un imbroglio in quanto la macchina veniva manovrata dall'interno da un giocatore umano.
AMT è infatti un "*Crowdsourcing Internet Marketplace*", un sito web collaborativo dotato di una interfaccia programmabile, che permette, a sviluppatori di ogni genere, di incorporare l'intelligenza umana nelle loro applicazioni.
Come già sappiamo il termine *"Crowdsourcing"* identifica una nuova metodologia di collaborazione attraverso la quale, imprese e non solo, richiedono un contributo attivo alla rete, delegando ad un insieme distribuito di persone lo sviluppo di un progetto.
Le motivazioni risiedono nel fatto che in questo modo è possibile risolvere quesiti che non possono essere affidati a un computer ma che necessitano dell'uomo.
I punti cardine attorno ai quali ruota il sistema di AMT sono quattro:

- *Human Intelligence Tasks (HITs)*, un HIT è un singolo e autonomo "Task", ad esempio data la parola "planet" individuare il relativo sinonimo.
- *Requesters*, coloro che creano progetti e a partire da questi "*Batch of HITs*" .
- *Workers*, persone che completano le HITs relative ai *Batch* disponibili.
- *Payment*, i workers che completano nel modo corretto i loro Task, secondo il giudizio dei Requesters, vengono retribuiti.

Quindi come Requester per prima cosa abbiamo creato una serie di progetti e ad ognuno di essi è stata associato un file contenente una lista di parole sulla base delle quali i Workers hanno completato una serie di task. Questi richiedevano di individuare un sinonimo, un antonimo o un iperonimo per ciascuna parola.
I dati da sottoporre agli utenti di AMT sono parole inglesi estratti da WordNet 1.6 in numero e in modo tale da avere come risultato undici file per ogni categoria sintattica - verbi, nomi e aggettivi - ognuno dei quali contenente undici parole esatte di cui 9 parole esistenti e 2 non-parole da utilizzare come indizio nell'individuare eventuali *scammers*(In questo caso, utenti con un profilo falso e che completano ugualmente Task disponibili solo a persone madrelingua inglese, cosa che loro non sono).
Analizzando i numeri, nel complesso sono stati creati undici progetti per ogni relazione semantica di ognuna delle tre parti del discorso. Ogni progetto creato, una volta caricato il file contenente l'elenco di parole corretto, è divenuto un "*Batch of HITs*" con un numero di *HIT* pari al numero di parole del file.
Una volta che i Workers hanno completato la serie di Task presenti in ogni Batch, il compito del Requester è stato quello di approvare o meno le risposte date.
Sulla base dei risultati ottenuti, dopo un processo di normalizzazione, i dati sono stati analizzati calcolando la frequenza con la quale gli utenti hanno risposto scegliendo un determinato termine come iperonimo, sinonimo o antonimo.

2.2 Creazione di un Batch di HITs

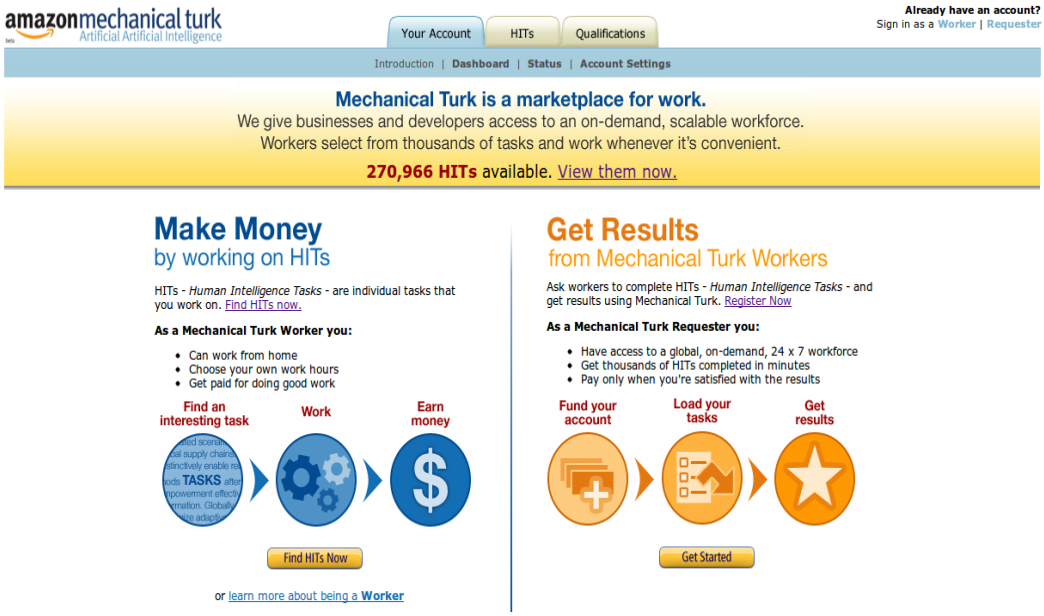


Figura 1. AMT home page

La Home Page di Amazon Mechanical Turk in Figura 1 consente di scegliere quale dei due possibili ruoli, "*Worker*" o "*Requester*", si abbia la necessità di interpretare.
Nel caso l'intenzione sia quella di creare un "*Batch of HITs*", ovvero una serie di *Tasks* differenti, il ruolo da rivestire sarà il secondo, l'intento invece quello di creare applicazioni "*task oriented*" da sottoporre all'utenza di AMT.
Il primo passo consiste nell'effettuare il login attraverso le credenziali fornite come Requester. Una volta ottenuto l'accesso ad Amazon Mechanical Turk, avremo piena interazione con quella che viene chiamata RUI, *Requester User Interface*.
Tramite l'interfaccia sarà possibile:

- *Definire le proprietà e disegnare il layout del nostro progetto.*
- *Pubblicare una volta completato un "Batch of HITs".*
- *Approvare o meno il lavoro sulle HITs lanciate.*

Selezionando il tab *Create*, posizionato all'interno della barra di navigazione, verremo indirizzati nella pagina che offre la possibilità di creare un nuovo progetto. Sotto il riquadro orizzontale, intitolato "Start a new project" , comparirà una lista di "*sample template*" da utilizzare come modello di base a seconda del tipo di progetto che si vuole creare (Figura 2). Ovviamente la selezione non è vincolante, in un secondo momento il template può essere modificato sulla base di una struttura differente progettata a seconda delle esigenze.

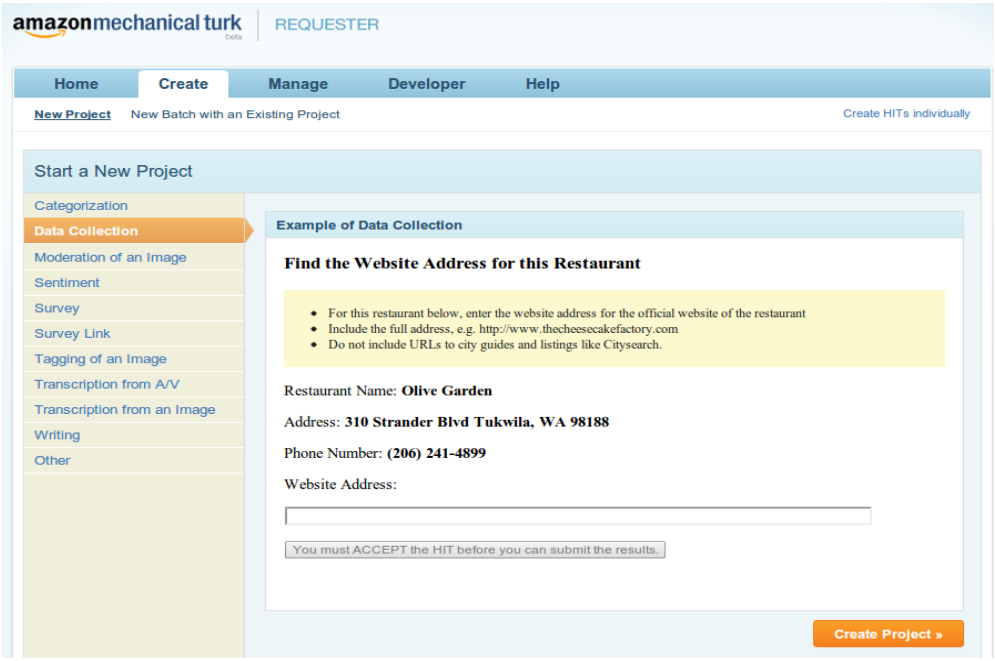


Figura 2. Scelta dei template

Nel selezionare il bottone in basso a destra "Create Project", si avrà accesso alla pagina adibita all'editing del progetto.
Questa fase consiste in tre passaggi fondamentali:

- *Define the projects properties / Definire le proprietà del progetto.*
- *Design the project's HTML layout / Disegnare il layout del progetto attraverso la possibilità di modificare il codice html.*
- *Preview the project/ Visualizzare un'anteprima del progetto.*

Il primo step "*Enter Properties*" in Figura 3 consente di inserire alcuni dati di base e si divide in due sezioni principali: "Describe your HIT to workers", "Setting up your HIT".

Innanzi tutto si dovrà scegliere il nome del progetto. Successivamente nella prima sezione vengono richieste determinate informazioni che saranno utili nel descrivere la HIT come il titolo, una breve descrizione generale e le relative keywords.

La seconda sezione prevede la modifica di alcuni valori di base associati alla HIT stessa ovvero l'ammontare della ricompensa dopo che i Workers hanno completato correttamente i Tasks da loro svolti, la data di scadenza e il tempo entro il quale le risposte degli utenti devono essere giudicate idonee o meno.

Infatti, nel caso il Requester non dovesse valutare l'operato dei Workers entro il tempo prestabilito, le risposte verranno approvate automaticamente da Amazon.

Sempre all'interno delle proprietà, sono presenti una serie di opzioni aggiuntive che hanno l'obbiettivo di restringere il numero di Workers partecipanti.

