



POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Gestionale

Christian Baratto

Metodi per l'individuazione dei Requisiti di prodotto
con algoritmi di text-mining: un'applicazione nel
settore della robotica collaborativa

Relatore:

Franceschini Fiorenzo

Correlatore:

Mastrogiacomo Luca

Ottobre 2021

Chi ha un perché abbastanza forte può superare qualsiasi cosa.

Friedrich Nietzsche

KEYWORDS

User-Generated Contents

Requirements Analysis

Collaborative Robots

Topic modelling

Small Data

Big Data

SOMMARIO

ABSTRACT	6
LISTA DEGLI ACRONIMI.....	7
1. INTRODUZIONE.....	8
1.1. LA ROBOTICA COLLABORATIVA	9
1.1.1. <i>Panoramica di mercato</i>	12
1.1.2. <i>Principali tendenze di mercato</i>	13
1.2. CHE COSA SONO I REQUISITI DI PRODOTTO.....	14
1.3. IL PROCESSO DI DEFINIZIONE DEI REQUISITI	16
1.4. LA CASA DELLA QUALITÀ	18
2. METODOLOGIA APPLICATA	29
2.1. PREPARAZIONE DELLA BASE DATI.....	29
2.1.1. <i>User Generated Content</i>	29
2.1.2. <i>Scraping</i>	31
2.1.3. <i>Data cleaning</i>	32
2.2. ANALISI E MODELLAZIONE DEI DATI.....	37
3. METODOLOGIA DI ANALISI	53
3.1. INTRODUZIONE AL PROBLEMA	53
3.2. SMALL DATA E BIG DATA.....	55
3.3. DISACCOPIAMENTO DEI REQUISITI FUNZIONALI DALLE FUNZIONALITÀ DEL PRODOTTO	60
4. APPLICAZIONE NEL SETTORE DELLA ROBOTICA COLLABORATIVA.....	62
4.1. REQUISITI DI PRODOTTO	62
4.2. FUNZIONALITÀ DEL PRODOTTO	74
4.3. CONFRONTO REQUISITI-FUNZIONALITÀ	84
4.4. CONFRONTO DELLA SOVRAPPOSIZIONE DEI TOPIC	86
4.4.1. <i>Approccio del matching tra requisiti e funzionalità</i>	87
4.4.2. <i>Approccio con indice di Jaccard</i>	89
4.4.3. <i>Approccio basato sulla frequenza delle keywords</i>	93
4.4.4. <i>Confronto tra metodi</i>	100
5. SVILUPPI FUTURI.....	102
5.1. APPROCCIO CON MARKOV CHAIN	102
5.2. APPROCCIO CON REGULAR EXPRESSION	106
6. CONCLUSIONI	110
7. ALLEGATI.....	112
ALLEGATO 1.....	112

ALLEGATO 2.....	120
ALLEGATO 3.....	121
ALLEGATO 4.....	124
ALLEGATO 5.....	125
ALLEGATO 6.....	128
ALLEGATO 7.....	131
ALLEGATO 8.....	133
8. BIBLIOGRAFIA	135
9. SITOGRAFIA	137

ABSTRACT

Il processo di definizione dei requisiti funzionali di un prodotto richiede considerevoli impieghi di risorse e di tempo, soprattutto per quanto riguarda la fase di acquisizione di informazioni. Attualmente, le informazioni disponibili sul web sono in crescita esponenziale. Secondo l'International Data Corporation (IDC), entro il 2025 saranno presenti 175 ZB (1 Zettabyte = 10^{21} bytes) di dati nel mondo, con una percentuale significativa di informazioni rilevanti per l'azienda provenienti da forme non strutturate, principalmente testuali.

In una realtà dominata dall'avvento di tecnologie intelligenti, tali contenuti web possono essere manipolati, in modo da risultare adatti a qualunque tipo di contesto.

Si è voluto studiare un metodo basato sull'analisi di contenuti web che potesse rendere più agevole la generazione dei requisiti funzionali in modo da velocizzare le fasi di studio e ingegnerizzazione del prodotto. La metodologia proposta è stata applicata al settore della robotica collaborativa per approfondire la collaborazione uomo-macchina al fine di poterla comprendere e migliorare.

LISTA DEGLI ACRONIMI

- **AMR** (*Autonomous Mobile Robot*)
- **API** (*Application Programming Interface*)
- **ARM** (*Arduino Robot Arm*)
- **CAGR** (*Compound Annual Growth Rate*)
- **COBOT** (*COllaborative roBOT*)
- **CSS** (*Cascading Style Sheets*)
- **CTO** (*Chief Technical Officer*)
- **DOM** (*Document Object Model*)
- **HOQ** (*House Of Quality*)
- **HRC** (*Human-Robot Collaboration*)
- **IOT** (*Internet of Things*)
- **IREX** (*International Robot EXhibition*)
- **NLP** (*Natural Language Processing*)
- **NRI** (*National Robotics Initiative*)
- **PMI** (*Piccole Medie Imprese*)
- **QFD** (*Quality Function Deployment*)
- **RBR50** (*Robotics Business Review*)
- **ROI** (*Return On Investment*)
- **ROS** (*Robot Operating System*)
- **RPCXML** (*Remote Procedure Call - eXtensible Markup Language*)
- **RTDE** (*Real-Time Data Exchange*)
- **STM** (*Structural Topic Modeling*)
- **UGC** (*User Generated Content*)
- **VBA** (*Visual Basic for Application*)

1. INTRODUZIONE

Gli avanzamenti nella tecnologia robotica sicura e collaborativa e la crescente disponibilità di hardware per robot hanno facilitato l'emergere di numerosi nuovi fornitori che entrano nel mercato con robot industriali collaborativi relativamente economici (*Ionescu, 2020*). Il settore manifatturiero tenderà a subire grossi cambiamenti nei prossimi decenni. Iniziative come l'Industria 4.0, Digital Agenda e l'IOT introdurranno nuove tecnologie che porteranno alla creazione di sistemi altamente connessi e integrati. Questi cambiamenti produrranno nuove metodologie di lavoro e genereranno nuove opportunità per gestire l'innovazione e la flessibilità dei processi industriali. Nello specifico, grazie allo sviluppo della robotica, sarà possibile far lavorare in modo collaborativo robot e operatore in modo da massimizzare i benefici dei processi manuali e dei processi automatizzati (*Mokaram et al., 2017*).

Nelle applicazioni industriali, l'eterogeneità dei sistemi robotici sta aumentando la necessità di competenze specializzate ostacolando la democratizzazione della tecnologia (ossia più accessibile a diversi utenti non esperti semplificando, senza omogeneizzare, la programmazione e riducendo in generale i costi di esercizio): uno dei principali problemi è che i linguaggi di programmazione di ogni ecosistema aziendale e gli ambienti proposti dal mercato sono raramente compatibili tra loro, generando un aumento della competitività che porta alla necessità di trovare possibilità di differenziazione da parte delle aziende del settore (*Ionescu, 2020*).

Per riuscire a far ciò si è voluto studiare un nuovo approccio alla definizione dei requisiti che permetta di snellire e velocizzare la fase di ingegnerizzazione di prodotto da parte delle aziende produttrici della tecnologia robotica collaborativa.

Una delle parti più onerose in termini di tempo e risorse impiegate è, infatti, la definizione dei requisiti utente, basata su uno studio approfondito delle esigenze di mercato ma che spesso può risultare fallimentare a causa di errori metodologici nell'acquisizione e manipolazione dei dati (*Jung, 1998*). Grazie all'attuale sviluppo tecnologico, con la nascita dell'intelligenza artificiale, le fasi di manipolazione dei dati sono sempre più precise, veloci e talvolta automatizzabili da algoritmi ben specifici.

In questo lavoro di tesi è stata analizzata la possibilità di generare un nuovo metodo di definizione dei requisiti funzionali che permettesse di coniugare precisione, velocità di esecuzione e automatizzazione algoritmica.

Si è adottata, nel contesto della robotica collaborativa, la metodologia proposta nei primi capitoli di questo documento per la definizione di una scheda di requisiti funzionali dei bracci robotici collaborativi.

Il documento nel primo capitolo analizza l'attuale metodo di definizione dei requisiti in modo da metterne in luce l'assetto algoritmico e gli aspetti migliorabili.

Nel secondo capitolo viene spiegata la metodologia operativa applicata per la prima fase di analisi attraverso l'utilizzo del text-mining, mentre nel terzo capitolo si spiega l'approccio adottato per l'analisi e per la definizione di una scheda dei requisiti robusta e ben definita.

Nel quarto capitolo vengono mostrati i risultati dell'analisi fatta nel mondo della robotica collaborativa e nel quinto capitolo vengono proposte tecniche di automatizzazione algoritmica dell'approccio proposto, implementabili in qualunque software open source; nello specifico, sono stati proposti dei codici in ambiente R.

Nel sesto capitolo vengono riportate le conclusioni del lavoro svolto e le possibili implicazioni che può avere in ambito aziendale e accademico.

1.1. LA ROBOTICA COLLABORATIVA

La robotica collaborativa è un termine generico che trasmette l'idea generale di prossimità tra macchine ed esseri umani per alcuni compiti utili in uno spazio condiviso, con una gamma di opzioni per il tempo (continuo, sincrono, alternato, ecc.). L'attuale stato dell'arte nella robotica collaborativa per applicazioni industriali è il risultato di una lunga eredità di ricerca e sviluppo in



Figura 1. Robot Collaborativi

Fonte: (<https://www.bastiansolutions.com/blog/collaborative-robots-part-2-benefits-and-expanding-capabilities/>)

principi e meccanismi di attuazione dagli anni 2000 in poi, insieme a un'abbondante letteratura sul miglioramento delle prestazioni di controllo, modalità di interazione intuitive e percezione con sensori (Vicentini, 2017).

I robot collaborativi (COBOT) vengono utilizzati per creare valore lavorando in un ambiente in cui vi è la presenza di operatori.

Il design del robot collaborativo è simile ai bracci robotici industriali, come si vede in Figura 1. La differenza sta nell'aver molti più assi, attualmente fino ad un massimo di otto assi, dunque un numero di movimenti consentito maggiore che permette di raggiungere al meglio i punti nello spazio del raggio d'azione e l'arrotondamento degli angoli per garantire una massima sicurezza in presenza dell'operatore.

L'aggiornamento 2017 alla strategia industriale dell'UE ha sintetizzato l'Europa come leader mondiale in molti settori, in particolare nei prodotti e servizi ad alto valore aggiunto, a basse emissioni di carbonio e sofisticati in cui i robot collaborativi possono aggiungere produttività e sostegno alla forza lavoro qualificata e di talento. I robot industriali tradizionali non sono del tutto utili per questo, infatti, non sono in grado di lavorare in modo cooperativo con gli esseri umani in uno spazio di lavoro condiviso. In questa visione, i robot collaborativi hanno un alto potenziale: i COBOT combinano la destrezza, la flessibilità e la capacità di risoluzione dei problemi degli esseri umani con la forza, la resistenza, la qualità, la coerenza e la precisione dei robot e possono essere ridistribuiti ad altri compiti molto più facilmente (Vanderborght, 2019). Per tale ragione, i COBOT sono molto adatti alle PMI perché hanno una configurazione rapida, non hanno bisogno di gabbie protettive, occupano poco spazio sul pavimento, sono facili da usare e consentono un impiego flessibile per molteplici usi e piccoli lotti (Vanderborght, 2019).

Nella Tabella 1 si riporta un confronto delle principali caratteristiche tra bracci robotici tradizionali e collaborativi.

Fattore di qualità	Robot Industriale Tradizionale	Robot Collaborativo
Sicurezza	Dannoso, richiede protezioni	Progettato per essere sicuro
Lavoro con l'uomo	Non consentito, richiede barriere	Possibile con le dovute valutazioni e precauzioni
Difficoltà programmazione	Elevata	Contenuta
Payload	Elevato	Ridotto
Sbraccio	Lungo	Limitato
Velocità	Elevata	Limitata
Adattabilità	Bassa: ottimo per grandi volumi e processi limitati	Elevata: ottimo per piccoli volumi e processi eterogenei
Prezzo	Elevato	Ridotto

Tabella 1. Principali differenze tra i robot tradizionali e quelli collaborativi
Fonte: (Giardinelli, 2018)

Secondo gli standard internazionali ISO 10218 parte 1 e parte 2, esistono quattro tipi di funzioni collaborative per i robot:

- Arresto monitorato di sicurezza: permette l'interruzione della mansione qualora un operatore si avvicini troppo allo spazio di manovra del COBOT
- Guida a mano: conosciuta come modalità *Free Drive*, viene utilizzata per insegnare rapidamente il percorso che il COBOT dovrà compiere per ogni tipo di mansione attraverso una sensoristica di movimenti e forze molto precisa
- Monitoraggio della velocità e dell'ambiente circostante: permette di diminuire la velocità d'esecuzione della mansione qualora in un certo raggio d'azione venga percepita la presenza di un operatore. In questo caso il robot collaborativo rallenterà la sua velocità, a differenza dell'arresto forzato precedentemente descritto che scatto qualora vengano violato i parametri di sicurezza del robot. Si pensi ad esempio ad un COBOT che percepisce un operatore a 5 metri di distanza; in questo caso la velocità d'esecuzione diminuirà per tutelare l'operatore. Diverso è il caso in cui il COBOT percepisce l'operatore a 1 metro di distanza, in

questo caso vengono violata i limiti di sicurezza che porteranno il robot collaborativo ad arrestare la mansione

- Limitazione di potenza e forza: il robot può percepire forze anomale ed ostacoli nel suo cammino. Una volta che vengono rilevati il robot è programmato per arrestarsi o adattare la corsa in base alla presenza dell'ostacolo.

(<https://blog.robotiq.com/what-does-collaborative-robot-mean>, Giardinelli, 2018).

1.1.1. Panoramica di mercato

Secondo un nuovo rapporto della società di consulenza sul mercato tecnologico globale ABI Research, il mercato dei COBOT è destinato a crescere notevolmente nel prossimo decennio. Il mercato ha avuto una valutazione globale di \$ 475 milioni nel 2020, che si espanderà a \$ 600 milioni nel 2021 e \$ 8 miliardi nel 2030, con un CAGR previsto del 32,5%. I governi, inoltre, stanno incoraggiando l'adozione della robotica prendendo iniziative a sostegno dello sviluppo delle moderne tecnologie nel mercato della robotica. Ad esempio, il governo federale degli Stati Uniti ha avviato un programma chiamato National Robotics Initiative (NRI) per rafforzare le capacità di costruzione di robot domestici nel paese e incoraggiare le attività di ricerca sul campo.

(https://www.roboticsbusinessreview.com/rbr/how_will_president_trump_affect_robotics_industry/)

Attualmente, nel settore della robotica collaborativa Cina e America hanno le quote di mercato più elevate:

- C'è stata un'adozione significativa della chirurgia robotica negli Stati Uniti. Il volume annuale della procedura per la chirurgia robotica è superiore a 5000000 negli Stati Uniti. Al fine di costruire l'immagine del marchio e il vantaggio competitivo, la chirurgia robotica è diventata il dispositivo medico più rapidamente adottato negli ultimi due decenni.
- Il motore principale per la crescita dei robot nella regione sono le industrie manifatturiere che sono attualmente tendono ad automatizzare i loro processi di produzione, al fine di rafforzare le industrie nei mercati nazionali ed internazionali.

