

Respondo: un sistema di risposta automatico a supporto del customer care

Pierpaolo Basile^a, Marco de Gemmis^a, Giovanni Semeraro^a, Vito Manzari^b, Franco Grieco^b

^a Dipartimento di Informatica, Università degli Studi di Bari Aldo Moro

^b QuestionCube s.r.l.

pierpaolo.basile@uniba.it, marco.degemmis@uniba.it, giovanni.semeraro@uniba.it
vito.manzari@questioncube.com, franco.grieco@questioncube.com

Abstract

Cercare all'interno delle Frequently Asked Questions (FAQ) è un compito cruciale per poter fornire in maniera efficiente agli utenti una risposta ad una serie di problematiche. In questo lavoro presentiamo un sistema di risposta in grado di interpretare la domanda posta in linguaggio naturale dall'utente e cercare all'interno di una lista di FAQ le domande che meglio rispondono al bisogno informativo espresso dall'utente. Mostreremo l'applicazione di questo sistema intelligente sia in un contesto aziendale sia nella pubblica amministrazione.

1 Introduzione

Cercare all'interno delle Frequently Asked Questions (FAQ) - la lista delle domande più frequenti - pubblicate sulle pagine web è un'operazione cruciale: i clienti/utenti potrebbero essere confusi dalle domande non rilevanti ed essere demotivati nel non riuscire a trovare la domanda più idonea al problema da risolvere. Molto probabilmente la domanda che cerca non è nella lista, ma è espressa in modo diverso dal bisogno informativo che l'utente ha nella sua mente.

L'operazione di ritrovare una domanda all'interno di una lista di FAQ in base al bisogno informativo dell'utente non è un compito semplice e richiede lo sviluppo di sistemi intelligenti capaci di interpretare la domanda e cercare all'interno delle FAQ le domande più simili.

In molti scenari reali, sia aziendali sia relativi alla pubblica amministrazione, viene fornito supporto ai clienti/cittadini attraverso la lista delle FAQ. In genere, questa lista di domande risulta poco accessibile e obbliga l'utente a leggere una lunga lista di domande per cercare quella più vicina al problema riscontrato. Sarebbe molto più semplice avere un sistema dove il cliente/cittadino possa esprimersi in modo naturale ponendo il suo problema sotto forma di domanda in linguaggio naturale.

Respondo è un sistema di risposta automatico in grado di interpretare la domanda posta dall'utente e ricercare all'interno delle FAQ le domande più simili ordinate per rilevanza. L'utente potrà così selezionare una delle domande proposte e leggere la relativa risposta. Questo sistema di interrogazione permette all'utente di esprimersi in maniera naturale attraverso l'uso del linguaggio e permette all'azienda di fornire ri-

sposte "certificate" recuperate da una lista di FAQ redatta e gestita dal personale del customer care. In questo modo si è sicuri di fornire all'utente delle risposte corrette. Questo è fondamentale nel caso della pubblica amministrazione dove è necessario fornire informazioni corrette se queste sono collegate a procedure o normative.

Tuttavia, l'interazione in linguaggio naturale pone delle sfide poiché è necessario interpretare sintatticamente e semanticamente la richiesta dell'utente. Infatti, l'utente potrebbe esprimersi utilizzando un linguaggio differente da quello utilizzato per redigere le FAQ, oppure l'utente ha scarsa conoscenza del dominio.

In questo lavoro presentiamo un sistema di risposta automatico chiamato Respondo, sviluppato attraverso una collaborazione tra il Dipartimento di Informatica dell'Università degli Studi di Bari Aldo Moro e la società QuestionCube¹. Alcuni dettagli sulla metodologia proposta saranno descritti nella Sezione 2, invece la Sezione 3 descrive due applicazioni di Respondo sia in ambito aziendale sia nella pubblica amministrazione.

2 Metodologia

La tecnologia di Respondo si basa su un motore di Question Answering (QA)[Molino e Basile, 2012; Molino *et al.*, 2015] che utilizza la tecnica del passage retrieval. Il motore di Question Answering è composto da tre differenti moduli:

1. **NLP Pipeline:** questa componente si occupa di elaborare il testo utilizzando tecniche di elaborazione del linguaggio naturale. La pipeline è in grado di aggiungere annotazioni lessicali, sintattiche e semantiche al testo. Sia i testi indicizzati dal motore di QA sia le domande poste dagli utenti sono elaborate utilizzando questa pipeline. La pipeline è in grado di elaborare testi in lingua italiana e inglese;
2. **Search Engine:** questa componente recupera i documenti e le porzioni di testo candidate in base alla domanda posta dall'utente. Il motore di ricerca è in grado di recuperare tutte le porzioni di testo che sono rilevanti rispetto alla domanda posta dall'utente. La rilevanza è calcolata combinando differenti modelli di information retrieval.

¹<https://www.questioncube.com>

3. **Answer re-ranking:** questa componente effettua un re-ranking delle porzioni di testo candidate recuperate dal Search Engine. Il re-ranking utilizza una serie di similarità calcolate tra la domanda utente e il passaggio di testo sfruttando le annotazioni prodotte dalla NLP Pipeline. Una delle similarità utilizza tecniche basate sui modelli di semantica distribuzionale [Molino *et al.*, 2012].