Nello specifico in questo modo abbiamo la possibilità di scegliere, in base alle necessità, quali di loro possano svolgere le HIT o meno. Ad esempio nel nostro caso era importante che solo gli utenti, la cui residenza fosse gli U.S.A, avessero la possibilità di lavorare al nostro progetto.

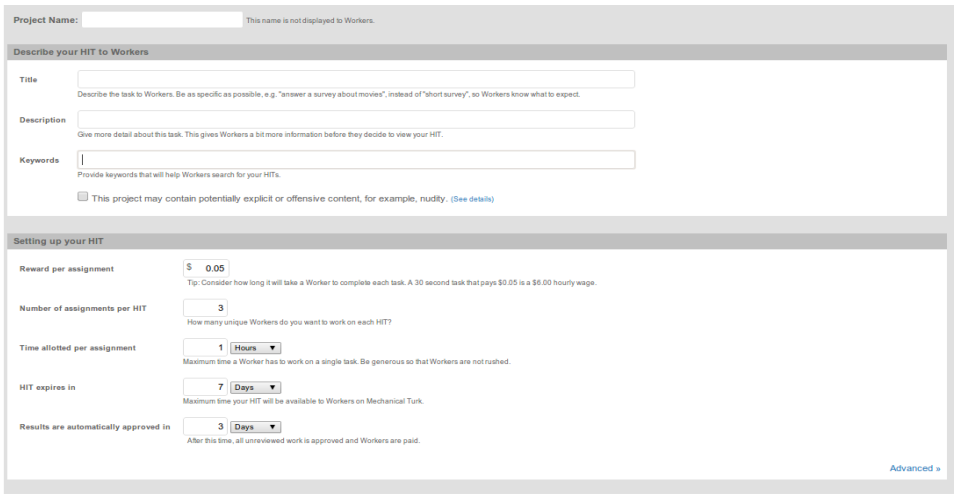


Figura 3. Enter Properties

Il secondo step "*Design layout*", Figura 4, permette la modifica del template di base, che avevamo selezionato in partenza, tramite un editor che consente, tra le altre cose, di visualizzare il codice sorgente del progetto stesso selezionando l'apposito tab presente sulla destra. Questo consente una modifica nel dettaglio in quanto rende possibile operare direttamente all'interno del codice html e css

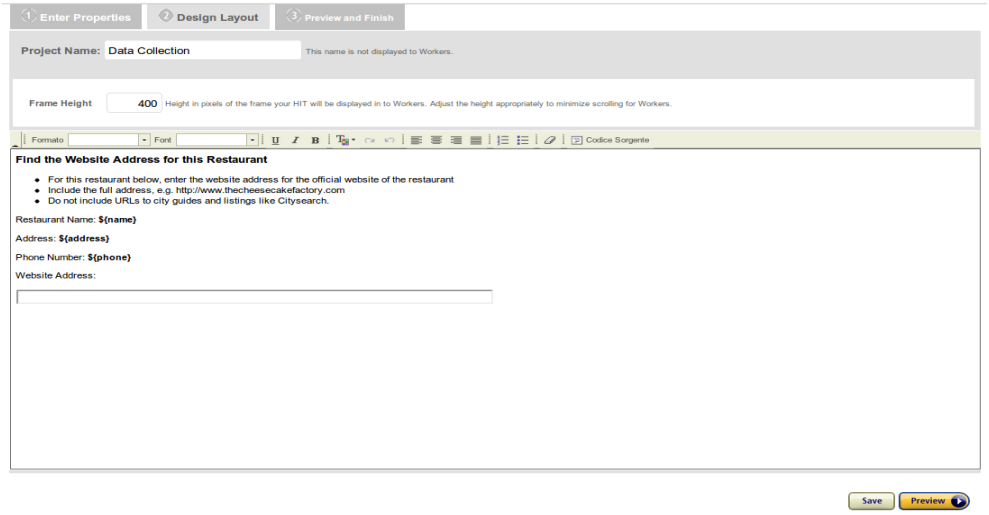


Figura 4. Design layout

L'ultimo passo "*Preview and finish*" mostra, in Figura 5, per prima cosa le informazioni relative alla nostra HIT impostate inizialmente, quindi il titolo, la ricompensa, la durata e le eventuali restrizioni imposte.

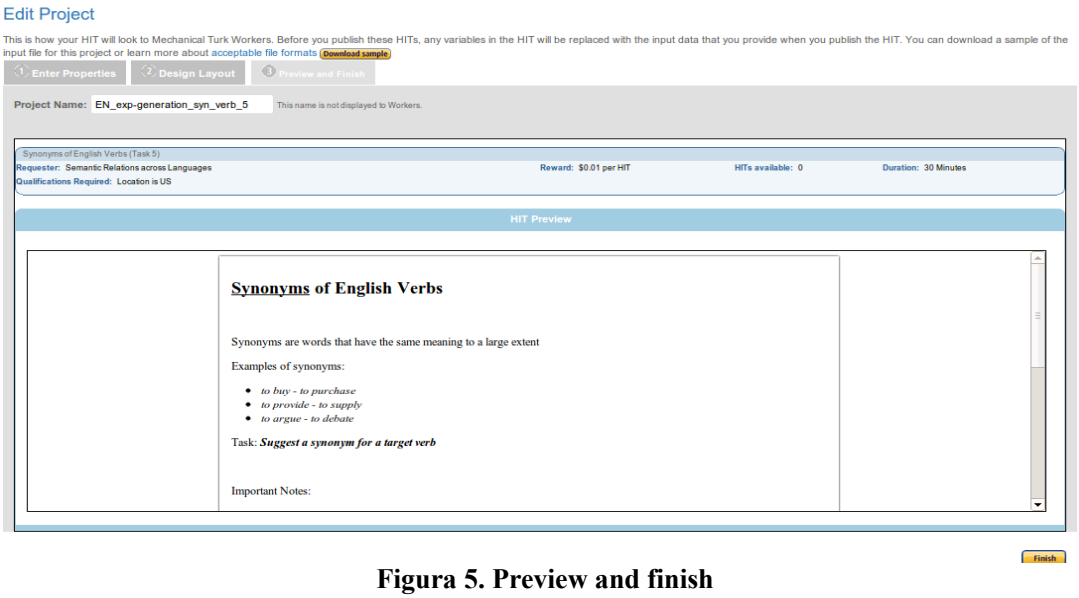


Figura 5. Preview and finish

Mentre immediatamente dopo viene mostrato il risultato finale dal punto di vista grafico, ovvero la visuale che avrà l'utente nel momento in cui dovrà svolgere il Task selezionato.

Una volta completato il progetto questo verrà salvato a parte senza essere reso immediatamente disponibile ai Workers presenti nella rete di AMT.

Selezionando nuovamente il tab Create apparirà una nuova schermata nella quale sarà presente un elenco, intitolato "*Start a New Batch with an Existing Project*", contenente tutti i progetti creati (Figura 6). Per ognuno di essi sarà possibile effettuare alcune operazioni come la modifica del template tramite il tab "edit" o semplicemente la cancellazione del progetto attraverso "*Delete*".

Il tab "*New Batch*" se selezionato creerà effettivamente il "Batch of HITs" che avevamo intenzione di realizzare.

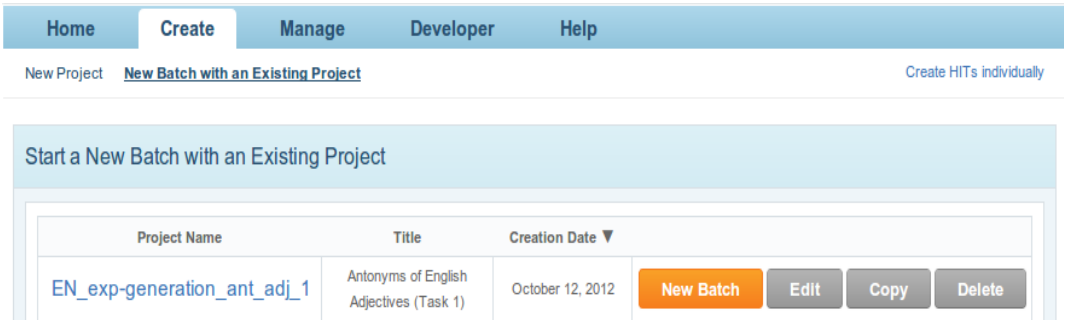


Figura 6. Elenco progetti creati

Nel nostro caso e in generale per tutti quei template nei quali si fa uso di variabili, selezionando lo stesso tab, si aprirà una finestra di dialogo nella quale viene indicato di scegliere un file .csv contenente le variabili che si vogliono utilizzare in quel determinato progetto, come mostrato in Figura 7.

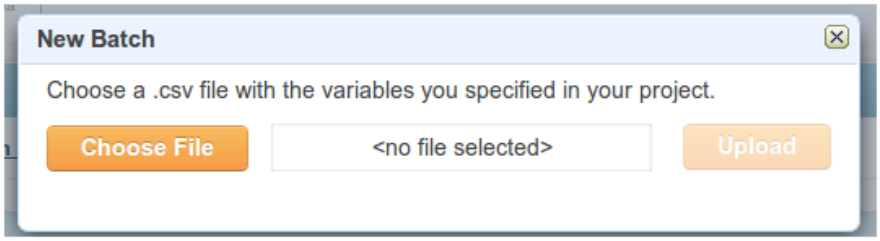


Figura 7. Finestra di dialogo

Una volta scelto il file, per completare la procedura, sarà possibile selezionare il bottone *Upload*, inizialmente caratterizzato da un effetto di trasparenza e non disponibile.
Il numero di HITs per progetto sarà pari al numero di parole presenti nel file .csv.
Per inserire una variabile all'interno di un template si utilizza la sintassi *${nomeVariabile}*. Affinché, nel nostro caso *${nomeVariabile}*, venga correttamente sostituita dalle parole presenti nei file .csv, lo stesso file deve avere, come parola iniziale, il nome della variabile utilizzata, in questo caso "nomeVariabile", seguita dalle altre parole precedute dalla virgola.