Universal Robots (UR) è attualmente l'attore dominante nel mercato, con il 50% delle vendite totali e un fatturato di \$ 219 per il 2020. Nel primo trimestre del 2021, UR ha guadagnato \$ 66 milioni dalle vendite di bracci COBOT, con un aumento del 32% annuo.

(<https://www.therobotreport.com/cobot-market-grow-8b-2030-report-finds/>)

Al momento, molte aziende europee di robotica hanno COBOT nei loro portafogli di prodotti, ad esempio Comau (robot Aura), Festo (Bionic Handling Assistant, BionicCobot), ABB (Yumi e Roberta, quest'ultima precedentemente di Gomtec), Bosch (APAS), Stäubli (TX2-60) e Mabi Robotic (Speedy 6 e 12) *(Vanderborght, 2019)*.

Le industrie che utilizzano i robot devono affrontare una pressione costante, principalmente per ridurre i costi operativi, aumentando la produttività e mantenendo i livelli di qualità richiesti, più convenienti, più facili da usare e meno complessi per scopi di formazione. Ciò offre principalmente più possibilità alle organizzazioni, aumentando la spinta e la domanda di questi robot. I robot collaborativi consentono alle industrie di tutte le dimensioni e scale di rimanere competitive poiché utilizzano i sensori più recenti, le tecnologie plug and play e la programmazione automatica dei robot dai dati CAD. In questo modo, le aziende riescono a migliorare il loro processo produttivo in modo estremamente efficiente e ad alto valore aggiunto sia per il cliente sia per il lavoratore. *(<https://www.cobottrends.com/china-us-to-drive-cobot-sales-post-covid/>, <https://www.automate.org/news/sepro-group-and-universal-robots-announce-new-cobot-partnership>)*

Inoltre, i fornitori sul mercato sono attivi anche in fusioni e acquisizioni che stanno portando a generare dei veri e propri ecosistemi tecnologici in grado di offrire soluzioni particolarmente complesse, all'avanguardia ed efficienti operativamente.

(<https://www.universal-robots.com/blog/the-benefits-of-an-ecosystem/>, <https://www.universal-robots.com/blog/tags/human-cobot-collaboration/>)

1.1.2. *Principali tendenze di mercato*

Applicando la procedura metodologica che verrà descritta nei successivi capitoli ai blog analizzati (si veda Allegato 4 per i link utili), è stato possibile sintetizzare alcune tendenze di mercato che sono di seguito riportate:

- Le soluzioni di automazione possono ridurre drasticamente il costo delle consegne. Supportate da questo ci sono le innovazioni robotiche in atto, che consentono alle imprese di vari settori di operare in modo autonomo, con capacità e capacità umane aumentate.
- La formazione e l'inserimento più lunghi, l'aumento dei benefici e dei tassi di compensazione e la carenza di manodopera sono fattori significativi che guidano la distribuzione. Sempre più strutture di magazzino, distribuzione e adempimento stanno investendo in soluzioni automatizzate. Man mano che la tecnologia migliora e le applicazioni diventano più ampie e flessibili, la robotica viene adottata da un numero più significativo di operazioni di produzione in tutti i contesti.
- I robot svolgono un ruolo cruciale nell'automazione industriale, con molte operazioni fondamentali nei settori gestiti da loro. I robot possono svolgere compiti complicati e ripetitivi con precisione, anche in ambienti pericolosi. Proprio per questo motivo, sono stati impiegati nelle unità di produzione della maggior parte dei settori.
- I robot industriali utilizzati nell'industria automobilistica sono più efficienti e utilizzati in modo conveniente, grazie ai loro nuovi livelli di velocità, accuratezza, precisione, flessibilità e agilità. La complessità della produzione di autovetture è aumentata negli ultimi dieci anni. Ciò ha portato all'utilizzo di robot di soluzioni di automazione per una parte sostanziale dei processi di produzione in questi giorni.

1.2. CHE COSA SONO I REQUISITI DI PRODOTTO

Tutti i processi di progettazione bene organizzati dovrebbero iniziare con la stesura di un documento di requisiti. Un *requisito* è una proprietà richiesta, oppure auspicabile, del prodotto. Il documento dei requisiti ha lo scopo di accogliere in forma sintetica una descrizione di tutte le proprietà desiderate. Dalla sua formulazione dovrebbe essere chiaro se un requisito esprime una proprietà obbligatoria, oppure soltanto suggerita o auspicabile.

I requisiti possono essere di vario tipo: i *requisiti funzionali* descrivono le funzioni che il sistema deve realizzare, i *requisiti non funzionali* descrivono proprietà che il prodotto dovrà possedere per renderlo all'avanguardia e adatto all'uso da parte di un generico utente. Lo scopo della definizione

dei requisiti è individuarli e descriverli nel modo più specifico e meno ambiguo possibile. Lo standard ISO 13407 suggerisce che, per identificare i requisiti rilevanti, in un processo di progettazione human-centred, si considerino i seguenti aspetti:

- ✓ le prestazioni richieste al nuovo sistema in relazione agli obiettivi operativi ed economico/finanziari
- ✓ i requisiti normativi o legislativi rilevanti, compresi quelli relativi alla sicurezza e alla salute
- ✓ la comunicazione e la cooperazione fra gli utenti e gli altri attori rilevanti
- ✓ le attività degli utenti, inclusa la ripartizione dei compiti, il loro benessere e le loro motivazioni
- ✓ le prestazioni dei diversi compiti
- ✓ la progettazione dei flussi di lavoro e dell'organizzazione
- ✓ la gestione del cambiamento indotto dal nuovo sistema, incluse le attività di addestramento e il personale coinvolto
- ✓ la fattibilità delle diverse operazioni, comprese quelle di manutenzione
- ✓ La progettazione dei posti di lavoro e la interfaccia uomo-computer

I requisiti vengono prodotti da persone che lavorano in stretto contatto con il committente per individuarne i bisogni in relazione al sistema da realizzare (o da migliorare, se si tratta di una riprogettazione). Possono essere stesi direttamente dai progettisti, o da persone che non necessariamente saranno coinvolte nel progetto successivo.

È importante non confondere l'attività di stesura dei requisiti con l'attività di progettazione: quando si specificano i requisiti di un prodotto si stanno *ponendo dei vincoli all'attività di progettazione*. In sostanza, lo scopo del documento è di indicare *che cosa* deve essere realizzato e *perché*, non *come* deve essere realizzato.

1.3. IL PROCESSO DI DEFINIZIONE DEI REQUISITI

La fase di definizione dei requisiti può essere suddivisa in tre attività fondamentali, come riportato in Figura 2, che si possono chiamare *esplorazione*, *organizzazione* e *revisione*.



Figura 2. Metodologia standard di definizione dei requisiti

Fonte: (Polillo, 2010)

1. *Esplorazione*: le persone incaricate di produrre il documento dei requisiti raccolgono il maggior numero possibile d'informazioni sugli obiettivi e sulle necessità riguardo al sistema da costruire. Le informazioni vengono raccolte da fonti diverse; in primo luogo, dal committente, cioè colui che ha avviato il progetto e che ne costituisce il riferimento principale; in secondo luogo, dalle interviste con gli *stakeholder* del prodotto, cioè tutti coloro che, in un modo o nell'altro, hanno qualche interesse nel prodotto, o la cui attività viene influenzata, direttamente o indirettamente, da esso; infine, dall'analisi della concorrenza, cioè di quei prodotti con i quali il prodotto in costruzione dovrà confrontarsi e competere.

La fase di esplorazione, nella quale vengono raccolti i requisiti, presenta spesso notevoli difficoltà. I problemi sono sostanzialmente di quattro tipi:

- *Problemi di ambito*. Generalmente, i "contorni" del sistema da progettare non sono ben definiti, ed esiste sempre il rischio di ampliare eccessivamente il campo di esplorazione. Restringendo troppo i temi da approfondire si rischia di tralasciare aspetti che potrebbero rivelarsi importanti nelle fasi successive. Inoltre, nelle conversazioni con gli stakeholder si è spesso tentati di abbozzare delle soluzioni di progetto. Solitamente in questa fase ci si deve limitare alla sola raccolta dei requisiti, lasciando le attività di progettazione alle fasi successive,

quando tutti i requisiti saranno stati individuati e organizzati in un insieme coerente.

- *Problemi di comprensione.* Da un lato, gli utenti hanno spesso una comprensione solo parziale dei loro bisogni, e una conoscenza piuttosto limitata delle possibilità offerte dalla tecnologia. Dall'altro, chi raccoglie i requisiti ha spesso una conoscenza limitata del dominio del problema, e utilizza un linguaggio differente da quello degli utenti e degli altri stakeholder. Ogni interlocutore tende a tralasciare gli aspetti che per lui sono ovvi, ma che potrebbero non esserlo affatto per interlocutori differenti.
- *Problemi di conflitto.* Stakeholder diversi possono avere punti di vista diversi sul sistema che dovrà essere progettato. Questi punti di vista potrebbero essere fra loro incompatibili e dovranno essere fatti emergere con chiarezza.
- *Problemi di volatilità.* I requisiti evolvono nel tempo. Infatti, il contesto del sistema può mutare anche molto rapidamente e in modo inaspettato. Questi cambiamenti possono riguardare le condizioni di mercato, ricambio del management o ristrutturazioni dell'organizzazione committente, acquisizioni o vendite societarie, evoluzioni della tecnologia, e così via. Tutti questi fatti possono modificare in modo rilevante le priorità dei diversi requisiti, o addirittura modificarli completamente nel corso del progetto.

2. *Organizzazione:* durante la fase precedente, vengono raccolti appunti e materiale informativo vario, che dovranno successivamente essere riesaminati, selezionati e organizzati. L'obiettivo principale di questa fase è costruire un *documento di specifica dei requisiti*, condiviso e approvato dal committente che sarà il riferimento principale per tutte le attività successive del progetto. Lo scopo di questo documento è di descrivere, nella forma più completa possibile, le richieste del committente e i vincoli che dovranno essere rispettati nelle fasi successive del progetto. Si analizza il materiale raccolto nella fase di esplorazione, lo si riordina, si risolvono eventuali contraddizioni (le persone intervistate potrebbero avere idee molto diverse su ciò che occorre fare), e si produce una prima bozza del documento dei requisiti.

3. *Revisione e approvazione*: la bozza del documento dei requisiti prodotta viene presentata al committente per la sua approvazione. Di solito, sarà necessario effettuare diversi aggiustamenti e revisioni del documento, prima che questo possa essere considerato sufficientemente consolidato e stabile per procedere alla successiva fase di progettazione. In un processo iterativo, come già più volte ricordato, il documento dei requisiti non potrà mai, comunque, considerarsi finale: in ogni momento successivo alcuni aspetti potranno essere rivisti e modificati, sulla base delle nuove informazioni acquisite in corso di progetto. Polillo (2010)

1.4. LA CASA DELLA QUALITÀ

Fra i vari strumenti e metodi utilizzabili nello sviluppo di nuovi prodotti vi è la Casa della Qualità che fa parte del Quality Function Deployment (QFD). In generale il QFD è considerato uno strumento per lo sviluppo del prodotto che traduce in modo sistematico i requisiti del cliente in linee guida per il processo di progettazione e produzione creando in modo sistematico nuovo valore per il cliente (Mazur, 2014).

Attualmente, la HOQ è lo strumento più utilizzato nella manifattura per analizzare in modo efficace ed efficiente i requisiti. Una delle ragioni di ciò risiede nel modo in cui QFD è stato introdotto nella cultura occidentale (Govers, 2001). QFD è stato introdotto come strumento di progettazione in Giappone, e la filosofia del miglioramento continuo (kaizen) incorporata nella cultura giapponese ha avuto un'influenza sulla sua diffusione. Questa filosofia non fu ereditata durante l'adozione del QFD nella cultura occidentale e per questo tale metodo si è ridotto ad un semplice (ma efficace) "strumento di progettazione" (Akao & Mazur, 2003), come riportato in Figura 3.

Mettendo a confronto il settore manifatturiero e quello dei servizi si notano delle sostanziali

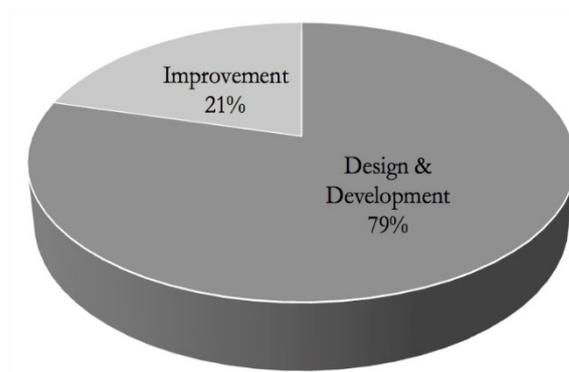


Figura 3. Applicazioni QFD – Design & Development
Fonte: (Erdil & Arani, 2018)

differenze. Mentre il QFD viene utilizzato nel settore manifatturiero principalmente nella progettazione e nello sviluppo, nel settore dei servizi sta aumentando la tendenza ad usare tale metodo come strumento di controllo per il miglioramento continuo, come mostrato in Figura 4. Ciò, è attribuibile al crescente tasso di adozione del QFD nelle aree non legate all'industria manifatturiera negli ultimi anni e ad un maggior numero di pubblicazioni sul QFD nell'industria dei servizi che hanno permesso di distaccarsi dalla teoria diffusa in occidente ed innovarla. (Erdil & Arani, 2018).