Il motore di QA così realizzato può essere utilizzato sia per l'indicizzazione di documenti e sia per l'indicizzazione delle FAQ. In particolare, quando è utilizzato per cercare all'interno delle FAQ l'interazione con l'utente avviene in due passaggi. L'utente inserisce una domanda e il sistema di QA invece di restituire una lista di possibili risposte, restituisce la lista delle domande più simili presenti all'interno delle FAQ. L'utente in seguito alla visualizzazione delle FAQ più rilevanti può selezionarne una e visualizzare la relativa risposta associata. In questo modo è possibile restituire una risposta "certificata", cioè redatta dal personale del customer care.

È importante sottolineare che sia in fase di indicizzazione che di ricerca il sistema di QA utilizza anche il contenuto associato alle domande per poter capire in maniera più precisa quale è la domanda, tra le FAQ, più simile a quella inserita dall'utente.

Inoltre, il sistema è in grado di collezionare il feedback lasciato dall'utente in seguito ad ogni interazione l'utente può specificare se la risposta restituita è soddisfacente o no. I feedback raccolti sono utilizzati da un sistema di learning to rank che impara a riordinare le risposte candidate anche in base al feedback degli utenti. Il modulo di learning to rank è integrato all'interno della componente Answer re-ranking.

3 Applicazioni

In questa sezione mostreremo due applicazioni del motore Respondo. La prima applicazione riguarda AQP Risponde² che è un servizio offerto dall'Acquedotto Pugliese (AQP) ai suoi clienti. AQP Risponde permette ai clienti AQP di porre delle domande e ottenere una lista di FAQ rilevanti. Come descritto nella Sezione 2, AQP Risponde raccoglie il feedback degli utenti e lo utilizza per migliorare le sue prestazioni. Inoltre, i feedback e la cronologia delle richieste sono utilizzate per fornire agli operatori del call center una serie di statistiche utili a capire quali sono le problematiche più richieste e le attività di intervento finalizzate a migliorare la qualità delle FAQ che ricevono pochi feedback positivi.

Analogo sistema è stato utilizzato nel contesto della pubblica amministrazione. Utilizzando le FAQ rilasciate come open-data dal sito LineaAmica³ è stato creato PA Risponde. PA Risponde è un sistema analogo ad AQP Risponde le cui FAQ ricadono nel dominio della pubblica amministrazione. In particolare, LineaAmica raccoglie una serie di FAQ più frequenti riguardanti numerosi argomenti relativi alla PA. Le FAQ sono costruite utilizzando le domande più frequenti rivolte al call center LineaAmica. Il progetto PA Risponde è frutto di una collaborazione tra il Dipartimento di Informatica dell'Università degli Studi di Bari Aldo Moro e il Team

per la Trasformazione Digitale⁴. Attualmente PA Risponde è in fase di beta privata e sarà a breve rilasciato e integrato all'interno del portale LineaAmica.

In Figura 1 è mostrato un esempio di risposta relativa alla domanda *Sono trascorsi due anni dall'acquisto della prima casa, se vendo perdo le agevolazioni?*. Come mostrato in figura le due FAQ più simili restituite dal sistema sono: *Quando si perdono le agevolazioni "prima casa"?* e *In quali casi si perde il beneficio dell'agevolazione prima casa?*. È importante sottolineare che in questo caso nella domanda c'è un riferimento esplicito al fatto che l'utente ha acquistato la casa due anni fa, infatti la risposta alla prima FAQ contiene esplicitamente il contenuto *l'abitazione è venduta o donata prima che siano trascorsi cinque anni dalla data di acquisto*. Questo dimostra che il sistema utilizza anche il contenuto della risposta per poter recuperare la FAQ più rilevante.

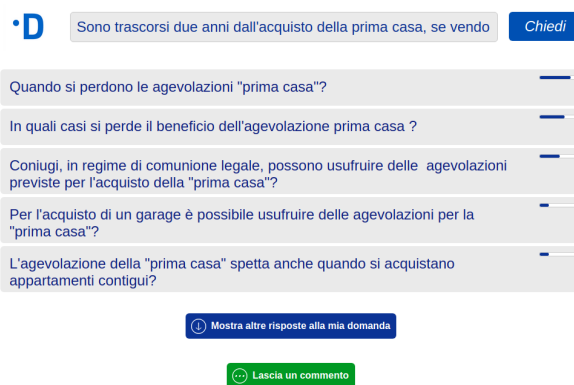


Figura 1: Esempio di interazione con PARisponde.

Riferimenti bibliografici

- [Molino *et al.*, 2012] Piero Molino, Pierpaolo Basile, Annalina Caputo, Pasquale Lops, e Giovanni Semeraro. Exploiting Distributional Semantic Models in Question Answering. In *Sixth IEEE International Conference on Semantic Computing, ICSC 2012, Palermo, Italy, September 19-21, 2012*, pages 146–153. IEEE Computer Society, 2012.
- [Molino *et al.*, 2015] Piero Molino, Pasquale Lops, Giovanni Semeraro, Marco de Gemmis, e Pierpaolo Basile. Playing with knowledge: A virtual player for Who Wants to Be a Millionaire? that leverages question answering techniques. *Artificial Intelligence*, 222(0):157 – 181, 2015.
- [Molino e Basile, 2012] Piero Molino e Pierpaolo Basile. QuestionCube: a framework for Question Answering. In Giambattista Amati, Claudio Carpineto, e Giovanni Semeraro, editors, *Proceedings of the 3rd Italian Information Retrieval Workshop, Bari, Italy, January 26-27, 2012*, volume 835 of *CEUR Workshop Proceedings*, pages 167–178. CEUR-WS.org, 2012.

²<http://aqprispnde.aqp.it/ask.php>

³<http://www.lineaamica.gov.it/>

⁴<https://teamdigitale.governo.it/>