In questo modo avremo creato un "Batches of HITs" e selezionando il tab "Manage" presente sulla barra di navigazione orizzontale accanto a "Create", è possibile visualizzare, subito dopo la riga "Batch in progress", lo status di completamento di tutti i "Batch" lanciati.
Una volta che gli utenti completeranno tutte le HITs, Amazon consentirà di visualizzare i risultati ottenuti i quali dovranno essere revisionati dallo stesso Requester. Una volta verificate le risposte date, i Batch verranno spostati nella sezione "*Batch already reviewed*" (Figura 8).

Figura 8. Batch already reviewed

Per ogni Batch approvato sarà possibile scaricare i file .csv contenenti tutti i dati ottenuti attraverso l'esperimento. Selezionando il bottone *Download CSV* verrà automaticamente scaricato sul pc un documento contenete tutte le informazioni relative al tipo di *HIT* completata, ai quesiti posti dai Requester, le relative risposte dei Workers e altre informazioni utili a seconda dei casi.

2.3 Estrazione dei dati da WordNet

WordNet nasce nel 1985 ad opera del linguista George Armitage Miller, professore presso l'Università di Princeton. Si tratta di un database lessicale online per la lingua inglese il quale si propone di organizzare i vocaboli secondo principi differenti da quelli lessicografici adottati fino ad allora.
Infatti per ogni categoria sintattica (nomi, verbi, aggettivi e avverbi), vengono elencati l'insieme di tutti i significati possibili di quel termine ognuno dei quali arricchito da una lista di sinonimi.

WordNet divide il lessico in 5 categorie sintattiche: sostantivi, verbi, aggettivi, avverbi e "*function words*" (Al momento, solo le prime quattro categorie sono contemplate). Gli elementi di ogni categoria sono organizzati all'interno di gruppi di sinonimi, chiamati "*synsets*", ciascuno dei quali rappresenta un concetto lessicale che sottintende. Ogni gruppo è collegato agli altri sulla base delle relazioni semantiche le quali, all'interno di *WordNet,* variano a seconda della categoria sintattica.
Ogni *synset* possiede una breve definizione e, in moti casi, una o più frasi che svolgono una funzione esplicativa dei risultati ottenuti (Miller, 1995).

Il procedimento di preparazione dei dati per gli esperimenti ha previsto l'estrazione di nomi, verbi e aggettivi, presenti su *WordNet*, versione 1.6, sulla base delle macrocategorie semantiche secondo le quali sono organizzate le tre categorie lessicali. Nello specifico è stato individuato un campione di 297 parole da impiegare come variabili all'interno delle HITs.

In un secondo momento sono state aggiunte al campione 66 non-parole per un totale di 363 elementi divisi in 11 file per ognuna delle tre categorie sintattiche.
Ciascun file doveva essere composto da 11 elementi di cui 9 parole e 2 non-parole.
Le non-parole sono state inserite con l'intenzione di verificare che i Workers avessero le credenziali per poter svolgere i Task.
Infatti durante la loro creazione è stata impostata una condizione per la quale solo gli utenti provenienti dagli U.S.A potessero completare le HIT.
Quindi nel caso in cui fosse stato inserito un sinonimo, un iperonimo o un antonimo di una non-parola, questo poteva essere un indizio della presenza di uno *scammer,* con la conseguenza che il Requester avrebbe dovuto non approvare l'intera serie di risposte di quell'utente.
Le macrocategorie semantiche di *WordNet* sono:

NOMI	*VERBI*	*AGGETTIVI*
Tops	Body	All
Animal	Change	Pert
Artifact	Cognition	
Attribute	Communication	
Body	Competition	
Cognition	Contact	
Communication	Creation	
Feeling	Emotion	
Food	Motion	
Group	Percepition	
Location	Possession	
Motive	Social	
Object	Stative	
Person	Weather	
Phenomenon		
Plant		
Possession		
Quantity		
Relation		
Shape		
Substance		
Time		

Nello scegliere le parole utili alla finalizzazione dell'esperimento su AMT sono stati utilizzati due parametri fondamentali: la frequenza e la polisemia.

Per quanto riguarda la frequenza sono state individuate tre fasce di valore :

- *Valore maggiore di 200 e minore o uguale a 3.000 , fascia di frequenza minima.*
- *Valore maggiore o uguale 3.000 e minore o uguale a 10.000, fascia di frequenza media.*
- *Valore maggiore di 10.000, fascia di frequenza massima.*

Per ottenere il valore esatto di ogni parola presente su *WordNet* e impostare la relativa fascia di frequenza, è stato usato *UkWaC*.
Si tratta di un corpus per l'inglese formato da circa 2 miliardi di parole costruito tramite la procedura di "*web crawling* " del *.uk Internet domain*.
Facendo un confronto tra le parole presenti in *WordNet* con quelle di *UkWaC* è stato possibile associare ad ogni parola la relativa fascia di appartenenza.
Il medesimo procedimento è stato impiegato per individuare la polisemia delle parole. Lo stesso *WordNet* infatti associa, ad ognuna di queste, un valore che identifica il numero di sensi che quella parola possiede a seconda della parte del discorso alla quale appartiene. Così ad esempio il termine "*spring*" può essere sia un nome, sia un verbo. Nel primo caso sono presenti sei significati differenti associati alla parola, ad esempio "*spring*" come Primavera o come molla. Nel secondo caso il verbo "*to spring*" ha cinque significati differenti come essere scaraventati lontano da un impatto ("*spring away from an impact*") o scoprire qualcosa improvvisamente, inaspettatamente ("*He sprang these news on me just as I was leaving*").
Quindi, sulla base del numero di significati che un termine possiede, sono stati individuati tre valori di polisemia fondamentali all'esperimento:
Polisemia 1, Polisemia 2 e Polisemia 3.
Le parole con frequenza minore di 200 e polisemia maggiore di 3 sono state scartate, le altre sono state suddivise, all'interno delle loro macrocategorie, in base a questi due valori. Utilizzando un esempio, all'interno della macrocategoria semantica *animal,* avremo nomi con frequenza minima e polisemia 1, frequenza minima e polisemia 2, frequenza minima e polisemia 3 e così via per gli altri casi come frequenza media e massima.

Sulla base della suddivisione effettuata sono stati scelti a caso i termini finali da utilizzare come stimoli all'interno di Amazon Mechanical Turk. Nello specifico sono state individuate 99 parole per ogni parte del discorso, 11 per ognuna delle 9 combinazioni polisemia, fascia di frequenza. All'interno di ciascuna delle 9 combinazioni le 11 parole sono state estratte in modo bilanciato dalle varie macrocategorie semantiche.

2.4 Creazione file .csv

Lo step successivo è stato quello di creare gli undici file .csv per ogni coppia relazione semantica - parte del discorso, ognuno dei quali composto da 11 elementi ciascuno. Questo ha previsto inizialmente l'impiego di uno script, creato tramite il linguaggio di programmazione Perl, con il compito di suddividere le parole estratte da *WordNet*. Il programma aveva il compito di verificare, tramite un'istruzione condizionale, i valori di frequenza e polisemia delle parole, creando output differenti sulla base di questi valori. Come risultato, ad esempio, i 99 aggettivi estratti da *WordNet* sono stati suddivisi all'interno di 9 file differenti per ognuna delle combinazioni tra frequenza e polisemia. La stessa cosa è avvenuta per i nomi e i verbi.

Successivamente è stato utilizzato un ulteriore script in Perl con il compito di generare i file .csv. Il procedimento base adottato dal programma è stato quello di scegliere, in maniera del tutto casuale, una singola parola da ognuno dei 9 file precedentemente divisi in base a frequenza e polisemia, posizionando il termine estratto in un nuovo file di output. In questo file inoltre sono state inserite due non-parole selezionate sempre casualmente dal programma da un elenco e, come primo elemento della lista, il nome della variabile utilizzata nei template di AMT.

Lo script avrà anche il compito di eliminare le parole scelte casualmente dal file di origine per evitare di inserirle nuovamente negli output successivi.

Il file sarà quindi formato come previsto dal nome della variabile, nel nostro caso *word*, seguito da 9 parole e 2 non-parole in ordine sparso.

Lo stesso procedimento è stato ripetuto fino ad ottenere 11 file di 11 elementi per ciascuna categoria lessicale.

2.5 Realizzazione dell'esperimento

Per prima cosa il compito del Requester è stato quello di creare undici progetti per ogni relazione semantica di ognuna delle tre parti del discorso da utilizzare: nomi, aggettivi e verbi. Per ogni progetto creato doveva essere associato un file .csv contenente ciascuno un elenco di undici parole appartenenti a una delle tre parti del discorso. Dopo aver avuto accesso alla pagina adibita alla creazione dei progetti, tramite il tab *Create*, al momento della selezione dei template di base, abbiamo utilizzato "*data collection*".

Dopo aver inserito il nome del progetto con un identificativo differente in base alla relazione semantica e alla parte del discorso – es: EN_exp-generation_syn_verb -, sono stati completati i campi relativi alla descrizione degli *HITs*.

Inoltre si è provveduto ad impostare una ricompensa minima per ogni Task approvato, trenta minuti come limite massimo entro il quale i Workers dovevano completare il compito scelto, quattro giorni il periodo nel quale ogni *HIT* sarebbe rimasta disponibile su *AMT*, e sette giorni come limite di tempo dopo il quale ogni *Task* non revisionato dal Requester sarebbe stato automaticamente approvato.

Per finire è stata inserita una restrizione per imporre un limite al tipo di Workers che potesse completare i nostri "*Batch of HITs*".

E' stato quindi impostato che solo persone la cui locazione fosse gli U.S.A avrebbero potuto completare il nostro lavoro.

Il template di base è stato sostituito interamente da un modello precostruito progettato con l'intento di raccogliere nel modo corretto i dati necessari. La sua struttura prevede un titolo sulla base del tipo di progetto. Ad esempio prendendo gli antomimi di aggettivi compare "Antonyms/Opposites of English adjectives", seguito da una piccola spiegazione relativa alla relazione semantica corrispondente e un indicazione evidenziata in grassetto che suggerisce il tipo di *Task* - es: *Suggest an antonym for a target adjective*. Sono stati inoltre messi alcuni esempi di coppie di parole legate dalla relazione oggetto del task, ad esempio sempre nel caso degli antonimi degli aggettivi: *large / small*, *know/ unknow*, *new/ old*.