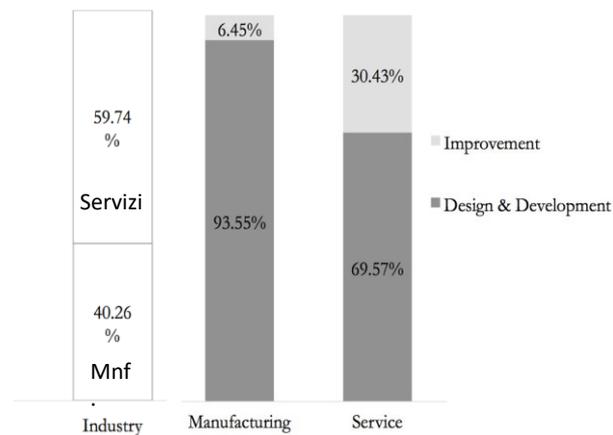


Figura 4. Applicazioni QFD - Industrie manifatturiere e dei servizi
Fonte: (Erdil & Arani, 2018)

Il Quality Function Deployment consiste nella compilazione di una serie di matrici correlate tra cui la più importante è la House of Quality (HOQ), che rappresenta la costruzione centrale del QFD. La costruzione della HOQ prevede sei passi (Grandi, 2017):

1. Identificare i bisogni (attributi) del cliente
2. Identificare le caratteristiche tecniche, cioè attributi espressi in linguaggio tecnico
3. Mettere in relazione gli attributi del cliente con le caratteristiche tecniche
4. Valutare i prodotti della concorrenza
5. Valutare le caratteristiche tecniche e i target di sviluppo, determinati in base agli ordini di importanza per il cliente
6. Determinare quali caratteristiche tecniche sviluppare nel prodotto

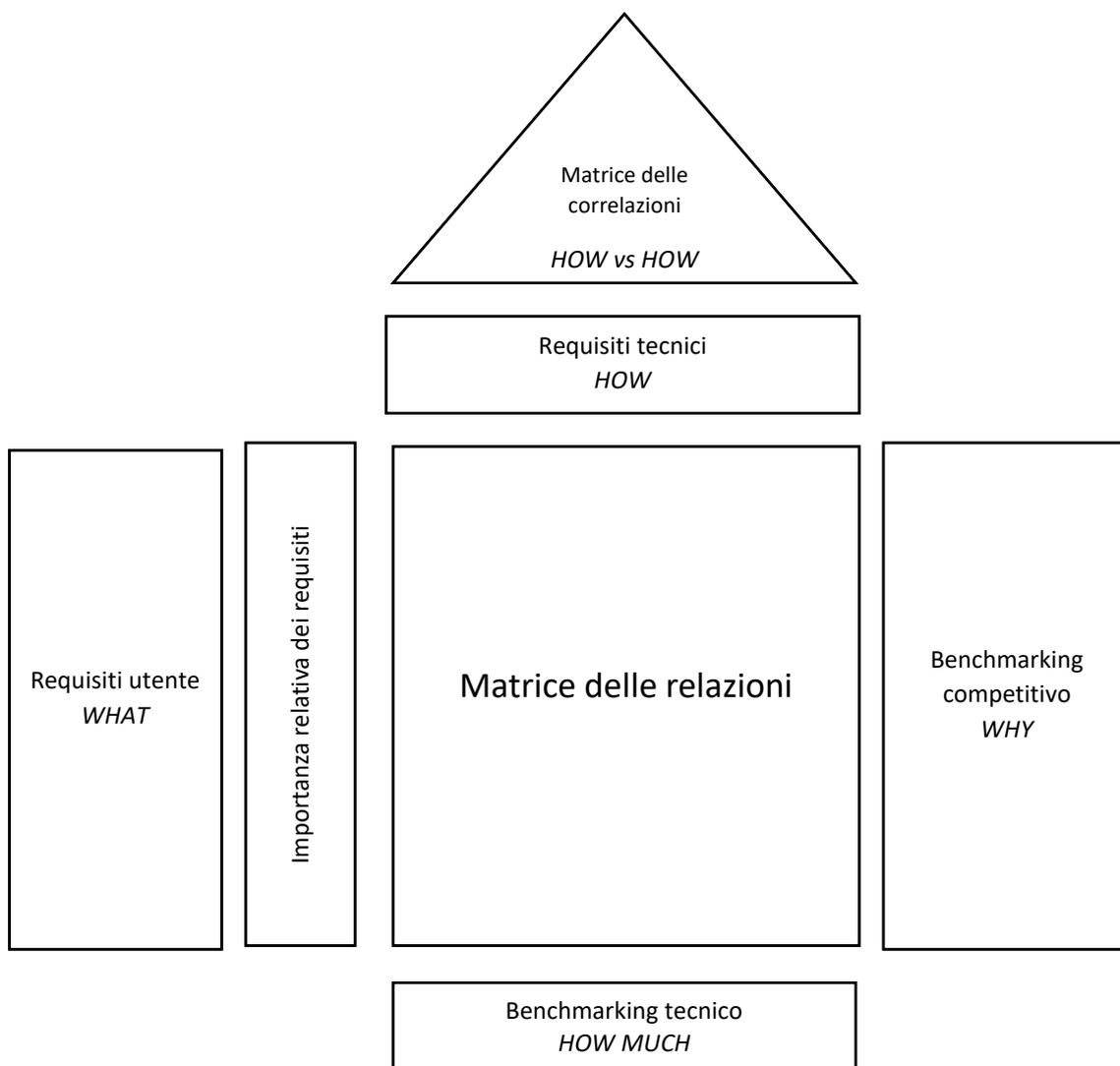


Figura 5. Matrici della Casa della Qualità

La Casa della Qualità è un diagramma a matrice come mostrato in Figura 5.

Di seguito si riporta una breve descrizione delle matrici componenti l'HOQ (Grandi, 2017, Franceschini et al., 2020):

- I Requisiti utente: è la parte sinistra della HOQ e contiene una lista strutturata dei bisogni del cliente per quel che riguarda il prodotto in considerazione.
- Il Benchmark Competitivo: permette di paragonare, in modo oggettivo, i dati riguardanti il prodotto della concorrenza rispetto a quello dell'azienda: è l'istanza destra della HOQ che permette di paragonare le performance del prodotto che il gruppo sta sviluppando, con la performance della concorrenza, e sviluppare una strategia volta sia ad ottimizzare l'abilità dell'azienda nel vendere il prodotto, sia ad ottenere la soddisfazione del cliente nel lungo periodo. In questa sezione sono presenti diverse voci:
 - a. Importanza per il cliente, relativamente ai requisiti identificati
 - b. Performance di soddisfazione del cliente: rappresenta la percezione del cliente di quanto il prodotto attuale, che si intende da modificare o sostituire, incontra i bisogni del cliente
 - c. Performance di soddisfazione dei prodotti della concorrenza
 - d. Obiettivi da raggiungere
 - e. Rapporto di miglioramento: serve per compiere il trade-off decisionale su quali caratteristiche andare a modificare massimizzando il posizionamento dell'azienda rispetto alla concorrenza. Viene calcolato nel seguente modo:

$$\text{Rapporto di Miglioramento} = \frac{\text{Obiettivi}}{\text{Performance di Soddisfazione del Cliente}} \quad (1.1)$$

Esempio:

In riferimento alla Figura 8, al Bisogno del Cliente "Ampiezza di copertura del sistema di protezione", il rapporto di miglioramento è così calcolato:

$$\text{Rapporto di Miglioramento} = \frac{4,60}{3} = 1,53$$

- f. Sales point, definiti come le informazioni che caratterizzano la capacità di vendere il prodotto, rappresentano delle metriche di valutazione per l'order qualifying

- g. Riga del Peso che rappresenta il sunto della rilevanza complessiva di ogni bisogno del cliente ed è calcolabile nel seguente modo:

$$\begin{aligned} \text{Riga del Peso} &= \text{Importanza per il cliente} * \text{Rapporto di Miglioramento} * \text{Sales Point} = \\ &= \text{Importanza per il cliente} * \frac{\text{Obiettivi}}{\text{Performane di Soddisfazione del Cliente}} * \text{Sales Point} \quad (1.2) \end{aligned}$$

Spesso può essere utile ridimensionare i valori in una scala da 0 a 1 attraverso la normalizzazione, come di seguito rappresentato:

$$\text{Riga del peso normalizzata} = \frac{\text{Riga del Peso}}{\sum \text{Riga del Peso}} \quad (1.3)$$

- Il Benchmarking Tecnico: descrive le caratteristiche tecniche che il nuovo prodotto dovrà avere e sono collocate nella parte bassa della HOQ. In questa sezione sono identificati i Target: rappresentano un insieme di valutazioni, relative alle caratteristiche tecniche che permettono di comprendere se l'azienda vuole impegnarsi ad affiancare, superare o ottenere una leadership per tale caratteristica in esame.
- La Matrice delle relazioni e le Priorità: in questa parte viene espressa una valutazione soggettiva dai progettisti sul legame esistente tra caratteristiche tecniche e bisogni del cliente. Solitamente si attribuiscono dei simboli ai quali viene dato un grado di rilevanza che andrà ad influenzare le Priorità, calcolabili nel seguente modo:

$$\text{Priorità}_j = \sum_i \text{Grado di rilevanza}_{ij} \quad (1.4)$$

dove i rappresenta il bisogno del cliente e j rappresenta la caratteristica tecnica.

- Matrice delle Correlazioni: situata nella parte alta della HOQ evidenzia le relazioni esistenti tra ogni coppia di caratteristiche tecniche. Sono utili per comprendere quali determinanti della qualità possono andare ad influenzarsi tra loro in modo positivo o negativo.

(Grandi, 2017, Franceschini et al., 2020)

Di seguito si riporta un esempio riguardante un prodotto elettronico: l'inverter vettoriale. Gli inverter sono dispositivi elettronici per l'alimentazione e il controllo di motori in corrente alternata monofase o trifase, applicabili nel mondo della robotica collaborativa e industriale.

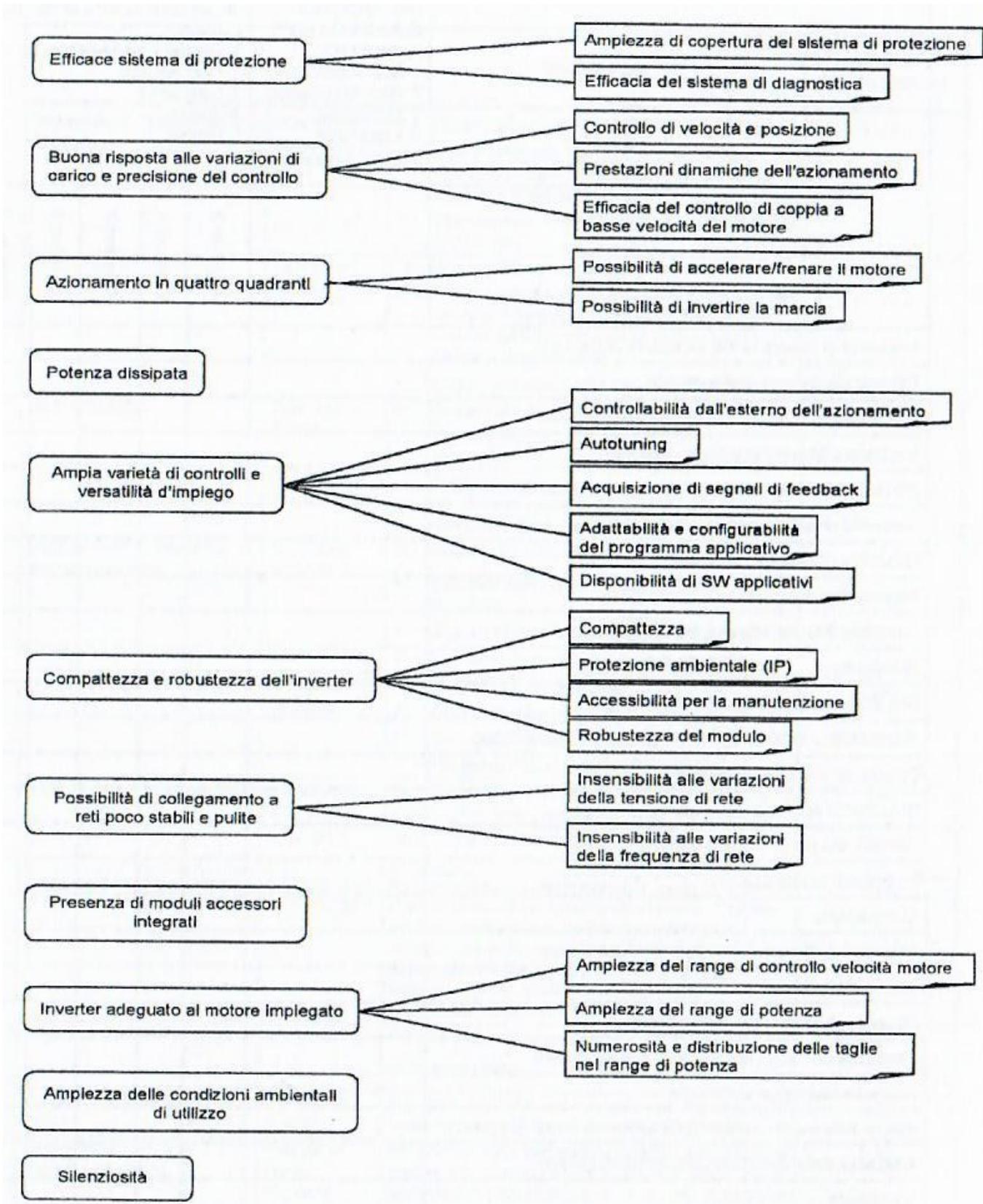


Figura 6. Esplosione ad albero dei bisogni dei clienti per un inverter vettoriale
 Fonte: (Grandi, 2017)

Attributi dell'inverter da valutare	Valutazione dell'importanza dell'attributo per l'intervistato: 1 Non Importante. 2 Poco Importante. 3 Sufficiente. 4 Importante. 5 Molto Importante.	Valutazione della rispondenza dei prodotti rispetto all'attributo indicato: 1 Del tutto Insufficiente. 2 Insufficiente. 3 Sufficiente. 4 Buono. 5 Ottimo.				
		Azienda X	Concorr.1	Concorr.2	Concorr.3	Concorr.4
Ampiezza di copertura del sistema di protezione						
Efficacia del sistema di diagnostica						
Precisione del controllo di velocità e di posizione						
Prestazioni dinamiche dell'azionamento						
Efficacia del controllo di coppia a basse velocità del motore						
Possibilità di accelerare/frenare il motore						
Possibilità di invertire la marcia						
Potenza dissipata						
Controllabilità dall'esterno dell'azionamento						
Autotuning						
Acquisizione di segnali di feedback						
Adattabilità e configurabilità del programma applicativo						
Disponibilità di SW applicativi						
Robustezza del modulo						
Accessibilità per la manutenzione						
Protezione ambientale (IP)						
Compattezza						
Insensibilità alle variazioni della tensione di rete						
Insensibilità alle variazioni della frequenza di rete						
Presenza di moduli accessori integrati						
Ampiezza del range di controllo velocità motore						
Ampiezza del range di Potenza						
Numerosità e distribuzione delle taglie nel range di potenza						
Ampiezza delle condizioni ambientali di utilizzo						
Silenziosità						

Figura 7. Questionario di valutazione dell'importanza dei requisiti del prodotto inverter vettoriale
Fonte: (Grandi, 2017)