Oltre a questo sono state poste anche alcune note importanti in tutti i layout, utili agli utenti che devono completare il loro task:

- *Only for native English Speakers.*
- *Please, complete all the 11 HITs.*
- *The test also contains some items that native speakers can identify as "invented words". In such case, please select the "I don't know" field below.*

L'ultimo punto specifica, in presenza di alcune "*non parole*", di selezionare il campo "*I don't know"* sopra il quale compare una variabile, identificata dalla sintassi *${word}*, seguita da un campo rettangolare all'interno del quale i Workers dovranno inserire, a seconda del Task richiesto, un antonimo, un sinonimo o un verbo.
Una volta associato il file .csv al progetto e reso attivo ai Workers, questi vedranno comparire al posto della variabile le parole contenute nel file caricato.
Una volta ultimato il layout, il terzo step "*Preview and Finish*" ci mostra la visuale che i Workers avranno al momento di svolgere il nostro lavoro (Figura 9). Selezionando il tasto "*Finish*" in basso a destra avremo ultimato il progetto.

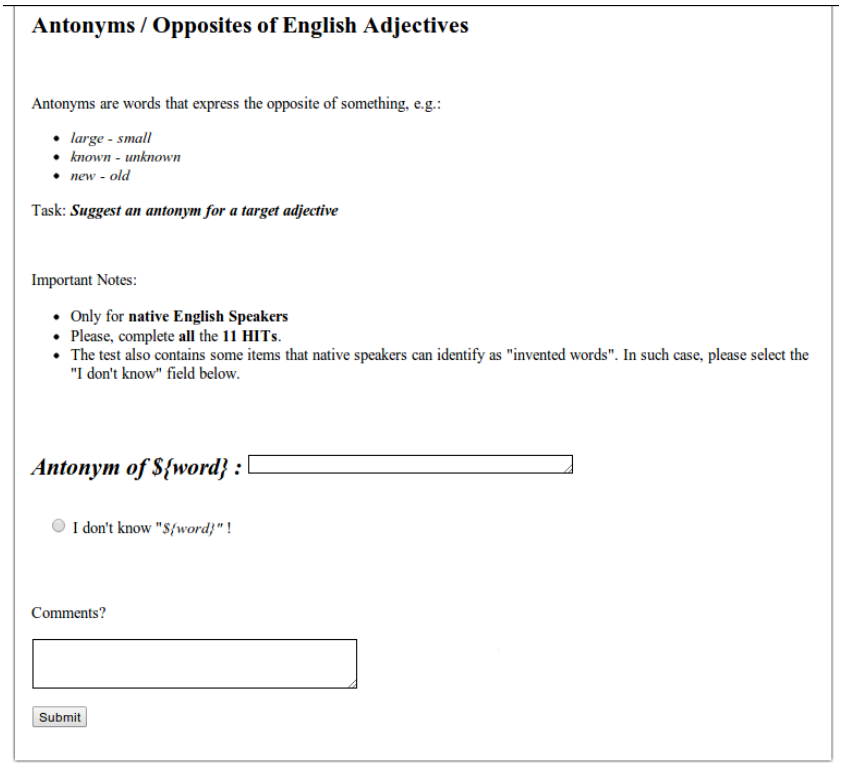


Figura 9. Visuale Workers

Nella pagina dove compare l'elenco di tutti i progetti creati fino a quel momento, abbiamo selezionato "*New Batch*" scegliendo il file .csv corrretto. A questo punto dopo aver creato a tutti gli effetti una serie di benchmarks task-oriented, non rimaneva che attendere il loro completamento.

2.6 Normalizzazione dei dati

Il procedimento consiste nel creare un file di testo raggruppando in un unico elenco tutte le risposte date dai Workers di ognuno degli undici progetti, come ad esempio i sinonimi dei nomi o gli iperonimi degli aggettivi. Quindi per ogni tipologia di progetto è stato creato un documento di testo da normalizzare, ovvero modificare secondo alcuni parametri necessari per una corretta analisi sui dati ottenuti. L'elenco di dati contenuti in ogni file doveva essere strutturato in modo tale da avere due colonne di cui la prima con tutte le parole dei file .csv, mentre la seconda con le risposte relative date dagli utenti.
Dopo la revisione da parte dello stesso "requester" di ogni progetto portato a termine dai Workers, Amazon Mechanical turk permette di scaricare, per ognuno di essi, un file .csv con le risposte date dagli utenti. Selezionando il tab Create, nella sezione di controllo dei Batches sotto *Batches already reviewed*, per ognuno di essi, selezionando *results* e successivamente *download csv* verranno scaricati i file necessari per un totale di 99 file .csv.

Ognuno di questi file .csv si presenta in questo modo:

	P	Q	R	S	T	U
1	WorkerId	AssignmentStatus	AcceptTime	SubmitTime	AutoApprovalTime	ApprovalTime
2	A3RHQUJHPKBB4B	Approved	Tue Nov 20 18:52:13 GMT 2012	Tue Nov 20 18:52:33 GMT 2012	Tue Nov 27 10:52:33 PST 2012	Tue Nov 20 23:49:28 PST 2012
3	ASWZD61ZKZ0RZ	Approved	Tue Nov 20 18:22:47 GMT 2012	Tue Nov 20 18:22:58 GMT 2012	Tue Nov 27 10:22:58 PST 2012	Tue Nov 20 23:49:27 PST 2012
4	A2NSWGSPODDQ5C	Approved	Tue Nov 20 20:30:38 GMT 2012	Tue Nov 20 20:31:17 GMT 2012	Tue Nov 27 12:31:17 PST 2012	Tue Nov 20 23:49:16 PST 2012
5	ADJG2Z2KT9ZI3	Approved	Tue Nov 20 21:56:13 GMT 2012	Tue Nov 20 21:56:22 GMT 2012	Tue Nov 27 13:56:22 PST 2012	Tue Nov 20 23:49:32 PST 2012
6	A2J40JEWL4EY5D	Approved	Tue Nov 20 21:23:50 GMT 2012	Tue Nov 20 21:23:56 GMT 2012	Tue Nov 27 13:23:56 PST 2012	Tue Nov 20 23:49:29 PST 2012
7	A26B18YJBK0RF5	Approved	Tue Nov 20 18:26:29 GMT 2012	Tue Nov 20 18:26:45 GMT 2012	Tue Nov 27 10:26:45 PST 2012	Tue Nov 20 23:49:29 PST 2012
8	A049499332VKQ31EB7MV5	Approved	Tue Nov 20 18:46:45 GMT 2012	Tue Nov 20 18:47:10 GMT 2012	Tue Nov 27 10:47:10 PST 2012	Tue Nov 20 23:49:26 PST 2012
9	A3VGZ0GXXNJIY8	Approved	Tue Nov 20 17:13:12 GMT 2012	Tue Nov 20 17:13:26 GMT 2012	Tue Nov 27 09:13:26 PST 2012	Tue Nov 20 23:49:23 PST 2012
10	A3HMBHM8HJLKRD	Approved	Tue Nov 20 19:51:06 GMT 2012	Tue Nov 20 19:51:24 GMT 2012	Tue Nov 27 11:51:24 PST 2012	Tue Nov 20 23:49:31 PST 2012
11	A1ZXA9517DY7BY	Approved	Tue Nov 20 20:06:18 GMT 2012	Tue Nov 20 20:07:07 GMT 2012	Tue Nov 27 12:07:07 PST 2012	Tue Nov 20 23:49:23 PST 2012
12	A26B18YJBK0RF5	Approved	Tue Nov 20 18:25:00 GMT 2012	Tue Nov 20 18:25:09 GMT 2012	Tue Nov 27 10:25:09 PST 2012	Tue Nov 20 23:49:32 PST 2012
13	A3HMBHM8HJLKRD	Approved	Tue Nov 20 19:50:08 GMT 2012	Tue Nov 20 19:50:29 GMT 2012	Tue Nov 27 11:50:29 PST 2012	Tue Nov 20 23:49:19 PST 2012
14	A1ZXA9517DY7BY	Approved	Tue Nov 20 20:07:10 GMT 2012	Tue Nov 20 20:07:51 GMT 2012	Tue Nov 27 12:07:51 PST 2012	Tue Nov 20 23:49:33 PST 2012
15	A3VGZ0GXXNJIY8	Approved	Tue Nov 20 17:07:45 GMT 2012	Tue Nov 20 17:08:08 GMT 2012	Tue Nov 27 09:08:08 PST 2012	Tue Nov 20 23:49:27 PST 2012
16	A049499332VKQ31EB7MV5	Approved	Tue Nov 20 18:44:45 GMT 2012	Tue Nov 20 18:45:10 GMT 2012	Tue Nov 27 10:45:10 PST 2012	Tue Nov 20 23:49:27 PST 2012
17	A3RHQUJHPKBB4B	Approved	Tue Nov 20 18:54:21 GMT 2012	Tue Nov 20 18:54:50 GMT 2012	Tue Nov 27 10:54:50 PST 2012	Tue Nov 20 23:49:17 PST 2012
18	A2J40JEWL4EY5D	Approved	Tue Nov 20 21:23:14 GMT 2012	Tue Nov 20 21:23:23 GMT 2012	Tue Nov 27 13:23:23 PST 2012	Tue Nov 20 23:49:28 PST 2012
19	ASWZD61ZKZ0RZ	Approved	Tue Nov 20 18:21:26 GMT 2012	Tue Nov 20 18:21:50 GMT 2012	Tue Nov 27 10:21:50 PST 2012	Tue Nov 20 23:49:27 PST 2012
20	AZT6P0RS7ZHCV	Approved	Tue Nov 20 17:37:44 GMT 2012	Tue Nov 20 17:37:55 GMT 2012	Tue Nov 27 09:37:55 PST 2012	Tue Nov 20 23:49:50 PST 2012
21	A3GJPHFUCNBO8J	Approved	Tue Nov 20 19:34:42 GMT 2012	Tue Nov 20 19:34:50 GMT 2012	Tue Nov 27 11:34:50 PST 2012	Tue Nov 20 23:49:24 PST 2012
22	A049499332VKQ31EB7MV5	Approved	Tue Nov 20 18:38:30 GMT 2012	Tue Nov 20 18:39:09 GMT 2012	Tue Nov 27 10:39:09 PST 2012	Tue Nov 20 23:49:21 PST 2012
23	A1ZXA9517DY7BY	Approved	Tue Nov 20 20:02:23 GMT 2012	Tue Nov 20 20:03:13 GMT 2012	Tue Nov 27 12:03:13 PST 2012	Tue Nov 20 23:49:29 PST 2012
24	A2J40JEWL4EY5D	Approved	Tue Nov 20 21:23:59 GMT 2012	Tue Nov 20 21:24:08 GMT 2012	Tue Nov 27 13:24:08 PST 2012	Tue Nov 20 23:49:23 PST 2012
25	A26B18YJBK0RF5	Approved	Tue Nov 20 18:26:48 GMT 2012	Tue Nov 20 18:27:18 GMT 2012	Tue Nov 27 10:27:18 PST 2012	Tue Nov 20 23:49:31 PST 2012
26	ASWZD61ZKZ0RZ	Approved	Tue Nov 20 18:22:13 GMT 2012	Tue Nov 20 18:22:44 GMT 2012	Tue Nov 27 10:22:44 PST 2012	Tue Nov 20 23:49:22 PST 2012
27	A3HMBHM8HJLKRD	Approved	Tue Nov 20 19:32:57 GMT 2012	Tue Nov 20 19:35:42 GMT 2012	Tue Nov 27 11:35:42 PST 2012	Tue Nov 20 23:49:25 PST 2012
28	A3RHQUJHPKBB4B	Approved	Tue Nov 20 18:58:02 GMT 2012	Tue Nov 20 18:58:44 GMT 2012	Tue Nov 27 10:58:44 PST 2012	Tue Nov 20 23:49:17 PST 2012
29	A3GJPHFUCNBO8J	Approved	Tue Nov 20 19:34:21 GMT 2012	Tue Nov 20 19:34:25 GMT 2012	Tue Nov 27 11:34:25 PST 2012	Tue Nov 20 23:49:25 PST 2012
30	A3VGZ0GXXNJIY8	Approved	Tue Nov 20 17:12:33 GMT 2012	Tue Nov 20 17:13:11 GMT 2012	Tue Nov 27 09:13:11 PST 2012	Tue Nov 20 23:49:23 PST 2012
31	A06764702RZDRUTR5RRE	Approved	Tue Nov 20 22:02:42 GMT 2012	Tue Nov 20 22:02:58 GMT 2012	Tue Nov 27 14:02:58 PST 2012	Tue Nov 20 23:49:20 PST 2012
32	ASWZD61ZKZ0RZ	Approved	Tue Nov 20 18:23:01 GMT 2012	Tue Nov 20 18:23:09 GMT 2012	Tue Nov 27 10:23:09 PST 2012	Tue Nov 20 23:49:22 PST 2012
33	A3HMBHM8HJLKRD	Approved	Tue Nov 20 19:39:04 GMT 2012	Tue Nov 20 19:48:08 GMT 2012	Tue Nov 27 11:48:08 PST 2012	Tue Nov 20 23:49:28 PST 2012

Figura 10. Esempio file .csv

	AB	AC	AD	AE
1	Input.word	Answer.Category	Answer.comment	Answer.suggestion
2	unrefined			debonaire
3	unrefined			refined
4	unrefined			Refined
5	unrefined			pure
6	unrefined			Refined
7	unrefined			cultured
8	unrefined			refined
9	unrefined			refined
10	unrefined			regal
11	unrefined			processed
12	dimensional			flat
13	dimensional			flat
14	dimensional	0		
15	dimensional			undimensional
16	dimensional			flat
17	dimensional			unidimensional
18	dimensional			One sided
19	dimensional			flat
20	dimensional			flat
21	dimensional			static
22	English			non-english
23	English			Spanish
24	English			Nonenglish
25	English	0	there is no opposite for English	
26	English			nonenglish
27	English			foreign
28	English		There is no real 'antonym' of English, assuming the capital e means the language, so I just put in any language.	Spanish
29	English	0		
30	English			foreign
31	English			Non-English
32	henevolent	0		

Figura 11. Esempio file .csv

Le informazioni presenti nel file sono molteplici e non tutte fondamentali al nostro obbiettivo. Per prima cosa sono state individuate solo le colonne che a noi interessavano ("*AssignmentStatus*" "*Input.word*" "*Answer.suggestion*") eliminando così quelle superflue.

Una volta completato tale procedimento sono state eliminate tutte le righe che alla casella appartenente la colonna "*AssignmentStatus*" avevano come valore "*Rejected*".
Il compito finale è stato la normalizzazione di ogni parola all'interno di ogni file.
I parametri da rispettare comprendevano la correzione ortografica, le maiuscole trasformate in minuscole, in presenza di parole composte divise da uno spazio vuoto è stato inserito l'underscore così come in presenza del trattino alto (-), in presenza di verbi doveva essere aggiunto il to seguito dall'underscore. Infine nel caso l'utente non abbia dato una risposta, lasciando così lo spazio vuoto, va inserito il numero 0.
Alcuni esempi di casi che necessitavano di normalizzazione sono stati:
to focus - to_focus
differ - differ
medical device - medical_device
movemant - movement
eart - earth

Un esempio di file normalizzato:

	A	B
1	Input.word	Answer.suggestion
2	planet	heavenly_body
3	planet	astronomical_object
4	planet	matter
5	planet	earth
6	planet	earth
7	planet	space
8	planet	earth
9	planet	celestial
10	planet	galaxy
11	planet	celestial_body
12	psychology	social_sciences
13	psychology	science
14	psychology	mind
15	psychology	science
16	psychology	science
17	psychology	science
18	psychology	social_behavior
19	psychology	science
20	psychology	behavioral_science
21	psychology	subject
22	knitwear	scarf
23	knitwear	clothing
24	knitwear	clothing
25	knitwear	shirt
26	knitwear	clothing
27	knitwear	clothing
28	knitwear	clothing
29	knitwear	clothing
30	knitwear	clothing
31	knitwear	clothing
32	chap	clothing
33	chap	person

Figura 12. File Normalizzato

Capitolo III

Analisi statistica dei risultati

L'ultimo passo consiste nell'analizzare statisticamente i risultati ottenuti.

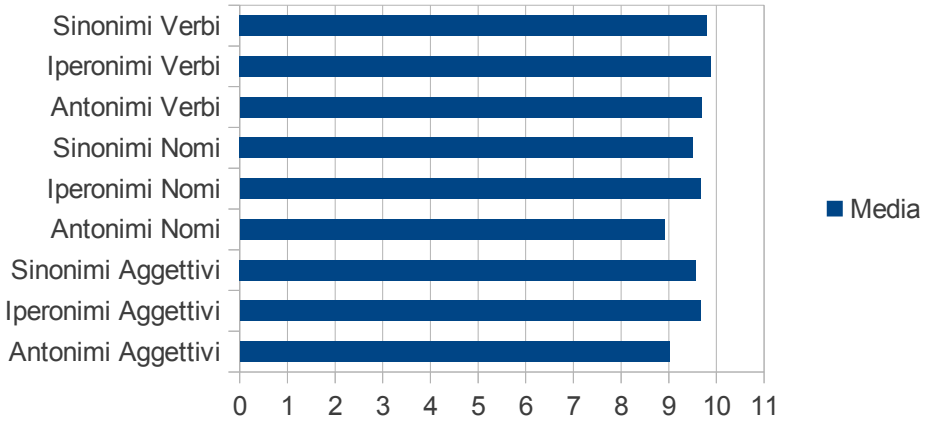
In particolare è stata calcolata la media e la deviazione standard.

La deviazione standard σ, detta anche scarto quadratico medio, si calcola come la radice quadrata della media dei quadrati degli scarti, dove scarto è la differenza tra un valore *Xi* e la sua media:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$$

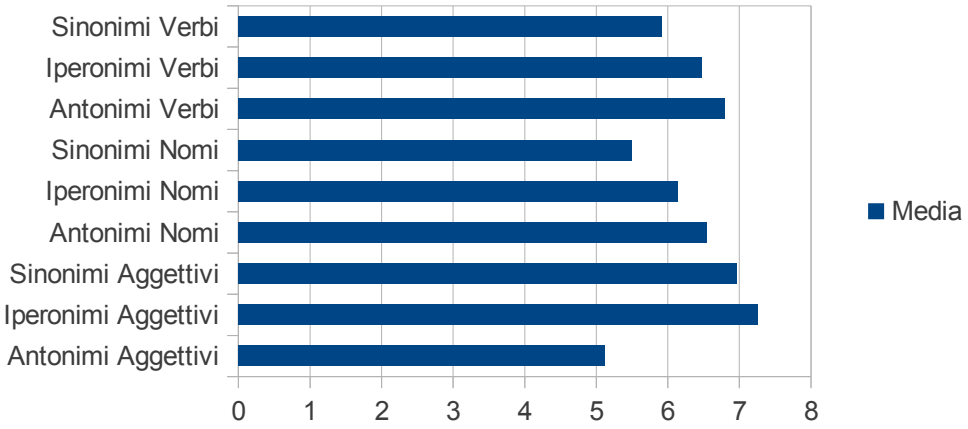
Per ogni relazione e ogni parte del discorso degli stimoli, è stata così calcolata la media e la deviazione standard delle risposte totali prodotte.