N. della domanda	Bisogni del Cliente	Importanza	Azienda X	Concorrente 1	Concorrente 2	Concorrente 3	Concorrente 4	Media	Obiettivi	Rapp. di miglioram.	Sales Point	Peso Assoluto	Peso Relativo
1	Ampiezza di copertura del sistema di protezione	5,00	3,00	4,80	5,00	4,60	4,20	4,32	4,60	1,53	1,20	9,18	0,06
2	Efficacia del sistema di diagnostica	5,00	2,43	4,60	4,50	3,00	4,20	3,75	4,50	1,85	1,20	11,10	0,07
3	Precisione del controllo di velocità e di posizione	5,00	3,71	4,00	4,00	4,00	3,57	3,85	4,00	1,08	1,50	8,10	0,05
14	Robustezza del modulo	5,00	4,14	4,80	3,50	4,14	4,14	4,14	4,50	1,09	1,50	8,18	0,05
6	Possibilità di accelerare/frenare il motore	5,00	3,86	4,20	4,00	3,86	4,40	4,10	4,10	1,06	1,20	6,11	0,04
7	Possibilità di invertire la marcia	4,80	4,14	4,20	4,50	4,00	4,14	4,20	4,20	1,01	1,00	4,85	0,03
9	Controllabilità dall'esterno dell'azionamento	4,80	3,71	5,00	4,00	5,00	4,50	4,44	5,00	1,35	1,20	7,78	0,05
8	Potenza dissipata	4,60	3,43	3,00	3,00	3,00	4,60	3,41	3,43	1,00	1,50	6,90	0,05
4	Prestazioni dinamiche dell'azionamento	4,60	3,57	4,00	4,50	4,00	4,60	4,13	4,13	1,16	1,00	5,34	0,03
10	Autotuning	4,60	1,00	4,80	4,00	4,00	3,71	3,50	3,50	1,08	1,00	4,97	0,03
15	Accessibilità per la manutenzione	4,60	3,14	4,00	3,00	3,71	3,14	3,40	3,40	1,08	1,20	5,96	0,04
18	Insensibilità alle variazioni della tensione di rete	4,40	3,14	4,40	3,50	3,29	4,00	3,66	4,00	1,27	1,20	6,71	0,04
19	Insensibilità alle variazioni della frequenza di rete	4,40	3,29	4,40	3,50	4,50	4,40	4,02	4,00	1,21	1,00	5,32	0,03
11	Acquisizione di segnali di feedback	4,40	3,71	4,60	4,00	3,71	5,00	4,20	4,50	1,21	1,00	5,32	0,03
22	Ampiezza del range di potenza	4,20	3,57	4,40	5,00	4,00	2,33	3,86	4,00	1,12	1,00	4,70	0,03
21	Ampiezza del range di controllo velocità motore	4,20	3,71	4,60	4,50	4,40	4,40	4,32	4,40	1,18	1,00	4,96	0,03
12	Attendibilità e configurabilità del programma applicativi	4,20	3,86	4,00	4,00	4,29	4,40	4,11	5,00	1,29	1,00	5,42	0,04
20	Presenza di moduli accessori integrati	4,20	1,86	4,60	4,00	4,60	3,14	3,64	3,00	1,61	1,00	6,76	0,04
16	Protezione ambientale (IP)	4,20	3,14	4,20	3,50	5,00	4,20	4,01	4,00	1,27	1,00	5,33	0,03
24	Ampiezza delle condizioni ambientali di utilizzo	4,20	3,57	4,80	4,00	3,00	3,57	3,79	3,57	1,00	1,20	5,04	0,03
23	Numerosità e distribuzione delle taglie nel range di potenza	4,00	3,29	4,40	4,00	4,40	4,00	4,02	4,40	1,34	1,00	5,36	0,03
5	Efficacia del controllo di coppia a basse velocità del motore	3,80	3,57	4,00	4,00	4,60	4,14	4,06	4,50	1,26	1,20	5,75	0,04
17	Compattezza	3,80	4,00	3,20	3,00	3,00	4,80	3,60	4,00	1,00	1,20	4,56	0,03
25	Silenziosità	3,80	3,71	4,20	4,00	4,20	4,00	4,02	4,00	1,08	1,00	4,10	0,03
13	Disponibilità di SW applicativi	3,60	4,29	3,60	3,00	3,57	4,00	3,69	4,30	1,00	1,50	5,40	0,04
	Somma totale dei valori											153,19	1,00

Figura 8. Benchmarking competitivo del prodotto inverter vettoriale

Fonte: (Grandi, 2017)

Risposta Tecnica (SQC)	Target	Motivazioni
Protezione di "Alimentazione di potenza"	Due nuove Protezioni	Per allinearsi alla concorrenza bisogna aggiungere: <ul style="list-style-type: none"> ✓ Rilevamento del tempo di precarica della tensione di barra (inverter con alimentatore c.a./c.c. Integrato). ✓ Segnalazione di eventuali guasti o anomalie in questa fase.
Protezione di "Velocità"	Due nuove Protezioni	Per allinearsi alla concorrenza: <ul style="list-style-type: none"> ✓ Presenza di variazione eccessiva di $d(v)/d(t)$ del segnale di velocità proveniente dall'encoder, ossia la velocità aumenta (o diminuisce) in maniera troppo rapida. ✓ Errata connessione dei canali dell'encoder.
Altre Protezioni	Due nuove Protezioni	Per allinearsi alla concorrenza: <ul style="list-style-type: none"> ✓ Presenza di stallo nel motore (carico meccanico eccessivo, eccessiva richiesta di accelerazione). ✓ Motore scollegato durante la ripresa in mancanza della tensione di rete.
Protezione di "Corrente e sovratemperatura"	Tre nuove Protezioni	Per allinearsi alla concorrenza: <ul style="list-style-type: none"> ✓ Corrente eccessiva nella resistenza di frenatura. ✓ Sovraccarico della resistenza di frenatura. ✓ Sovratemperatura dei circuiti di controllo.
Protezione della "Scheda di Controllo"	Tre nuove Protezioni	Per allinearsi alla concorrenza: <ul style="list-style-type: none"> ✓ Guasto di schede opzionali. ✓ Intervento protezione esterna. ✓ Fallimento della fase di Autotuning.
Numero di ingressi analogici	Introdurne uno in più	Per allinearsi alla concorrenza ed espandere il numero di segnali elaborati o inviati dall'inverter.
Corrente Nominale	2,5+750 A	Come conseguenza dell'aumento del numero di taglie.
Potenza nominale / Numero di taglie	1,5+500 kVA su 19 taglie	Per avvicinarsi maggiormente alle taglie dei motori che normalmente si trovano sul mercato. Il cliente risparmia avendo a disposizione inverter con potenze più vicine a quelle dei motori che utilizza. Si può usare il Declassamento se i vantaggi, in termini di prezzo per il cliente, non sono significativi, viceversa si usa la Riprogettazione.
Potenza motore consigliata	1+405 kW	Per potenze maggiori ai 400 kW esiste già un'altra fam. di inverter.
Numero di uscite digitali ON/OFF	Introdurne 1 in più	Per allinearsi alla concorrenza ed espandere il numero di segnali elaborati o inviati dall'inverter..
Tastierino di program./segnalaz.	Introdurre come standard	Per allinearsi alle performance della concorrenza e perché è la Risposta Tecnica più importante per la soddisfazione dei clienti e in previsione di vendere l'inverter come prodotto sciolto.
Numero di uscite in frequenza	Introdurne una	Per superare la concorrenza ed espandere il numero di segnali elaborati o inviati dall'inverter.
Grado di protezione IP	IP 20	Per allinearsi alla concorrenza. Questo comporta un aumento della protezione dell'inverter dalla penetrazione da parte di corpi solidi.
Alimentatore c.a./c.c.	Su tutte le taglie	Per allinearsi alle performance della concorrenza e in previsione di vendere l'inverter come prodotto sciolto.
Peso e Ingombri	non stimabile	Esse non sono definibili a priori, ma si misureranno quando il prodotto sarà finito. In particolare, il peso non è molto importante visti i settori in cui l'azionamento opera.

Figura 10. Risposte tecniche ai requisiti di prodotto identificati dai bisogni dei clienti del prodotto inverter vettoriale

Fonte: (Grandi, 2017)

Come si può vedere dall'esempio, il metodo rappresenta una buona linea guida per identificare le caratteristiche tecniche più rilevanti del prodotto e comprendere quali migliorare in modo strategico ed economicamente conveniente, come si riporta in Figura 10.

L'obiettivo di questo documento è di proporre una metodologia che possa portare alla generazione dei requisiti in modo più rapido e veloce, evitando complicazioni procedurali che possono portare ad una troppo lenta e talvolta sbagliata definizione dei requisiti funzionali. L'approccio proposto utilizza informazioni provenienti da contenuti web che verranno poi processate da algoritmi di intelligenza artificiale, nello specifico il text-mining.

2. METODOLOGIA APPLICATA

2.1. PREPARAZIONE DELLA BASE DATI

2.1.1. *User Generated Content*

Per procedere operativamente con l'analisi dei requisiti si è ricorso all'utilizzo di UGC (User Generated Content) provenienti da diverse fonti web.

Definiamo UGC un contenuto accessibile liberamente nel web e generato da utenti che in modo volontario contribuiscono con dati, informazioni o media. (*Krumm et al., 2008*).

Gli UGC godono delle seguenti proprietà:

- Appartengono ad un dominio pubblico, dunque sono accessibili in modo libero da una qualsiasi piattaforma pubblica
- Sono veritieri in quanto provengono direttamente dalle esigenze delle diversità del web
- Garantiscono contenuti alternativi, dunque non strettamente connessi a tradizionali strutture di produzione di contenuti, come agenzie, redazioni o professionisti o qualunque altra struttura mossa da logiche di profitto

Secondo uno studio di *Olapic* (<https://www.olapic.com/resources/consumer-trust-usage-attitudes/>) realizzato su un campione di 4.500 utenti web compresi tra 16 e 49 anni e residenti negli Stati Uniti e nei principali paesi europei, viene accordata maggior fiducia ai contenuti pubblicati dagli utenti:

- Immagini UGC (52%);
- Video UGC (27%);
- UGC testuali (12%);
- Classici messaggi pubblicitari (5%)

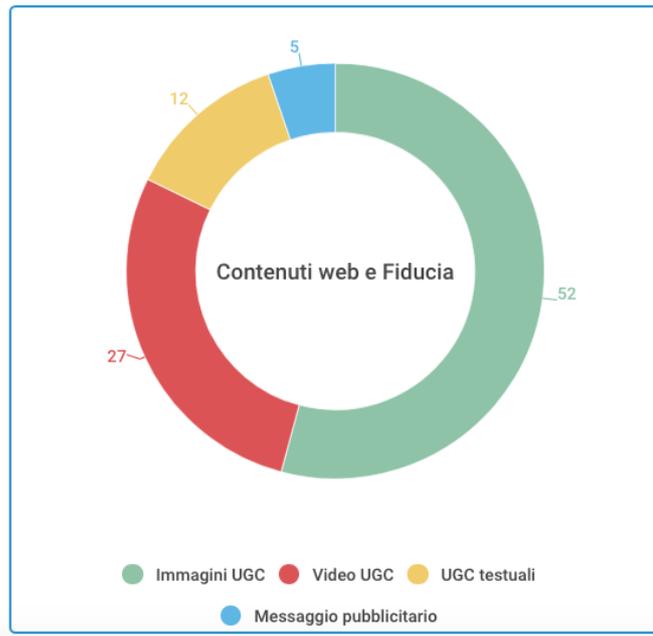


Figura 11. Contenuti web e fiducia
 Fonte: (<https://it.semrush.com/>)

Il 40% degli intervistati ha ammesso di aver prodotto contenuti “taggando” il loro brand preferito, nel 34% dei casi lo ha fatto perché ha giudicato positiva l’esperienza di brand. In tutto questo, solo il 14% degli intervistati ha detto di aver utilizzato hashtag proposti direttamente da un brand.

Il 76% ha affermato che gli User Generated Content sono “più onesti” della comunicazione tradizionale. Una percentuale variabile tra il 53 e il 70% (a seconda del paese di residenza) ritiene che gli User Generated Content ben generati possano aumentare la propria predisposizione all’acquisto del prodotto o servizio descritto.

Nel caso di definizione di requisiti funzionali, gli UGC contengono informazioni preziose e soprattutto di facile accesso. Non sempre gli utenti sono in grado di spiegare in dettaglio quali sono le modalità di uso desiderate per il prodotto nella loro attività quotidiana. Potrebbero anche avere un’immagine distorta di come si comportano nelle varie situazioni d’uso reali: normalmente un utente non ha interesse a conoscere in dettaglio la natura e la frequenza dei compiti che svolge quotidianamente, ma li vuole solamente eseguire nel modo più semplice possibile. Uno studio sul campo per apprendere come gli utenti si comportano nella realtà può quindi essere molto istruttivo ma risulta essere molto costoso, considerando anche la possibile varietà delle diverse tipologie di utenti.

Gli UGC in questo contesto facilitano di molto il processo di acquisizione di informazioni e

soprattutto permette di comprendere le reali esigenze del prodotto, dunque i suoi possibili punti di miglioramento (*Krumm et al., 2008*).

2.1.2. Scraping

Il problema di maneggiare grandi moli di informazioni risiede nella difficoltà di identificazione e di accesso ad esse (*Wyrwoll, 2014*). Lo scoglio di maggiore rilevanza dell'utilizzo degli UGC deriva dall'eterogeneità di fonti da cui possono essere generati; avendo a che fare con contenuti provenienti dal web, ciascuno di essi viene dunque referenziato da un sito web, dunque inserito in un corpo HTML differente di volta in volta. (*De Sirisuriya, 2017, Saurkar et al., 2018*)

Per accedere a tali contenuti web si adotta una tecnica di estrazione dati chiamata *Scraping*.

In generale, lo scraping dei dati Web può essere definito come il processo di estrazione e combinazione di contenuti di interesse dal Web in modo sistematico (*Marres & Weltevrede, 2013*). In tale processo, un software imita l'interazione di navigazione tra i server Web e l'utente in un normale attraversamento Web. Passo dopo passo, il bot accede a tutti i siti Web necessari, ne analizza i contenuti per trovare ed estrarre i dati di interesse e struttura tali contenuti come desiderato.

Solitamente i software per lo scraping implementano le API e i framework di Web scraping in modo da poter garantire agli utenti un software flessibile e con un'interfaccia grafica di semplice comprensione.

Il processo di scraping si articola nelle seguenti fasi:

1. *Accesso al sito*: il software di estrapolazione di dati Web stabilisce la comunicazione con il sito Web di destinazione tramite il protocollo HTTP, un protocollo Internet basato su testo senza stato che coordina le transazioni di richiesta-risposta tra un client, in genere un browser Web, e un server Web. In HTTP, i "metodi" di richiesta più frequenti sono GET, utilizzato nelle richieste di risorse, e POST, utilizzato per l'invio di moduli e il caricamento di file. Anche lo "User-Agent" è un'importante intestazione della richiesta, perché il server lo esamina per scoprire che tipo di programma sta accedendo ai suoi contenuti (utente presso un browser rispetto ai bot) e, infine, differenziare le risposte degli utenti.

2. *Analisi HTML ed estrazione dei contenuti*: una volta che il documento HTML è stato recuperato, il web data scraper può estrarre i contenuti di interesse. A tal fine, la corrispondenza delle espressioni regolari, da sola o in combinazione con una logica aggiuntiva, è ampiamente adottata. In alternativa, esistono librerie di analisi HTML (che lavorano sulla struttura DOM delle pagine Web) e linguaggi basati su selettori, come XPath (<http://www.w3.org/TR/xpath20/>) e la sintassi del selettore CSS. Come linea guida generale, si consiglia di mantenere le espressioni di corrispondenza il più generali possibile, per rendere i robot meno vulnerabili alle modifiche nel documento HTML.

3. *Costruzione dell'output*: l'obiettivo principale è trasformare i contenuti estratti in una rappresentazione strutturata adatta per ulteriori analisi e archiviazione. Sebbene questo passaggio finale sia marginale rispetto al Web scraping, alcuni strumenti sono a conoscenza della post-elaborazione dei risultati, fornendo strutture di dati in memoria e soluzioni basate su testo, come stringhe o file (in genere file XML o CSV). (Saurkar ET AL., 2018)

2.1.3. Data cleaning

I dati in output dal software di scraping solitamente devono essere formattati in modo da essere poi manipolabili per determinati scopi successivi.

Quando si combinano più origini dati, ci sono molte probabilità che i dati vengano duplicati o estratti in modo errato. Se i dati non sono corretti, i risultati e gli algoritmi non sono affidabili, dunque i modelli di analisi non potranno poi essere validati positivamente. Non esiste un modo assoluto per prescrivere i passaggi esatti nel processo di pulizia dei dati perché i processi variano da set di dati a set di dati, ma è fondamentale stabilire un modello per il processo di pulizia dei dati in modo da avere la consapevolezza della formattazione finale da raggiungere.

La pulizia dei dati è il processo che rimuove i dati che non possono appartenere al set di dati. La trasformazione dei dati è il processo di conversione dei dati da un formato o struttura in un altro, o meglio, da un modulo dati "grezzo" in un formato ad-hoc per l'analisi.

Con questa tecnica è possibile correggere e rimuovere dati errati, danneggiati, formattati in modo errato, duplicati o incompleti all'interno di un set di dati.