Stimoli	Media	Deviazione Standard
Antonimi Aggettivi	9,02	1,74
Iperonimi Aggettivi	9,67	0,91
Sinonimi Aggettivi	9,57	1,06
Antonimi Nomi	8,91	1,79
Iperonimi Nomi	9,66	0,81
Sinonimi Nomi	9,51	1,22
Antonimi Verbi	9,69	0,68
Iperonimi Verbi	9,89	0,48
Sinonimi Verbi	9,81	0,6



Successivamente è stato effettuato lo stesso calcolo non più utilizzando come parametro il numero di risposte totali date ma il numero di risposte diverse per ogni stimolo.

Stimoli	Media	Deviazione Standard
Antonimi Aggettivi	5,12	2,27
Iperonimi Aggettivi	7,26	1,58
Sinonimi Aggettivi	6,97	1,93
Antonimi Nomi	6,54	2,01
Iperonimi Nomi	6,14	2,13
Sinonimi Nomi	5,5	2,03
Antonimi Verbi	6,8	1,92
Iperonimi Verbi	6,47	1,96
Sinonimi Verbi	5,92	1,98

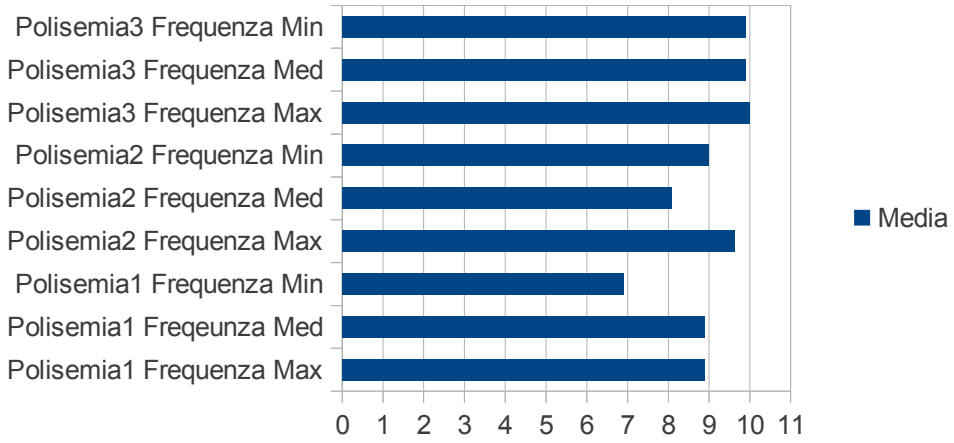


Lo stesso tipo di analisi statistica è stata effettuata sulla base dei valori di polisemia e frequenza che uno stimolo possiede. Così ad esempio abbiamo calcolato media e deviazione standard di ogni stimolo, appartenente agli antonimi degli aggettivi, con polisemia 1 e frequenza massima, successivamente gli stimoli con polisemia 1 e frequenza media e così via per tutte le possibili combinazioni. Il risultato è stato quello di ottenere nove grafici differenti a seconda della relazione semantica e della categoria sintattica.

Per quanto riguarda gli antonimi degli aggettivi, i valori riportati nella tabella relativa mostrano come, in questo caso, gli stimoli, in particolare con un grado di polisemia pari a 3, abbiano ottenuto un numero elevato di risposte da parte degli utenti.
Al contrario gli stimoli con un grado di polisemia 1 e frequenza minima hanno registrato il numero di risposte più basso.
Gli stimoli con polisemia 2 invece hanno ottenuto risultati leggermente migliori rispetto agli elementi con polisemia 1 ad eccezione della frequenza media e polisemia 2 che possiede un valore tra i più bassi.
Dal punto di vista della deviazione standard se da un lato, come era facile prevedere, gli stimoli con un grado di polisemia 3 presentano valori particolarmente bassi, dall'altro lato, gli stimoli con polisemia 1 e 2 presentano una variabilità di risposte maggiore.

Antonimi degli Aggettivi

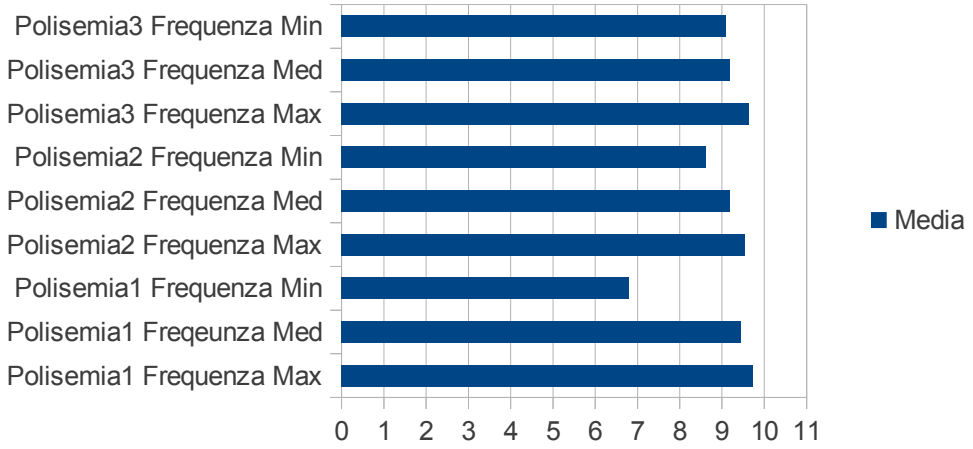
Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	8,9	1,31
Polisemia 1 Frequenza Med.	8,9	0,99
Polisemia 1 Frequenza Min.	6,9	2,46
Polisemia 2 Frequenza Max.	9,63	0,48
Polisemia 2 Frequenza Med.	8,09	2,5
Polisemia 2 Frequenza Min.	9	1,53
Polisemia 3 Frequenza Max.	10	0
Polisemia 3 Frequenza Med.	9,9	0,28
Polisemia 3 Frequenza Min.	9,9	0,28



Come nel caso precedente anche i risultati ottenuti per gli antonimi dei nomi mostrano come gli stimoli che hanno polisemia 3 presentano una media di risposte superiore. E' evidente però un sostanziale equilibrio sia tra le medie degli stimoli di polisemia 2 con frequenza media e massima, sia tra le medie degli stimoli di polisemia 1 con la stessa frequenza. La deviazione standard presenta la stessa situazione. Gli stimoli che possiedono frequenza minima e polisemia 1 così come quelli con frequenza minima e polisemia 2 possiedono una variabilità maggiore relativa al numero di risposte. Da notare come il valore della deviazione standard riferito alla polisemia 3 frequenza minima sia di gran lunga superiore al valore relativo agli antonimi degli aggettivi.

Antonimi dei Nomi

Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	9,72	0,44
Polisemia 1 Frequenza Med.	9,45	0,98
Polisemia 1 Frequenza Min.	6,81	2,85
Polisemia 2 Frequenza Max.	9,54	0,89
Polisemia 2 Frequenza Med.	9,18	1,26
Polisemia 2 Frequenza Min.	8,63	2,26
Polisemia 3 Frequenza Max.	9,63	0,64
Polisemia 3 Frequenza Med.	9,18	1,4
Polisemia 3 Frequenza Min.	9,09	1,62

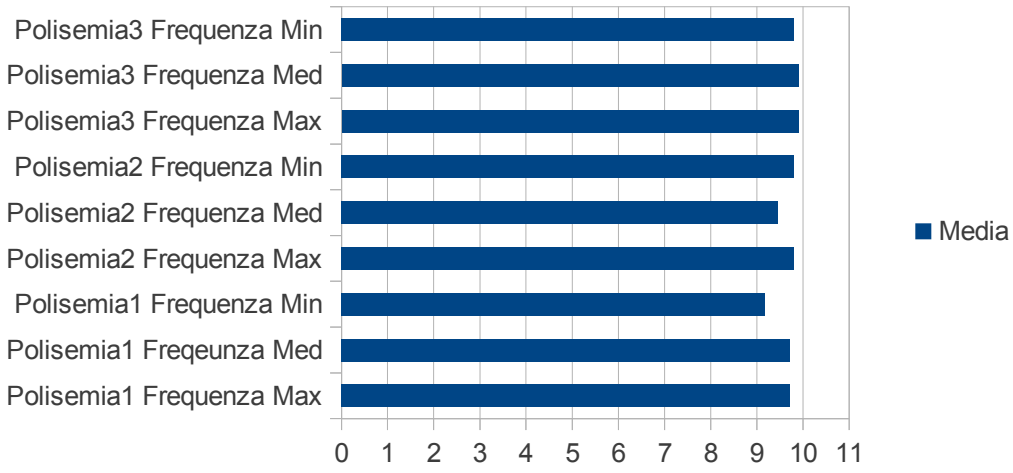


Anche per gli antonimi dei verbi si presenta lo stesso andamento. Le medie relative agli stimoli con polisemia 3 continuano ad essere molto alte così come per la frequenza sia media che massima di polisemia 1 e 2. Gli stimoli con polisemia 1 e frequenza minima invece hanno una media di risposte che risulta essere anche la più bassa.

Non emergono valori particolarmente elevati relativi alla deviazione standard tranne nel caso polisemia 1 e frequenza minima e comunque si tratta di un valore nel complesso abbastanza contenuto.