L'approccio che è stato adottato nell'analisi descritta al successivo capitolo si basa sull'identificazione di cinque caratteristiche di qualità del dato che sono state rispettate:

1. Validità, intesa come il grado di conformità dei dati alla tematica presa in considerazione
2. Precisione
3. Completezza
4. Consistenza
5. Uniformità

Nonostante non esista un approccio univoco per la pulizia dei dati, di seguito vengono generalizzati i passi seguiti:

1. Sono stati rimossi i dati indesiderati, dunque le osservazioni duplicate e le osservazioni irrilevanti. Tali errori di dati sono da imputare sia all'eterogeneità della provenienza del contenuto informativo sia a possibili errori nell'estrazione delle informazioni dal software di scraping.

Una seconda fase, più onerosa in termini di tempo, è stata un'attenta identificazione dei documenti che non rientrano nel problema specifico di analisi, come viene riportato nel seguente esempio; in questo modo la successiva fase di analisi è risultata più efficiente e coerente all'obiettivo di indagine principale, oltre a creare un set di dati più gestibile e più performante.

Esempio:

Commento1: Thank you

Commento2: Thank you very much for the help!

In questo caso i commenti dovrebbero essere entrambi eliminati in quanto non hanno nessun contenuto informativo ai fini dell'obiettivo di ricerca.

2. Sono stati identificati e rimossi gli errori strutturali intesi come convenzioni di denominazione strane, errori di battitura, simboli non pertinenti o a basso contenuto informativo, immagini, tabulazioni, spaziature in eccesso.... Questa fase risulta di fondamentale importanza qualora i dati debbano essere passati ad algoritmi di text mining, dunque basati proprio sull'analisi tra la struttura testuale e semantica del contenuto.

Esempio:

Commento1: The UR5 arm doesn't work, help pls!! 😞

Commento1_Corretto: The UR5 arm doesn't work, help!!

La tabulazione, l'emoticon e l'abbreviazione "pls" sono state rimosse in quanto creano distorsioni nell'utilizzo dell'algoritmo di text-mining non appartenendo al linguaggio NPL.

3. Sono stati adattati i dati in base al contenuto informativo richiesto. Ogni tipo di dato deve essere associabile univocamente ad un certo tipo di contenuto informativo (esempio: un UGC non può essere paragonabile alla data in cui esso è stato generato, devono essere differenziati). In questa fase sono state utilizzate strategie diverse, ad-hoc per ogni tipo di insieme di dati per distinguere al meglio ed in modo completo il contenuto informativo dei vari tipi di dati.

Il modo in cui rendere coerenti i dati è problem dependent, pertanto, deve essere valutato dall'analista in fase di data cleaning. Nell'attuale analisi, per esempio, la presenza di richiami testuali negli UGC (si pensi ad esempio in un forum alle risposte che riprendono il testo della domanda) avrebbe portato ad una distorsione nel campione dei dati e dunque alla sua successiva manipolazione; per tale ragione, sono stati generati dei codici finalizzati a fare dei confronti intra-stringa per eliminare i richiami testuali.

Nell'allegato 1 si riportano alcuni frammenti di codici utilizzati per completare tale fase.

Esempio:

Commento1: Nov' 2020 The cobot dresspack doesn't slither correctly in the arm, what can I do? Thank you, Chris

Commento2: Dec' 2020 @Chris said: The cobot dresspack doesn't slither correctly in the arm, what can I do? Thank you, Chris Answer: Try to call service assistance of the industry, factory defects are frequent!

Commento2_Corretto: Try to call service assistance of the industry, factory defects are frequent!

La ripresa della domanda nel Commento2 porta a deformazioni nell'applicazione dell'algoritmo di text-mining, deformandone l'analisi. È opportuno eliminare tali strutture come nel Commento2_Corretto. È poi doveroso estrarre la data all'interno

del commento in modo da distinguere le due tipologie di dato nell'analisi (tipo DATA e tipo STRING).

4. La gestione dei dati mancanti all'interno del set di dati è risultata fondamentale per determinare il livello di profondità dell'analisi successiva. I dati mancanti non possono essere ignorati perché in genere gli algoritmi non li accettano e in ogni caso porterebbero distorsioni.

Gli approcci che sono stati seguiti sono di seguito riportati:

- a. Dove il contenuto informativo del dato non è considerato totalmente rilevante per la successiva analisi o la quantità del tipo di dato disponibile non è sufficientemente elevata, si è optato per la rimozione del dato dal set di dati.

Esempio:

Si consideri la seguente tupla:

Commento: The cobot dresspack doesn't slither correctly in the arm, what can I do? Thank you, Chris

Anno: 2020

Mese:09

Giorno: -

Ora: -

La non disponibilità di giorno e ora renderà l'analisi sensibile al mese, per tale ragione è possibile omettere dall'analisi tali dati senza intaccare la validità dell'analisi.

- b. Dove il contenuto informativo del dato è considerato rilevante e la disponibilità di tale tipo di dato è sufficientemente elevata, si è optato per la generazione di tale dato in base alla distribuzione che meglio approssima tale set di dati.

Esempio:

Si consideri un campione di commenti estratti in ordine cronologico sequenziale:

Anno1: 2019

Anno2:2020

Anno3: -

Anno4:2020

La non disponibilità dell'anno 3 è facilmente prevedibile, in particolare sarà 2020. Qualora il campione di commenti non sia sequenziale occorre generare un numero che rispetti l'andamento della distribuzione empirica del campione:

2019 - 33%

2020 - 66%

Estraendo in modo casuale il numero da tale distribuzione empirica è possibile completare l'anno mancante.

- c. Cambiare tipologia di estrazione dei dati qualora il contenuto informativo risultasse non sufficientemente attendibile.

Nei primi due casi l'approccio non è ottimale in quanto va ad inficiare nel primo caso sulla completezza del set di dati, mentre nel secondo caso può portare a distorsioni. La terza opzione risulta la più corretta ma oltre ad essere molto onerose in termini di tempo e risorse utilizzate può non portare ad un miglioramento del contenuto informativo dei dati.

5. Alla fine del processo è stata fatta una revisione dei dati e ove vi fosse un esito positivo, sono stati uniformati in un'unica base dati eliminando l'eterogeneità iniziale.

	A	B	C	D	E	F
1						
2		i..Link	Commento	Giorno	Mese	Anno
3	1	https://github.com/Unpinning	as Kinetic has been s	2	3	2018
4	2	https://github.com/Both	base >tool0 and base_link >	17	12	2018
5	3	https://github.com/Is	this still an issue @maxime pe	29	3	2019
6	4	https://www.github.com/Have	the same issue. Did you si	16	4	2019
7	5	https://www.github.com/can	you ping controller from you	16	6	2018
8	6	https://www.github.com/this	is the mfp screen first powe	6	4	2021
9	7	https://www.github.com/ok	i check the pin cable and back	5	4	2021
10	8	https://www.github.com/750	mm/sec in non collaborativ	7	4	2017
11	9	https://www.github.com/Thank	you! This looks very prom	23	3	2017
12	10	https://www.github.com/I	found my issue. Thanks again J	18	9	2017
13	11	https://www.github.com/Ah	brilliant. Thank you for the li	24	9	2016
14	12	https://www.github.com/Thanks	this is what I wanted to h	20	9	2016
15	13	https://www.github.com/Hi	there, are there any news on	8	4	2020
16	14	https://www.github.com/You	can write me to dalvar_5@y	8	12	2020
17	15	https://www.github.com/Hi	, is there any news about this	23	6	2020
18	16	https://www.github.com/If	i can fix it, I will let you know!	28	1	2020
19	17	https://www.github.com/Yes	, that s the correct way to do	20	4	2020
20	18	https://www.github.com/Thanks	. Thats very helpful and v	15	8	2020
21	19	https://www.github.com/maybe	you could use: ² or	18	8	2020
22	20	https://www.github.com/Thank	you Sam for your kind inte	19	9	2020
22	21	https://www.github.com/We	are interested by this featur	22	7	2020

Figura 12. Esempio di dataset di analisi con i dati formattati e uniformati riferito al caso applicativo sulla robotica collaborativa

2.2. ANALISI E MODELLAZIONE DEI DATI

Per l'analisi è stato utilizzato un algoritmo di text mining basato su complesse strutture matematico – informatiche dell'intelligenza artificiale.

Il text mining viene definito come il processo di analisi di grandi quantità di documenti per mettere in evidenza nuove informazioni o aiutare a rispondere a domande di ricerca specifiche. (Gupta & Lehal, 2009)

Una volta estratte, queste informazioni vengono convertite in una forma strutturata che può essere poi analizzata in modo più approfondito.

L'estrazione del testo impiega una varietà di metodologie per elaborare il testo, una delle più importanti è il *Natural Language Processing (NLP)*. (Bird et al., 2009)

La comprensione del linguaggio naturale aiuta le macchine a "leggere" il testo (o un altro input come il parlato) simulando la capacità umana di comprendere una lingua naturale come l'italiano, l'inglese, lo spagnolo, il cinese.... L'elaborazione del linguaggio naturale include sia la comprensione del linguaggio naturale che la generazione del linguaggio naturale, che simula la capacità umana di creare testo in linguaggio naturale, ad esempio per riassumere informazioni o prendere parte a un dialogo.

In questo contesto, si è verificato che negli odierni sistemi di elaborazione del linguaggio naturale possono analizzare quantità illimitate di dati basati su testo senza fatica e in modo coerente e imparziale; possono comprendere concetti all'interno di contesti complessi e decifrare ambiguità del linguaggio per estrarre fatti e relazioni chiave o fornire riepiloghi (*Hearst, 2003*).

Per far ciò, gli algoritmi devono basarsi su quello che è l'apprendimento automatico.

L'apprendimento automatico è una tecnologia di intelligenza artificiale (AI) che fornisce ai sistemi la capacità di apprendere automaticamente dall'esperienza senza la necessità di una programmazione esplicita e può aiutare a risolvere problemi complessi con precisione che può rivaleggiare o addirittura superare a volte gli umani.

La specificazione di un'ontologia include un vocabolario di termini e vincoli formali al suo utilizzo. L'elaborazione del linguaggio naturale pronto per le aziende richiede una gamma di vocabolari, ontologie e strategie correlate per identificare i concetti nel loro contesto corretto:

- Thesauri, vocabolari, tassonomie e ontologie per concetti con termini noti;
- Approcci basati su modelli per categorie quali misurazioni, mutazioni e nomi chimici che possono includere termini nuovi (invisibili);
- Identificazione, annotazione e trasformazione di concetti specifici del dominio e basati su regole;
- Integrazione dei vocabolari dei clienti per consentire annotazioni personalizzate;
- Ricerca avanzata per consentire l'identificazione di intervalli di dati per date, valori numerici, area, concentrazione, percentuale, durata, lunghezza e peso.

Sfruttando la tecnologia precedentemente descritta, si è utilizzato un approccio metodologico per determinare una tabella di requisiti funzionali generati dall'analisi di contenuto.

Di seguito viene spiegato nel dettaglio la procedura adottata.

1. *Lettura dei dati*

L'algoritmo prende in ingresso un file contenente un foglio di calcolo contenente i covariati del testo in cui esiste una formattazione dei dati di testo univoca, ottenuta dalla precedente fase di data cleaning. In Figura 12 viene riportato un estratto del database uniformato.

2. *Pre-elaborazione*

Per il corretto funzionamento dell'algoritmo in R, è utile adottare alcuni tipi di elaborazione dei dati prima di modellarli. Le tecniche che sono state utilizzate sono lo *stemming* (dunque la riduzione delle parole alla loro radice, estraendone le desinenze), l'eliminazione della punteggiatura e la rimozione delle parole di arresto.

Esempio:

Nei commenti possono essere presenti le seguenti parole: working, work, works.

Attraverso lo stemming vengono ripulite e trattate come un'unica parola: work.

3. *Preparazione del dataset di analisi*

Dopo la lettura dei dati, viene fatta un'elaborazione che permette di valutare il corretto caricamento del dataset preparato, l'avvenuta lettura del dataset da parte dell'algoritmo e la rimozione delle parole a bassa frequenza, considerate irrilevanti data una certa soglia fissata ragionevolmente dopo la valutazione dei risultati di diversi scenari iterativi analizzati. In questa fase vengono indicizzate tutte le relazioni tra i metadati e i documenti inseriti, con la possibilità che alcuni di questi vengano cancellati in quanto ritenuti non sufficientemente a valore aggiunto dall'algoritmo, per esempio perché contenevano parole rare.

Esempio:

Commento: While I was working with my UR5, my dog started barking,so crazy thing!

Commento_Ripulito: While I was work with UR5 start thing

Il commento ripulito, seppur di difficile comprensione è depurato delle parole non pertinenti all'analisi; in questo caso, l'algoritmo pone il suo focus sull'utilizzo dell'UR5 da parte dell'utente.

4. *Stima del numero di Topic ottimale*

Come introdotto nel precedente paragrafo, i metadati possono essere inseriti nel modello di argomento in due modi: prevalenza Topica e contenuto Topico. I covariati di metadati per la prevalenza Topica consentono ai metadati osservati di influire sulla frequenza con cui viene discusso un argomento. I covariati nel contenuto Topico consentono ai metadati osservati di influenzare l'uso della frequenza delle parole all'interno di un determinato argomento, cioè il modo in cui viene discusso un particolare argomento. Il modello è stato impostato per l'esecuzione di massimo 75 iterazioni. La convergenza è monitorata dalla variazione del limite inferiore variazionale approssimativo. Una volta che il limite ha una modifica abbastanza piccola tra le iterazioni, il modello viene considerato convergente.

La scelta del numero ottimale di Topic che approssima al meglio il contenuto informativo del dataset iniziale deve essere poi valutato in base a determinati criteri.

In Figura 13 si riportano alcuni indicatori estratti dal caso studio sulla robotica collaborativa. Per l'analisi sono stati considerati i seguenti criteri, riportati in ordine di importanza attribuita per il trade off decisionale:

1. Stimatore di massima verosimiglianza (Maximum Held-out likelihood)

Si basa sul confronto delle specifiche del modello per vedere quanto bene ogni modello riesce a prevedere le parole all'interno del documento. In questo caso, viene definito come la stima della probabilità che le parole che appaiono all'interno di un documento vengano predette quando tali parole sono state rimosse dal documento nella fase di stima. La sua definizione è nota per essere molto impegnativa, in quanto equivale a stimare una probabilità marginale nella selezione del modello Bayesiano, per questo ne viene

fornita un'approssimazione valida e veloce da calcolare (*Scott Baldridge, 2013*).

II. *Coerenza semantica*

La coerenza semantica è massimizzata quando le parole più probabili in un dato argomento spesso ricorrono insieme, ed è una metrica che si correla bene con il giudizio umano sulla qualità dell'argomento. Avere un'elevata coerenza semantica è relativamente facile, tuttavia, se si hanno solo pochi argomenti dominati da parole molto comuni, il risultato potrebbe essere notevolmente distorto.

III. *Esclusività*

Misura il grado in cui le parole principali di un argomento non vengono visualizzate come parole principali in altri argomenti, ovvero la misura in cui le parole principali sono "esclusive". Il valore è la media, su ogni parola, della probabilità di trovare quella parola nell'argomento divisa per la somma delle probabilità di trovare quella parola in tutti gli argomenti:

$$x_k = \frac{\sum_i p_{ij}}{n * p_i} \quad (2.1)$$

Dove k rappresenta il numero di iterazioni comprendenti un set di parole principali di un generico argomento j composto da n parole e i rappresenta la generica parola in considerazione. L'esclusività è correlata negativamente al conteggio delle iterazioni e spesso indica anche argomenti più vaghi e generali.

5. *Topic modelling*

L'obiettivo del Topic modeling è quello di mettere in risalto tematiche di discussione inerenti ad uno o più argomenti e stimare la loro relazione con i metadati dei documenti. Gli output del modello possono essere utilizzati per condurre test di ipotesi su queste

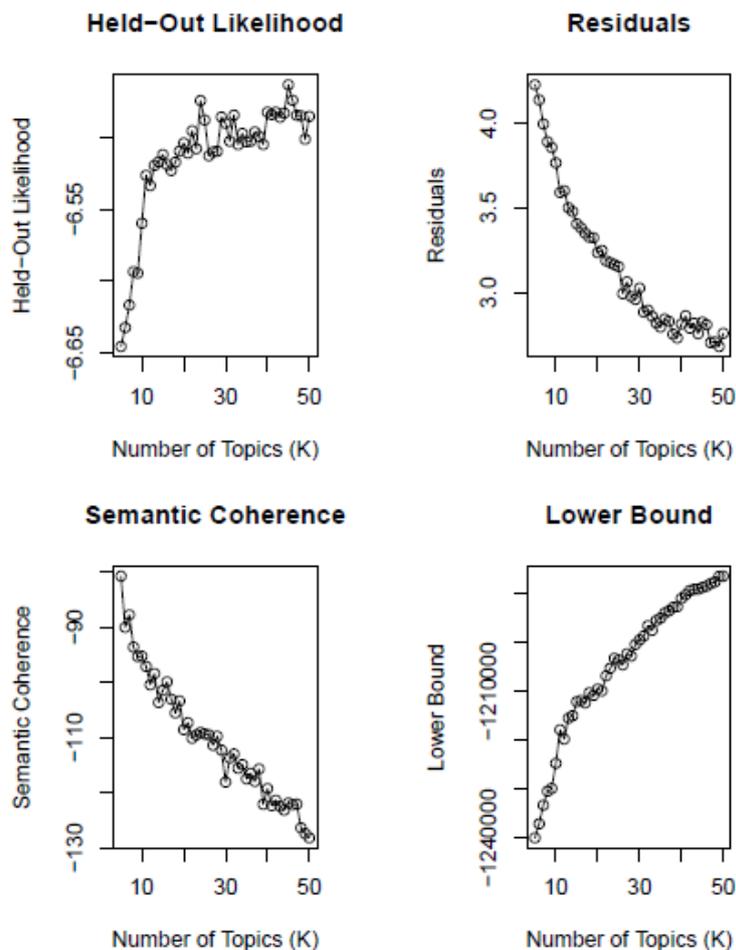


Figura 13. Esempio dell'andamento degli indicatori al variare del numero di Topic riferito al caso applicativo sulla robotica collaborativa

relazioni che a sua volta potranno generare delle relazioni complesse e difficilmente visibili dopo uno studio approfondito da un soggetto umano.

Per affrontare questo problema in modo algoritmico e computazionale si è fatto riferimento al linguaggio R utilizzando l'interfaccia grafica, di facile comprensione, proposta dalla sua estensione Rstudio (<https://www.rstudio.com/>).

La ragione di tale scelta è dettata dal fatto che R è un software e un linguaggio di programmazione gratuito, open source e soprattutto adatto ad elaborare dati statistici complessi.

L'algoritmo in R utilizza il pacchetto STM che fornisce anche strumenti per facilitare il flusso di lavoro associato all'analisi dei dati testuali.

L'algoritmo STM è un modello generativo di conteggio delle parole, dunque, definisce un processo di generazione dei dati per ogni documento preso in input e utilizza tali dati per trovare i valori più probabili per i parametri all'interno del modello. (Roberts et al., 2019)

Il modello inizia a definire distribuzioni di documenti-argomenti e argomenti-parole che generano documenti a cui sono associati metadati, come riportato in Tabella 2 e Tabella 3. In questa sequenza di relazioni, un argomento è definito come un agglomerato di parole in cui ognuna ha una probabilità di appartenere a un dato argomento; un documento, invece, è un insieme di argomenti, il che significa che un singolo documento può essere composto da più argomenti. Di conseguenza, essendo ogni relazione rappresentata in forma percentuale, la somma delle proporzioni dell'argomento in tutti gli argomenti di un documento è pari a 1 e la somma delle probabilità di parola per un determinato argomento è pari a 1.

Documento	Covariati del Topic
@lakshmip001 As David mentioned, if you have a 2D CAD of the part, you can print it in a 1:1 scale and use it to teach the part to the vision system. Other than reflections created by the shiny part, this part should work fairly well with the Robotiq Wrist Camera. I would try using a 2D drawing of the part to teach the part to the vision system. Try printing the part on a colored piece of paper (Yellow would work well for this), this way you will have a high contrast between the black printed lines and the background. Once you have taken the 5 pictures, remove the piece of paper and put the real part in the camera's field of view for the scanning process.	0,68523
@Student I agree with @LoÃ~c if you aren't able to chamfer the edge of the gauge you will either need to increase the precision you are using to find the hole or you will need to use an FT sensor. I can say we are using a camera to find a 4 mm hole in a part and then place that hole onto a peg that's about 3.8 mm square. Â We do not use any type of search method to do this either. Â To achieve the precision that we have the hole taking up about 70% of the FoV of	0,67632

the camera, it's literally placed within 75 mm of the part where the hole is.	
@Etienne_Samson I want to locate a M3 hole on aluminium parts, the camera should locate the hole pick up a screw and put it in the hole.	0,66045
matthewd92 So, It means I can use Asrily feeder with the Robotiq Wrist Camera. But Asrily feeder vibration can only be obtained when object is not detected and the other features of Camera cannot be controlled. And as the Robotiq camera takes minimum 10 20 secs to detect an object process is slow. And Robotiq camera can detect only one object per cycle. As shown in the video can Robot pick randomly placed screws on the flat surface?? So, UR camera can work this feeder. Am I right with the conclusion. Thanking you in advance.	0,64885

Tabella 2. Esempio di quattro documenti con i relativi covariati del Topic "Wrist Camera" riferito al caso applicativo sulla robotica collaborativa

Topic	Keywords	Prevalenza Topica
Wrist Camera	<p><u>Highest Prob</u>: work, tri, problem, now, seem, just, still, error, chang, solv</p> <p><u>FREX</u>: problem, tri, seem, solv, work, fine, got, figur, now, doesnt</p> <p><u>Lift</u>: broke, fine, solv, problem, helloim, stuck, mistak, seem, tri, appar</p> <p><u>Score</u>: work, tri, problem, broke, now, solv, seem, error, fine, still</p>	0,02616

Tabella 3. Esempio del processo di etichettatura riferito al caso applicativo sulla robotica collaborativa del Topic "Wrist Camera" con associate le keywords descrittive e la prevalenza Topica

Per determinare tali relazioni strutturali il modello utilizza i covariati calcolati ad ogni iterazione, come riportato in Figura 14.

Nelle statistiche, un covariato è definito come una variabile che è predittiva in modo probabilistico del risultato esaminato. Un covariato può essere di interesse diretto o può essere una variabile confondente o interagente per la definizione delle relazioni strutturali. L'algoritmo permette di utilizzare covariati di prevalenza Topica, covariati di contenuto Topico, entrambi o nessuno dei due.

La prevalenza Topica si riferisce alla quantità di un documento associata a un argomento mentre il contenuto Topico si riferisce alle parole utilizzate all'interno di un argomento. Quindi i metadati che spiegano la prevalenza Topica sono indicati come covariati di prevalenza Topica e le variabili che spiegano il contenuto Topico sono indicate come covariati di contenuto Topico.

Il modello prende in input un dataset di documenti e permette di preparare il contenuto informativo e successivamente di valutare le relazioni, comprendere le strutture logiche tra i vari dati e visualizzarne i risultati, come riportato in Tabella 2 e Tabella 3.

	A	B	C	D	E	F	G	H	I	J	K
3		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
4	16509	0.0037	0.0041	0.00271	0.00412	0.00622	0.00103	0.02038	0.01615	0.00424	0.0029
5	15444	0.00853	0.00537	0.00634	0.01069	0.01037	0.00333	0.04446	0.12664	0.00884	0.0132
6	12448	0.02184	0.00613	0.01473	0.01135	0.04407	0.00198	0.02835	0.04275	0.01058	0.0080
7	15231	0.02443	0.02187	0.0061	0.01396	0.01885	0.00394	0.08533	0.0321	0.02718	0.0063
8	13297	0.01534	0.01533	0.00737	0.01862	0.01459	0.003	0.05108	0.03801	0.01547	0.0045
9	14450	0.00664	0.02302	0.00272	0.0046	0.00827	0.00268	0.05558	0.02398	0.0048	0.0027
10	4702	0.01123	0.01458	0.01161	0.02851	0.01835	0.00496	0.05764	0.02343	0.03577	0.0087
11	13577	0.0078	0.01004	0.01133	0.04148	0.05322	0.00323	0.02701	0.07375	0.00943	0.0060
12	15614	0.0219	0.00495	0.00852	0.00705	0.02723	0.00459	0.06878	0.03116	0.00439	0.0146
13	8515	0.01043	0.01313	0.00503	0.00975	0.01571	0.00455	0.03812	0.03777	0.00811	0.0064
14	10977	0.01172	0.01163	0.01093	0.01154	0.01619	0.00793	0.22031	0.02621	0.01255	0.0077
15	16470	0.01007	0.00139	0.10244	0.00403	0.19831	0.0017	0.00966	0.06608	0.00245	0.0211
16	13767	0.01089	0.02579	0.00563	0.01074	0.01595	0.00689	0.1631	0.02195	0.0112	0.0058
17	8361	0.00926	0.01697	0.00901	0.01939	0.01841	0.00578	0.05902	0.03621	0.01469	0.0070
18	15315	0.06576	0.0055	0.004	0.01238	0.00532	0.00291	0.01385	0.03985	0.0054	0.0080
19	2960	0.01633	0.0128	0.01085	0.03315	0.01768	0.00604	0.05189	0.0378	0.02522	0.0183
20	14128	0.00914	0.01046	0.02717	0.01315	0.02445	0.00619	0.12763	0.06389	0.00845	0.0076
21	3232	0.01323	0.01291	0.01901	0.07335	0.02941	0.00829	0.05563	0.03286	0.02676	0.0133

Figura 14. Esempio del calcolo dei covariati effettuato dall'algoritmo di text-mining riferito al caso applicativo sulla robotica collaborativa

6. Labelling

L'algoritmo STM permette di restituire in output le relazioni tra documenti e Topic, associando a questi ultimi un set di parole più rilevanti per descrivere tale argomento, come riportato in Figura 15 e in Tabella 3. Le parole per ogni Topic sono classificate in quattro categorie come di seguito riportato:

1. *Highest probability*: parole generate in base alla probabilità più alta di essere generate per un argomento
2. *FREX*: parole generate in base alla frequenza e all'esclusività
3. *Lift*: parole generate da un punteggio calcolato dividendo la distribuzione Topic-parola per la distribuzione empirica della probabilità del conteggio delle parole
4. *Score*: parole con il punteggio più alto in base alla classificazione LDA

(Roberts et al., 2019)

L'etichettatura dei Topic è un processo che attualmente richiede un contributo umano *(Matrogiacomo et al., 2021)* in quanto necessita della piena comprensione dell'argomento in questione per definire un'etichetta il più pertinente possibile. *(Mei et al., 2007)*

Per l'etichettatura dei Topic è stata seguita una metodologia volta a massimizzare la comprensione dell'argomento in riferimento al dataset iniziale. Per ogni Topic sono stati ordinati in modo decrescente i valori delle probabilità di associazione a tale Topic (covariati) di ogni documento del database iniziale, sono stati selezionati un numero ragionevole di documenti più pertinenti per tale Topic e ciascuno di essi è stato compreso al fine di individuare quale fosse l'argomento principale che li collegasse tra di loro.

Esempio:

Si supponga di avere un set di 20.000 documenti e di volerne leggere cinque per ogni Topic in modo da comprendere direttamente di cosa trattano.

Di seguito si riporta un generico insieme di documenti con i relativi covariati riferiti ad un generico Topic V2 (come da Figura 15).

Output algoritmo di text-mining:

Documento1 – 0,8725

Documento2 – 0,0097

Documento3 – 0,0002

Documento4 – 0,0969

Documento5 – 0,7524

...

In questo ordine, i documenti non rappresentano nel miglior modo possibile il Topic V2 in quanto non tutti dimostrano dei covariati sufficientemente elevati: ad esempio, il Documento2, il Documento3 e il Documento4 non risultano significativi per il contenuto del Topic V2.

Creando un ordine decrescente dei covariati, sul set dei 20.000 documenti, è possibile, invece, ottenere i documenti più pertinenti a descrivere il Topic V2:

Ordinamento compiuto:

Documento 1997 – 0,9997

Documento 1 – 0,8725

Documento 18466 – 0,8235

Documento 8741 – 0,8124

Documento5 – 0,7524

...

Leggendo questo campione si è certi di analizzare i cinque documenti con il contenuto più rappresentativo del Topic V2, minimizzando la possibilità di compiere errori in fase di labelling.

Infine, l'etichetta è stata confrontata con le parole associate ad ogni Topic al fine di poterla poi validare. In Tabella 3 viene riportato un esempio del risultato dell'etichettatura riferita al Topic "Wrist Camera".

	A	B	C	D	E	F	G	H	I
4		FREX: insid, back, noth, anoth, sometim, pull, replac, rememb, bracket, put							
5		Lift: feb, forth, bracket, anoth, insid, noth, back, kawasaki, pull, timeout							
6		Score: feb, back, anoth, insid, null, noth, replac, put, timeout, mechan							
7	V2								
8		Highest Prob: code, python, java, urscript, daemon, sampl, function, use, process, ex							
9		FREX: urscript, daemon, sampl, java, python, languag, xmlrpc, rpc, librari, code							
10		Lift: invest, rpc, xmlrpc, deamon, urscript, languag, daemon, invok, sampl, xml							
11		Score: daemon, python, java, urscript, invest, sampl, xmlrpc, rpc, xml, code							
12	V3								
13		Highest Prob: forc, tool, use, tcp, base, axi, frame, sensor, rotat, plane							
14		FREX: forc, frame, tool, rotat, torqu, axi, vector, transform, coordin, measur							
15		Lift: euler, marcantoinelacass, vector, poseadd, transform, wrt, forcemod, rxryrz, pos							
16		Score: forc, tool, frame, axi, rotat, plane, tcp, sensor, coordin, marcantoinelacass							
17	V4								
18		Highest Prob: urcap, featur, use, polyscop, function, add, user, implement, current, r							
19		FREX: featur, request, polyscop, add, urcap, develop, implement, api, releas, jbm							
20		Lift: benefici, mbirkholz, elabor, request, sam, mbush, jbm, news, sko, featur							
21		Score: urcap, featur, benefici, polyscop, api, user, request, implement, sdk, develop							

Figura 15. Esempio dell'output dell'algoritmo di Topic modeling riferito al caso applicativo sulla robotica collaborativa

7. Verifica dei risultati

La verifica dei risultati è una fase compiuta prima o al massimo contemporaneamente all'etichettatura. Emergono spesso a questo punto errori che danno via ad una lunga serie di iterazioni volte a migliorare l'output dell'algoritmo STM.