Antonimi dei Verbi

Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	9,72	0,611
Polisemia 1 Frequenza Med.	9,72	0,74
Polisemia 1 Frequenza Min.	6,18	1,11
Polisemia 2 Frequenza Max.	9,81	0,57
Polisemia 2 Frequenza Med.	9,45	0,89
Polisemia 2 Frequenza Min.	8,81	0,38
Polisemia 3 Frequenza Max.	9,9	0,28
Polisemia 3 Frequenza Med.	9,9	0,28
Polisemia 3 Frequenza Min.	9,81	0,38



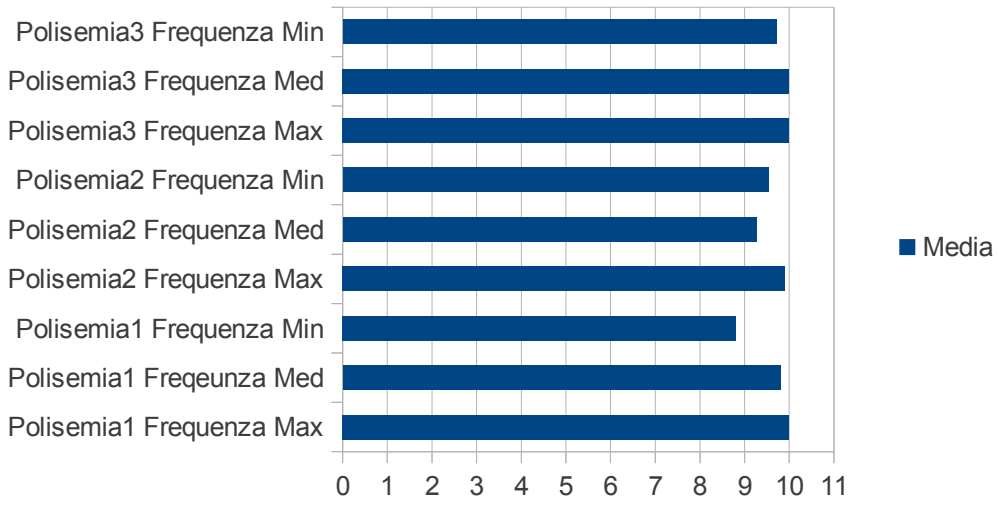
Per quanto riguarda gli iperonimi degli aggettivi , se si analizza la maggior parte delle possibili combinazioni tra polisemia e frequenza, le medie ottenute sono molto elevate. L'unica eccezione risiede nelle risposte date per gli stimoli con polisemia 1 e frequenza minima.

Nella deviazione standard i valori non sono particolarmente elevati , in alcuni casi come quello relativo alla polisemia 3 e frequenza media il calcolo ottenuto è stato pari a 0. In questo caso il numero di risposte per stimolo è stato il medesimo.

Il valore più alto ancora una volta appartiene agli stimoli con polisemia 1 e frequenza minima seguiti immediatamente dopo da quelli con polisemia 2 e frequenza media.

Iperonimi degli Aggettivi

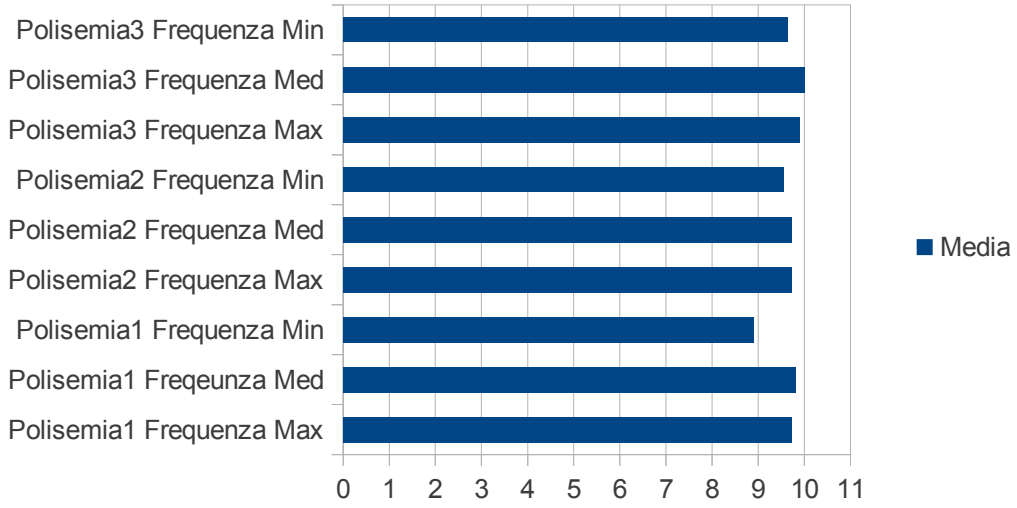
Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	10	0
Polisemia 1 Frequenza Med.	9,8	0,28
Polisemia 1 Frequenza Min.	8,81	1,64
Polisemia 2 Frequenza Max.	9,9	0,28
Polisemia 2 Frequenza Med.	9,27	1,48
Polisemia 2 Frequenza Min.	9,54	0,89
Polisemia 3 Frequenza Max.	10	0,28
Polisemia 3 Frequenza Med.	10	0
Polisemia 3 Frequenza Min.	9,72	0,44



Le medie relative alle risposte effettuate per gli iperonimi dei nomi indicano un elevato numero di risposte non soltanto per le parole con polisemia 3 e polisemia 2 ma anche nel caso di stimoli con polisemia 1 e frequenza sia media che massima.
Come già visto nelle analisi precedenti i termini appartenenti alla categoria polisemia 1 e frequenza minima continuano ad avere una media di risposte più bassa.
Non solo la media ma anche la variabilità nel numero di risposte superiore rispetto agli stimoli appartenenti alla categoria polisemia 2 e 3, costituisce un tratto ricorrente della categoria parola con polisemia 1 e frequenza minima

Iperonimi dei Nomi

Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	9,72	0,61
Polisemia 1 Frequenza Med.	9,81	0,38
Polisemia 1 Frequenza Min.	8,90	1,50
Polisemia 2 Frequenza Max.	9,72	0,61
Polisemia 2 Frequenza Med.	9,72	0,44
Polisemia 2 Frequenza Min.	9,54	0,89
Polisemia 3 Frequenza Max.	9,90	0,51
Polisemia 3 Frequenza Med.	10	0
Polisemia 3 Frequenza Min.	9,63	0,88

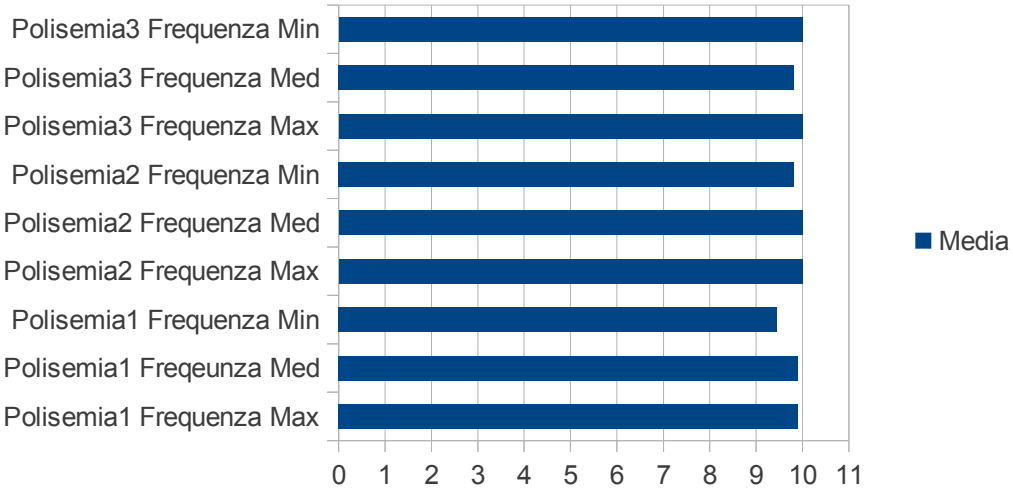


Gli iperonimi dei verbi differiscono in maniera significativa da quanto visto fino ad ora. Il tratto che ha caratterizzato le precedenti analisi, ovvero la discrepanza tra i valori di media e deviazione ottenuti per gli stimoli relativi alla classe polisemia 1 e frequenza minima, non è presente in questo caso.

Tutti i valori ottenuti sono sostanzialmente elevati e lo scarto tra un valore e l'altro è relativamente basso. Nonostante i valori siano sorprendentemente omogenei tra di loro, gli stimoli con polisemia 1 e frequenza minima rimangono ugualmente quelli con la media più bassa e la deviazione standard più alta.

Iperonimi dei Verbi

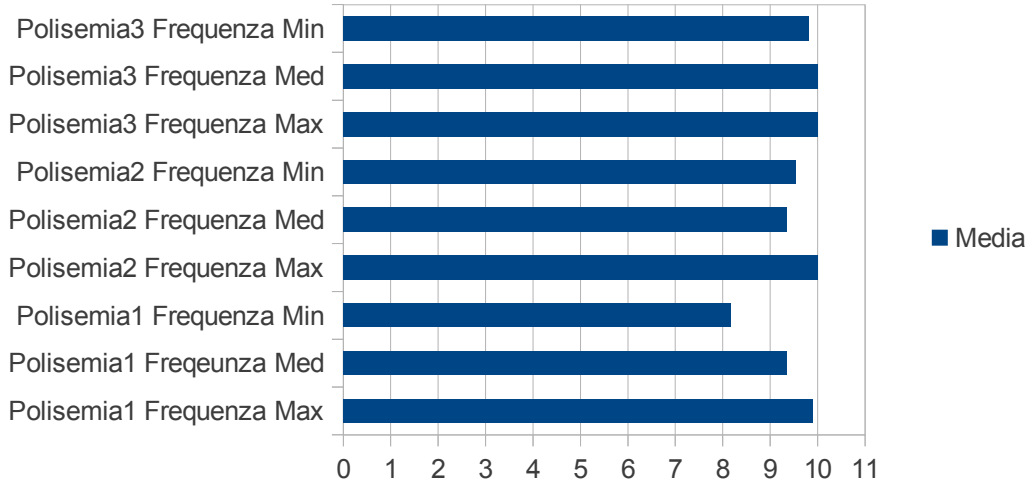
Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	9,90	0,28
Polisemia 1 Frequenza Med.	9,90	0,28
Polisemia 1 Frequenza Min.	9,45	0,89
Polisemia 2 Frequenza Max.	10	0
Polisemia 2 Frequenza Med.	10	0,57
Polisemia 2 Frequenza Min.	9,81	0,38
Polisemia 3 Frequenza Max.	10	0
Polisemia 3 Frequenza Med.	9,81	0,38
Polisemia 3 Frequenza Min.	10	0,42



Analizzando i sinonimi degli aggettivi possiamo notare che in tre casi la media è pari a 10 e la deviazione standard invece equivale a 0. Questo significa che gli stimoli con polisemia 2 e frequenza massima, polisemia 3 e frequenza media, polisemia 3 e frequenza massima hanno ottenuto da parte degli utenti sempre il numero massimo di risposte possibili confermando un repertorio semantico migliore per quanto riguarda i sinonimi delle parole da parte degli utenti stessi.
La deviazione standard continua ad essere discretamente alta nel caso della polisemia 1 e frequenza minima ma anche per la polisemia 2 frequenza media.