In questa fase occorre verificare le parole in output per ogni Topic, verificarne l'utilità al fine dell'obiettivo d'indagine e dove possibile, vedere quanto sono coerenti al Topic associato e quanto valore aggiungono possono portare per una migliore descrizione.

Esempio:

Parole come thing, thank, you, think, etc... sono irrilevanti per l'analisi e devono essere rimosse per rendere l'analisi più precisa possibile. In questo modo si lascia la possibilità all'algoritmo di considerare parole pertinenti all'indagine.

Vengono, quindi, adottate due strategie:

- i. Apportare manualmente modifiche al dataset iniziale in modo da correggere e ripulire ulteriormente i dati testuali da valori considerati ridondanti o superflui

- ii. Generare un dataset di parole che verranno inserite come input all'algoritmo in R al fine di poterle rimuovere in modo autonomo durante l'esecuzione del codice

Entrambi gli approcci conducono al risultato voluto; nel primo caso vengono richiesti impieghi di tempo onerosi in quanto ciascuna parola deve essere filtrata in base alle sue desinenze, automatizzabile con la procedura di stemming a livello computazionale. Il primo approccio risulta comunque l'unico possibile qualora in fase esecutiva del codice appaiano delle eccezioni generate dal contenuto informativo della base dati iniziale che possono portare a bug del codice non facilmente modificabili.

8. Validazione dei risultati

La validazione dei risultati è stata effettuata estraendo un campione casuale di commenti per ogni Topic in modo da poterne vedere in che modo il contenuto informativo dei documenti iniziale fosse ben rappresentabile dalle etichette. Per questa fase sono stati scelti campioni casuali di commenti con una percentuale di pertinenza col Topic relativamente alta, dunque documenti che potessero associarsi in modo sufficientemente pertinente ad un dato Topic (es. un documento con basse probabilità di generare ogni Topic non è stato considerato una metrica attendibile di validazione del modello).

L'indicatore utilizzato per tale fase è l'*accuratezza* (a) definita come il rapporto tra osservazioni correttamente previste (c) e osservazioni totali (t) (Mastrogiacomo, 2021):

$$a = \frac{\sum_i c}{\sum_i t} \quad (2.2)$$

Nella Tabella 4, viene riportato un campione di commenti estratti in modo casuale riferiti al Topic "Controllo Velocità" (riportato nella Tabella 5). Ogni commento è stato analizzato e ad ognuno è stato associato un nome di un Topic, tra quelli elencati nella Tabella 5, che lo rappresentasse al meglio. Qualora il nome del Topic ritenuto più significativo fosse "Controllo Velocità" allora il commento può essere ritenuto validato. In questo modo, come riportato in Tabella 4, il numero di osservazioni correttamente previste (c) è di 6 mentre le osservazioni totali (t) sono 8; applicando la Formula 2.2 si trova:

$$a = \frac{6}{8} = 0,75$$

Stabilendo una soglia minima di accuratezza è possibile validare o respingere il labelling del Topic di riferimento; in questo caso il valore di 75% è considerato accettabile, dunque, l'etichetta può essere validata e confermata.

Commento estratto casualmente	Validazione del commento
Setting the speed in RoboDK allows you to specify joint speed and linear speed: https://robodk.com/doc/en/Robot Programs.html#InsSpeed Joint speed & acceleration is used for joint movements and linear speed & acceleration is used for linear movements.	Sì
Hi, I think the velocity profile follows the equation as the following to achieve the triangle velocity profile. $V^2 = a * S$ (V is the velocity, S is the moving distance) Whether the profile is triangle or trapezoid depends on the S is longer enough to speed up to the desired constant velocity with the given acceleration.	Sì
If you use break when, only the current motion will be cancelled, but there may be multiple motions have been queued and these queued motions will be executed. You can cancel all the motions you have sent (current motion and the queued motions) by <code>CodeIbr.getController().getExecutionService().cancelAll()</code>	No
Hi @guido.sacson, to reconstruct a movej, you need to slice your path in time windows, calculated with the desired joint speed, e.g. 60 deg/s on the fastest joint. as you generate the path with a fixed frequency other than using the robot controller clock (2ms on e Series) i would suggest to keep the lookahead_time high to smoothen the path. The higher your frequency, the closer the servoj path gets to the movej path, but without using the controller frequency it would never be the same	Sì
Thanks for the question. It might be counter intuitive to use speedJ for linear moves because the nature of the move is very different and more simple. For linear moves the robot makes more calculations and typical more joints need to turn to make such linear move. And I have not seen an example of using SpeedJ for linear moves. In any motion it is important to handle acceleration and deceleration between the cruising speed because this will also take time also when there is a turn. To make smooth moves it is important to handle the timing for the motion by controlling the speed.	Sì
Hi Gavin Yes when providing the command <code>movej([0, 0, 0, 0, 0.0, 0.0], a=0.2, v=0.2)</code> my robot does move to joint positions 0,0,0,0,0 which is the same position expressed as Cartesian position X= 457.02mm, Y= 194.01mm, Z=2.52mm, Rx=91.16 deg (1.5910 Rad), Ry=0.01 deg (0.0062 Rad), Rz=0.43 deg (0.0056 Rad). Notice the difference in Joint positions and Cartesian position (Pose) way expressing a position. It is not possible to put the robot into Pose 0,0,0,0,0 because that will mean the robot move into itself whereas it is possible to put the robot into joint position 0,0,0,0,0 which is with the arm at a horizontal stretch where the wrist 3 is pointing downwards. To move the robot in Cartesian coordinates there needs to be a p in front of the list like this <code>movej(p[X, Y, Z, Rx, Ry, Rz])</code> , but again the robot cannot move to X=0, Y=0, Z=0 because that's inside itself.	No

It is vibrating while moving, with the speed of 100%, and Joint speed is 80 mm/s, and Acceleration 100 mm/s ²	Sì
Hi, I need to move UR5 at constant tool speed and on a straight line. Tool speed is constant when I use movep. But I dont if I can achieve linear motion with movep .Can you please help me if I can use some commands to move robot arm with a tool constant speed and linear displacement?	Sì

Tabella 4. Esempio della procedura di validazione del labelling riferito al caso applicativo sulla robotica collaborativa del Topic “Controllo velocità”

La validazione del labelling può essere perfezionata considerando la possibilità di compiere errori in fase di valutazione del campione di documenti estratto.

Si consideri:

- α = rischio di classificare un commento valido come non valido (falso difettoso)
- β = rischio di classificare un commento non valido come valido (falso buono)
- n = numerosità del campione estratto

La percentuale apparente di commenti non validati può essere calcolata richiamando la Formula 2.2, applicando la seguente formula:

$$p_{APP} = \frac{\sum_i t - \sum_i c}{\sum_i t} = 1 - a \quad (2.3)$$

Il numero di falsi difettosi può essere così calcolato:

$$Falsi\ difettosi = n * (1 - p_{REA}) * \alpha \quad (2.4)$$

Analogamente, il numero di falsi buoni è pari a:

$$Falsi\ buoni = n * p_{REA} * \beta \quad (2.5)$$

Il numero corretto di unità difettose è quindi pari a:

$$Difettosi\ reali = difettosi\ apparenti - n * (1 - p_{REA}) * \alpha + n * p_{REA} * \beta \quad (2.6)$$

Da cui, dividendo per la numerosità del campione, è possibile ricavare la seguente formula:

$$p_{REA} = \frac{p_{APP} - \alpha}{1 - \alpha - \beta} \quad (2.7)$$

(Franceschini et al., 2020)

Con riferimento alla Tabella 4, è quindi possibile correggere l'accuratezza del 75% trovata precedentemente assumendo un rischio di prima specie (α) del 1% e un rischio di seconda specie (β) del 5%.

$$a_{REA} = 1 - \frac{0,25 - 0,01}{1 - 0,01 - 0,05} = 0,74$$

Considerando la possibilità che in fase di valutazione possano essere fatti degli errori, si può notare che l'accuratezza diminuisce di un punto percentuale.

3. METODOLOGIA DI ANALISI

3.1. INTRODUZIONE AL PROBLEMA

L'obiettivo dell'analisi è quello di generare una scheda di requisiti funzionali del prodotto COBOT che permetta di fungere da strumento decisionale per miglioramenti strategici del prodotto, volti a migliorarlo e/o ottenere poi un vantaggio competitivo tecnico.

La tecnica operativa come precedentemente spiegato è quella del text-mining.

Per definire i requisiti funzionali si è partiti da una base dati riferita a commenti generati da utenti su forum di vario genere relativi alla tematica della robotica collaborativa.

Con la tecnica precedentemente descritta si è identificato un campione di possibili siti web, in questo caso forum, che potessero ben rappresentare l'obiettivo della ricerca. In questa fase si è fatta molta attenzione a trovare documentazione appartenete alla categoria degli UGC, dunque un contenuto libero e disinteressato.

Con tale insieme di fonti di dati è stato fatto scraping e data cleaning per ottenere una base dati ricca di contenuto e coerente con la tematica trattata. Per raggiungere tale obiettivo sono stati generati degli algoritmi di scraping ad hoc per ogni fonte e funzioni di pulizia per uniformare e formattare i dati in modo corretto e uniforme.

Il risultato finale è un database di partenza di circa 20.000 commenti ai quali è stata associata la relativa data di pubblicazione.

La base dati è stata poi manipolata ed analizzata dal software di text-mining in R e come precedentemente descritto, è stato possibile avere in output una nuova base dati di commenti filtrati per il loro contenuto informativo, una tabella di probabilità di generare un Topic da un dato commento e il set di parole associate ad ogni commento. Per raggiungere un output ragionevolmente corretto, questa procedura è stata più volte eseguita secondo un iter che partiva dalle fasi di verifica e validazione dei risultati e ritornava alla pulizia del dataset di dati iniziale. In questo modo, seppur oneroso in termini di tempo, è stato perfezionato l'output in modo da farlo passare ai test di verifica e di validazione.

In questa fase è emersa la necessità di velocizzare il processo iterativo in cui gran parte del tempo impiegato era dovuto al fatto di dover unificare diverse tipologie di file, con contenuti di informazione totalmente differenti (es. valori tabellati in fogli di calcolo .csv e file di testo in documenti testuali .txt).

Dopo aver testato la correttezza funzionale del processo sono stati creati dei blocchi di codice che hanno permesso di automatizzare l'importazione dei dati ogni qualvolta fosse richiesto e di predisporre l'ambiente di lavoro in modo che risultasse agevole per le successive analisi.

Nell'allegato 1 vengono riportati i codici generati. L'ambiente di lavoro utilizzato è stato VBA (Visual Basic for Application) in Excel.

Il numero ottimale di Topic per l'iterazione finale, la più accurata, è di 34 ed è stato stabilito eseguendo il trade-off tra held-out likelihood, coerenza semantica ed esclusività. In particolare, si è preferito utilizzare il primo valore che rendesse stazionaria la curva della held-out likelihood e che avesse il più alto valore di coerenza semantica ed esclusività, come si può verificare in Figura 13.

Questi Topic sono stati poi etichettati: per ogni Topic si è studiato un campione dei commenti più pertinenti e ne è stata poi associata un'etichetta rappresentativa di tale argomento percepito dall'algoritmo STM.

Il risultato che è emerso da una prima analisi è l'individuazione di tematiche molto tecniche (Tabella 5) e specifiche per diversi settori, collegate a tematiche meccaniche ed informatiche. La ragione di tale output è da ricercare nella base dati analizzata, molto tecnica, ma l'unica disponibile per l'identificazione di UGC validi, in quanto, il mercato di tale prodotto risulta specifico ed ancora di nicchia, dunque non accessibile ad un campione di utenti elevato.

Con la fase di etichettatura si è potuto avere a disposizione un set di argomenti classificati in base alla loro frequenza di generazione dal dataset di documenti iniziale. Gli algoritmi di text-mining si fermano solitamente a questo risultato che, per la generazione dei requisiti funzionali di prodotto non è soddisfacente.

Il set di argomenti, in questo contesto, deve essere considerato come un cluster di requisiti di prodotto che non garantisce nessun contenuto informativo su di essi.

Si è reso necessario, quindi, adottare un approccio innovativo per la generazione di requisiti utenti a partire dai risultati di un algoritmo di text-mining.

Per l'individuazione di tale metodologia è risultato utile approcciarsi al mondo degli *Small Data*. Il data mining, invece, viene solitamente eseguito su grandi moli di dati per determinare i cosiddetti *Big Data*.

3.2. SMALL DATA E BIG DATA

In genere, gli analisti definiscono i big data con la regola delle "tre V": volume, varietà e velocità. In realtà, le tre V non sono caratteristiche dei soli big data ma sono ciò che rende i big data e gli small data diversi l'uno dall'altro (Porru, 2018, Rocchi, 2020).

- Volume: il volume dei dati è l'enorme quantità di dati che devi elaborare. I big data implicano maggiori quantità di informazioni mentre gli small data sono, non sorprendentemente, più piccoli. I big data sono spesso usati per descrivere enormi blocchi di informazioni non strutturate. Gli small data, d'altra parte, implicano metriche più precise e minuscole.
- Varietà: la varietà dei dati si riferisce al numero di tipologie di dati. Un esempio intuitivo preso dalla letteratura è il seguente: se si sta analizzando il traffico verso il sito web di un'azienda, i "big data" potrebbero riferirsi all'intero numero di visitatori, indipendentemente da come hanno raggiunto il sito o dalle loro qualità demografiche. Gli small data tendono a concentrarsi su un tipo di dati, quindi, gli "small data" potrebbero essere un'analisi di tutti i visitatori che hanno trovato la tua attività tramite la pubblicazione sui social media.
- Velocità: si intende la velocità con cui le informazioni vengono acquisite ed elaborate. In genere, i big data coinvolgono enormi blocchi di informazioni importate e analizzate in contenuti periodici.

(Porru, 2018, Rocchi, 2020)

Si è deciso di basare l'approccio sull'utilizzo degli small data per i seguenti vantaggi:

- Gli small data sono ovunque
- Gli small data si traducono immediatamente in business intelligence e spesso derivano da analisi compiute da esseri umani piuttosto che da algoritmi di data mining
- Se combinati con i big data permettono di garantire una comprensione dai dati superiore e di raggiungere in modo più efficiente il cliente finale ai quali sono riferiti i dati
- Si basano spesso su esperienze sensoriali e sperimentali degli utenti finali su un dato prodotto, dunque, sono maggiormente veritieri
- Pongono il focus direttamente sul cliente finale

- Sono semplici da comprendere

Seppur ben definiti dalla letteratura, gli small data non sono ancora in grado di essere compresi totalmente e ad oggi, non esistono approcci che permettono di modellarli. La ragione principale di tale limite è da associare sia al fatto che sono stati recentemente scoperti, sia perché non possono essere univocamente definiti. Risultano dunque difficili da individuare in quanto richiedono un'analisi approfondita per poterne determinare un metodo operativo di identificazione. Raggiunta questa fase, i vantaggi precedentemente descritti sono validi. (Hu, 2014)

Si pensi, per esempio, al caso LEGO (<https://www.blog-management.it/2018/09/14/case-study-sneakers-lego-small-data/>) in cui l'azienda è stata salvata dagli Small Data, grazie ad un dettaglio che è emerso durante un'analisi ma il cui metodo d'indagine non è generalizzabile e valido per altri contesti.

Nel caso studio della robotica collaborativa, gli small data sono identificati da determinate keywords (o da combinazioni di esse) nei vari Topic.