Sinonimi degli Aggettivi

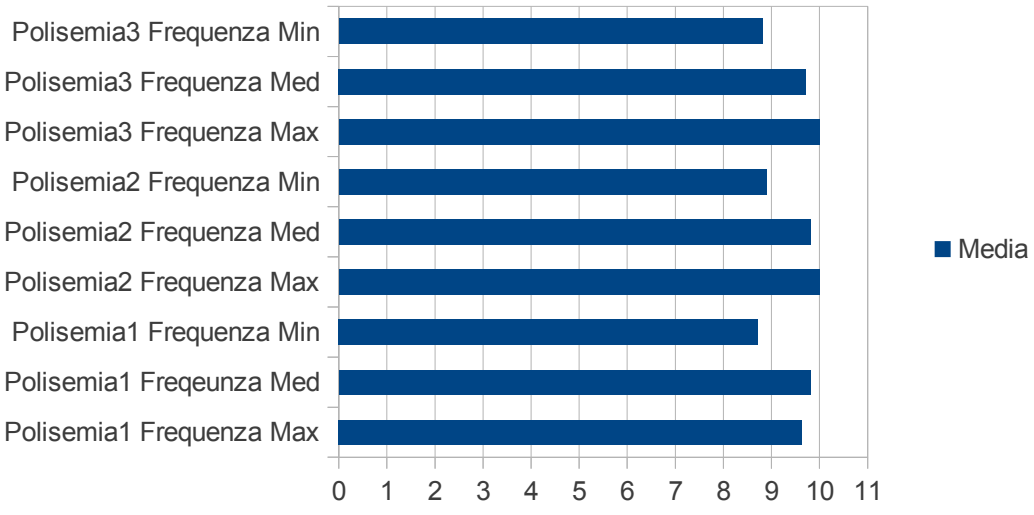
Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	9,9	0,28
Polisemia 1 Frequenza Med.	9,36	0,64
Polisemia 1 Frequenza Min.	8,18	1,84
Polisemia 2 Frequenza Max.	10	0
Polisemia 2 Frequenza Med.	9,36	1,43
Polisemia 2 Frequenza Min.	9,54	1,15
Polisemia 3 Frequenza Max.	10	0
Polisemia 3 Frequenza Med.	10	0
Polisemia 3 Frequenza Min.	9,81	0,38



I risultati ottenuti per i sinonimi dei nomi sono abbastanza omogenei se si analizzano le stesse classi di frequenza ma diversi valori di polisemia. Infatti ad esempio le medie relative alle risposte ottenute per gli stimoli con frequenza minima di ognuna delle tre fasce di polisemia possiedono valori molti simili. La stessa cosa è evidente anche analizzando gli altri casi relativi alla frequenza media e massima delle diverse polisemie. Così come per la media anche la deviazione standard assume lo stesso atteggiamento e varia maggiormente per la frequenza minima mentre la frequenza media e quella massima hanno valori molto vicini tra di loro.

Sinonimi dei Nomi

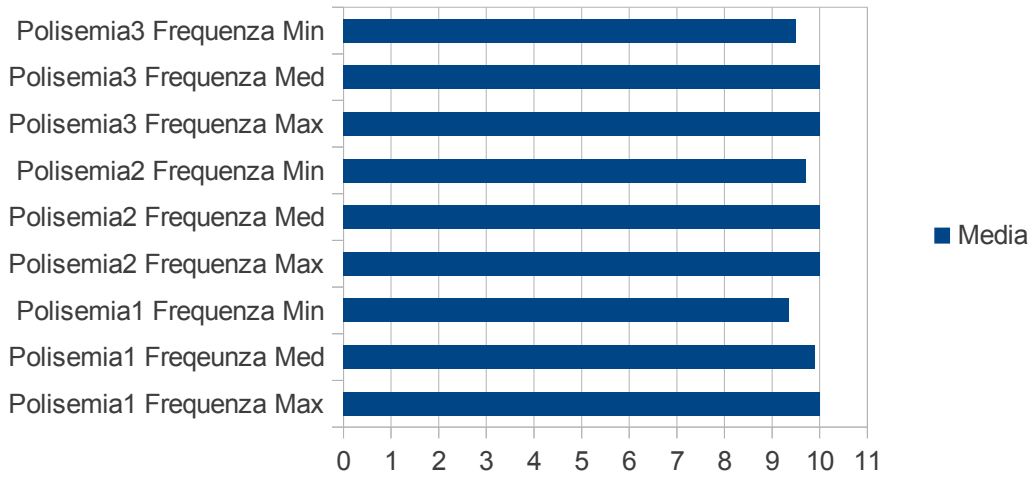
Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	9,63	0,48
Polisemia 1 Frequenza Med.	9,81	0,38
Polisemia 1 Frequenza Min.	8,72	1,48
Polisemia 2 Frequenza Max.	10	0,57
Polisemia 2 Frequenza Med.	9,81	0,57
Polisemia 2 Frequenza Min.	8,90	2,06
Polisemia 3 Frequenza Max.	10	0,60
Polisemia 3 Frequenza Med.	9,72	0,44
Polisemia 3 Frequenza Min.	8,81	1,74



I sinonimi dei verbi presentano medie molto alte per tutte le possibili combinazioni tra frequenza e polisemia. Anche i valori che riguardano la deviazione standard in generale sono omogenei, i più elevati non che i più simili sono i valori ottenuti per la frequenza minima delle tre polisemie.

Sinonimi dei Verbi

Polisemia e Frequenza	Media	Deviazione Standard
Polisemia 1 Frequenza Max.	10	0,6
Polisemia 1 Frequenza Med.	9,9	0,28
Polisemia 1 Frequenza Min.	9,36	0,77
Polisemia 2 Frequenza Max.	10	0,28
Polisemia 2 Frequenza Med.	10	0,42
Polisemia 2 Frequenza Min.	9,72	0,74
Polisemia 3 Frequenza Max.	10	0
Polisemia 3 Frequenza Med.	10	0,42
Polisemia 3 Frequenza Min.	9,5	0,67



In generale tutti i casi analizzati presentano una costante comune:
la classe che produce meno relazioni semantiche è quella con la frequenza minore. Questo vuol dire che per le parole più rare è mediamente più difficile generare relazioni semantiche.

Capitolo IV
Conclusioni

In questo lavoro si è cercato di approfondire la natura e le proprietà delle relazioni semantiche e di muovere qualche passo in avanti nello studio delle metodologie da adottare per creare "*benchmark task oriented*" per la valutazione di Modelli Semantici Distribuzionali.
Un'attenzione particolare è stata dedicata ad indagare, attraverso l'uso di strumenti di crowdsourcing, la capacità degli utenti di generare parole semanticamente correlate ad parole stimolo.

In particolare, nell'ambito di questo lavoro si è cercato di analizzare e comprendere il funzionamento dei più comuni e recenti sistemi di valutazione di un DSM nel settore del Natural Language Processing.
Abbiamo così menzionato il "*TOEFL synonym detection task*", il "*WordSim 353 data set*" e il "*dat set di Almuhareb-Poesio*".
Da questa ricerca è emerso chiaramente che benchmark standard come il "*TOEFL task*" , limitato per altro ad un solo tipo di relazione semantica, non sono stati creati appositamente con l'intento di valutare Modelli Semantici Distribuzionali e per questo motivo devono essere modificati opportunamente.

L'obiettivo del nostro progetto è stato quindi quello di creare un "*benchmark task oriented*" per la valutazione di *DSM*, raccogliendo relazioni semantiche lessicali da soggetti attraverso il metodo di *crowdsourcing*. La piattaforma che sfrutta questa metodologia, impiegata nel nostro esperimento linguistico, è stata Amazon Mechanical Turk .
Durante l'esperimento i "*Workers*" di AMT sono stati sottoposti a *Task*, nel totale 99, che richiedevano di generare sinonimi, antonimi e iperonimi sulla base di una serie di parole stimolo.

Un’attenzione particolare è stata rivolta, in questo contesto, al lavoro che ha previsto la scelta e l'estrazione degli stimoli da impiegare nell'esperimento.
Questi sono stati estratti da WordNet 1.6, database lessicale online per la lingua inglese, sulla base delle macrocategorie semantiche, della frequenza e della polisemia.

I risultati quantitativi ottenuti, illustrati nella parte finale del lavoro con un calcolo della media e della deviazione standard sulla base delle risposte ottenute, confermano nel nostro caso l'importanza di porre un'attenzione sempre maggiore, soprattutto attraverso l’utilizzo di strumenti di crowdsourcing, alle abilità nell'individuare diverse relazioni semantiche da parte degli utenti presenti sulla rete.
In particolare, l’esperimento mostra la capacità da parte di persone comuni di generare parole legate allo stimolo da una specifica relazione semantica.

Bibliografia

- M. Lynne Murphy. 2003. *Semantic Relations and the Lexicon.*
- Alessandro Lenci. 2009. *Spazi di parole. Metafore e Rappresentazioni semantiche.*
- Alessandro Lenci, Marco Baroni. 2011. *How we BLESSed distributional semantic evaluation.*
- *Abdulrahman Almuhareb. 2006. Attributes in Lexical Acquisition.* Phd thesis, University of Essex.
- Miller G. A. and Charles W. G. (1991). *Contextual correlates of semantic similarity. Language and Cognitive Processes.*
- George A. Miller. 1995.*WordNet: A Lexical Database for English.*
- Landauer Th. K. and Dumais S. T. (1997). *A Solution to Plato's problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge.*
- Robert Munro, Hal tily. 2011. *The Start of the art: An introduction ti Crowdsourcing Technologies for Language and Cogniion Studies.*