Le parole associate ad ogni Topic permettono all'algoritmo STM di generare correlazioni per definire un certo argomento, ma sono anche utilizzate per aiutare il ricercatore nell'etichettatura degli argomenti.

Nel caso della ricerca di requisiti funzionali di prodotto, le parole, se inserite in un contesto discorsivo prendono un significato non percepibile apparentemente che può essere tradotto in uno o più requisiti di prodotto.

La condizione necessaria per tale approccio è avere un output dell'STM preciso e robusto in termini di consistenza, coerenza e correttezza.

L'approccio algoritmico agli small data che è stato adottato per questa analisi è di seguito riportato:

1. *Preparazione dei dati necessari all'identificazione del requisito*

Identificare per ogni argomento il set di parole associate e i documenti più pertinenti. Avere a disposizione un output STM ben strutturato è sufficiente per semplificare questa prima fase.

2. *Comprendere l'origine del requisito*

Comprendere i documenti richiesti e associare ciascuna keyword al contesto argomentativo considerato. In questo caso, etichette ben definite possono velocizzare il processo di identificazioni di tali correlazioni tra keywords e Topic. In questa fase è importante considerare non solo le keywords, ma gli aggettivi associati ad esse; la ragione di tale

sottigliezza risiede nella necessità di generare requisiti utente robusti ed univoci, privi di ogni tipo di ambiguità.

Un esempio che è emerso nell'analisi è il seguente:

Combinazioni di keywords portano a generare requisiti del tipo “velocità del braccio robotico”. Questo tipo di requisito se considerato in quanto tale, può essere fortemente distorsivo per l'analisi se non privato di ogni possibile ambiguità. La velocità del braccio robotico inerente al tema sicurezza è associata a parole quali “diminuzione, riduzione, limitazione...”; mentre, se associata al suo processo di lavorazione è associata a “incremento, aumento...”.

Nell'esempio precedente si nota un risultato fortemente distorsivo e contraddittorio, per tale ragione, ogni parola deve essere attentamente analizzata ed associata al suo contesto in modo da renderla significativa e priva di ambiguità. I Topic, come precedentemente detto, risultano in modo evidente come cluster di requisiti che possono essere tra loro aggregati per formare macro-tematiche di requisiti funzionali del prodotto.

Da qui l'importanza di vedere le keywords come uno small data che si inserisce e descrive al meglio il Topic definito dunque come un big data.

È importante sottolineare le seguenti relazioni che vengono a generarsi:

- i. Non tutti i Topic sono associati ad almeno un requisito

Esempio:

In riferimento alla Tabella 8 e alla Tabella 9, il Topic “Spinta alla transizione al 5.0” non è associato a nessuna funzionalità di prodotto.

- ii. Una parola può essere associata ad un solo requisito per Topic

Esempio:

In riferimento alla Tabella 5 e alla Tabella 6, la parola “drive” relativa al Topic “Passaggio Free Drive - Remote Control” è associata solamente al requisito “Free Drive Mode”.

- iii. Una parola può essere associata a più requisiti considerando Topic differenti

Esempio:

In riferimento alla Tabella 5 e alla Tabella 6, la parola “speed” è associata a due requisiti diversi referenziati da due Topic distinti. Al Topic “Controllo velocità” corrisponde il requisito “Velocità operativa” mentre al Topic “Sicurezza” corrisponde il requisito “Regolazione velocità di esecuzione in presenza di operatore”.

- iv. Una parola può essere associata a nessun requisito

Esempio:

In riferimento alla Tabella 6, la parola “robot” non è associata a nessun requisito.

- v. Un requisito è associato ad un solo Topic

Esempio:

In riferimento alla Tabella 5 e alla Tabella 6, il requisito “Velocità operativa” è associato solamente al Topic “Controllo velocità”.

- vi. Un requisito può essere associato a più parole

Esempio:

In riferimento alla Tabella 5 e alla Tabella 6, il requisito “Semplicità di definizione degli waypoint” può essere associato a più parole come “waypoint, point, posit(ion), move, frame”

La scelta di quali parole considerare e quali Topic sono considerati pertinenti per una generazione dei requisiti è una scelta che richiede un input umano, dunque soggetta ad arbitrarietà e soggettività.

3. Definire il requisito

Definire un’etichetta pertinente per ogni requisito che permetta di renderlo comprensibile ed univoco per un lettore esterno. Il requisito deve essere comprensibile e adattato al linguaggio naturale considerato per l’analisi e ai tecnicismi del settore di riferimento

L’analisi può essere perfezionata stabilendo un ordine nella scheda dei requisiti.

Nel caso specifico è stata adottata una procedura che permettesse di clusterizzare i requisiti in blocchi di pertinenza. È stato fatto questo aiutandosi con un risultato dell’algoritmo in R che ha

permesso di generare cluster di argomenti in base alla correlazione generata da ogni Topic. L'algoritmo STM consente di evidenziare correlazioni tra argomenti. Correlazioni positive tra gli argomenti indicano che entrambi gli argomenti saranno probabilmente discussi all'interno di un documento. Gli argomenti vengono poi riassunti in grafici che possono essere a scelta dell'utente; nel caso in analisi si è scelto un plotting associato ad un grafo, come si vede in Figura 16.

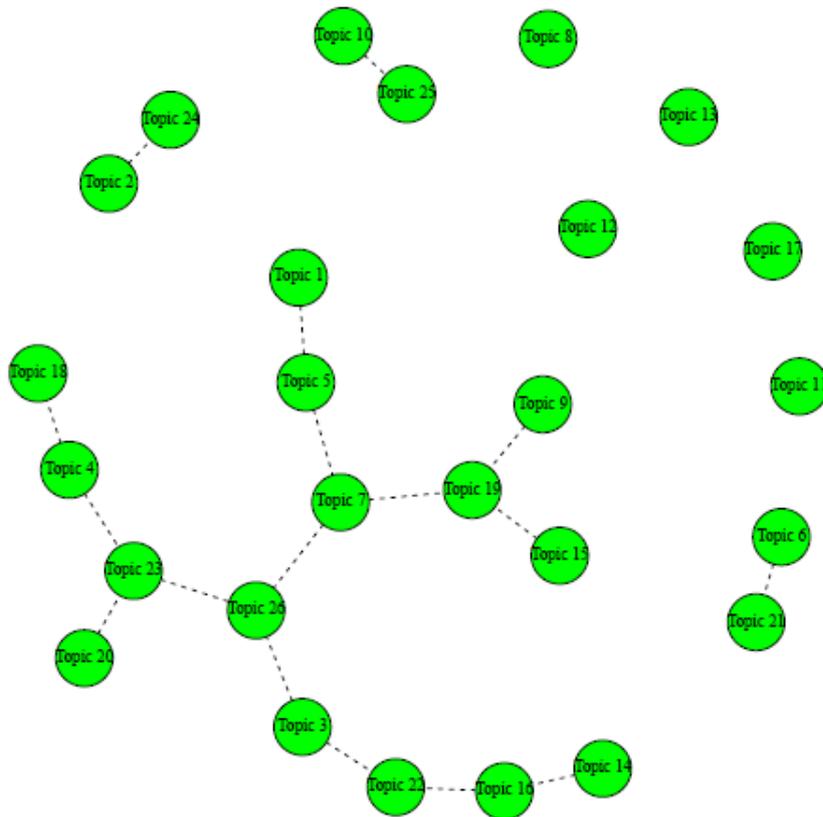


Figura 16. Grafo delle correlazioni riferito alla clusterizzazione dei requisiti nel caso applicativo sulla robotica collaborativa in cui emergono le correlazioni dei Topic

Il grafo delle correlazioni viene filtrato in base ad un valore minimo di soglia che deve essere rispettato per considerare i Topic tra loro correlati. Definire valori di soglia molto alti significa azzerare i legami di correlazione, viceversa considerare valori di sogli molto bassi può riempire il grafo di correlazioni minimizzando il contenuto informativo apportato da tale processo. Il valore soglia deve essere scelto dall'utente in una procedura iterativa in cui si verifica di volta in volta il valore informativo aggiunto all'analisi con l'output ottenuto.

Risulta necessario evidenziare che questo processo di clusterizzazione non deve essere necessariamente seguito per l'analisi ma piuttosto deve fungere da conferma per rilevare il grado di coerenza del lavoro svolto con i risultati ottenuti.

Il processo di generazione di requisiti utente, infatti, pone l'enfasi principale sul legame parola-argomento; una volta definito il set di requisiti e associandoli ciascuno ad un Topic è possibile definire set di cluster in modo autonomo senza utilizzo di algoritmi specifici. In questo modo si attribuisce al risultato finale una soggettività che non incide in modo rilevante sul risultato finale. L'aiuto del clustering da parte dell'algoritmo serve non tanto ad aggiungere un contenuto informativo, ma piuttosto a verificare se la soggettività inserita nell'analisi ha un riscontro positivo da parte dell'algoritmo di text-mining; in questo modo la scheda di requisiti aumenta di molto la qualità in termini di coerenza strutturale e argomentativa.

Nella Tabella 6 e nella Tabella 9 vengono riportati rispettivamente i requisiti funzionali e le funzionalità dei robot collaborativi, inerenti al caso studio, clusterizzati grazie all'adozione del grafo delle correlazioni.

3.3.DISACCOPPIAMENTO DEI REQUISITI FUNZIONALI DALLE FUNZIONALITÀ DEL PRODOTTO

Il risultato finale di quanto detto fino ad ora è una scheda dei requisiti funzionali di prodotto raggruppati in cinque insiemi.

In questo modo, una scheda dei requisiti utente può essere diversamente letta e percepita dall'analista per creare valore aziendale; attualmente il mondo della robotica collaborativa, seppur estremamente diversificato, appare molto simile in termini di prestazioni di prodotto e percezione della qualità del prodotto.

Per aggiungere valore all'analisi finora conseguita, si è voluto definire una metrica di confronto che riuscisse a stabilire in che modo i requisiti trovati potessero essere ulteriormente filtrati per determinare possibili fonti di differenziazione per le aziende produttrici nel mercato di riferimento. L'ipotesi di partenza deriva dal fatto che un requisito funzionale può essere diversamente percepito dal mercato. In questo senso, si è voluto adottare il concetto di *funzionalità del prodotto* percepita dal mercato, in modo da avere un benchmark di confronto per l'analisi.

Con il termine *funzionalità* si intende il modo in cui il prodotto risponde a determinate esigenze. In questo contesto, si disaccoppia il requisito funzionale (funzionalità richiesta dal cliente) dalla

funzionalità del prodotto (requisiti funzionali del cliente percepiti dal mercato). La tecnica di analisi consiste nel parallelizzare lo studio lato domanda e lato offerta del mercato di riferimento e vedere in che modo i requisiti vengano percepiti.

In questa fase è stata seguita nuovamente la procedura precedentemente descritta, partendo però da una base dati differente che potesse rappresentare le percezioni il mercato e allo stesso tempo fosse di dominio pubblico. È risultato fondamentale distaccarsi dal concetto di UGC, in particolare per quello che riguarda il disinteressamento dell'informazione contenuta. Il nuovo dataset è stato costruito su un contenuto informativo promozionale e divulgativo, dunque influenzato dalle conoscenze delle aziende nel mercato di riferimento, come per esempio blog, notizie promozionali, casi applicativi descrittivi...

Successivamente è stato eseguito il data cleaning in modo personalizzato per ogni insieme di documenti estratti da diverse pagine web.

La base di dati finale è risultata di circa 2100 post prettamente correlati alla tematica della robotica collaborativa.

Il numero ottimale di Topic identificati dall'algoritmo STM è di 82.

L'elevato numero di Topic è da attribuire al fatto che i contenuti dei post sono ricchi di informazioni e tali informazioni sono estremamente variegata in quanto sfociano in diversi ambiti culturali: da possibili tendenze di sviluppo del prodotto agli ambiti di applicabilità, da contesti politici a informazioni prettamente tecniche.

In un processo iterativo, gli 82 argomenti, identificati dall'algoritmo STM in R, sono stati poi etichettati (con la tecnica precedentemente descritta) in modo sempre più accurato fino a superare i test di verifica e di validazione.

4. APPLICAZIONE NEL SETTORE DELLA ROBOTICA COLLABORATIVA

4.1. REQUISITI DI PRODOTTO

Come precedentemente spiegato, i Topic generati dall'algoritmo di text-mining sono 34. Nella Tabella 5 si riportano i risultati ottenuti.

Prevalenza Topica	Nome del Topic	Descrizione del Topic
0.00364	Style guide e subroutine generiche	Integrabilità dei software di programmazione con il linguaggio HTML
0.01275	Servizio di supporto delle aziende produttrici	Efficacia del servizio di supporto post-vendita
0.012869	Cobot dress pack	Usabilità dei rivestimenti dei bracci robotici per particolari tipologie di lavorazione (es. rivestimento isolante per lavorazioni che richiedono altissime temperature)
0.017557	Remote Procedure Call	Possibilità di programmazione "a distanza" su computer remoti
0.018306	Diversificazione terminali	Diversificazione nelle tipologie dei gripper di lavorazione
0.018592	Installazione packages ambienti di supporto	Efficienza nell'interoperabilità del software principale con ambienti di supporto
0.020178	Definizione waypoint	Programmazione delle funzioni di riconoscimento del braccio robotico nello spazio tridimensionale
0.021128	Script operativo	Codici di programmazione operativa generali

0.021245	Documentazione completa mondo UR	Efficienza nella comunicazione da parte delle aziende produttrici, intesa come la capacità di rendere l'operatore autonomo nella gestione del proprio braccio robotico senza costi di consulenza aggiuntivi
0.022935	Driver ROS	Possibilità di interagire con un insieme di framework per lo sviluppo e la programmazione di robot; in particolare ROS che è quasi uno standard per l'ambiente della robotica
0.024498	Passaggio Free Drive - Remote Control	Possibilità di guidare il braccio robotico in remoto attraverso la programmazione o tramite l'apprendimento automatico in base al percorso insegnato dall'operatore
0.02483	Parallelismo esecuzione dei Thread	Possibilità di poter svolgere un numero maggiore possibile di thread (blocchi di codice) in parallelo per evitare l'insorgere di bug
0.026159	Wrist Camera	Problematiche inerenti alla sensoristica visiva dei bracci robotici. Combinando tipologie differenti di gripper e telecamere si ottiene una grande flessibilità nelle mansioni e nella sensibilità del robot.
0.026524	Implementazione alberi n-ari	Utilità nell'adozione di tecniche di programmazione chiamate alberi n-ari per garantire flessibilità, snellimento ed organizzazione del codice
0.027882	Ambiente KUKA	Problematiche inerenti ai modelli IIWA dell'ecosistema KUKA
0.028831	Compatibilità versioni software	Errori di comunicazione tra versioni differenti di software
0.029612	Movimenti braccio_Script Operativo	Condivisione di frammenti di codice inerenti la programmazione del movimento dei bracci,

		dunque, la fluidità del movimento, la velocità di rotazione...
0.029873	Modellazione concettuale operativa	Condivisione di modelli concettuali per l'utilizzo di un braccio robotico in nuovo processo aziendale/lavorativo
0.030055	Generalità script	Correzioni/errori di frammenti di codice dovuti ad una sintassi non corretta
0.031653	Sicurezza	Sicurezza in presenza e senza l'operatore del braccio robotico
0.032713	Interfacce di comunicazione	Ottimizzazione della comunicazione, ad esempio attraverso l'utilizzo di maschere di sottorete, dell'interfaccia RTDE e delle altre porte di comunicazione
0.037387	Controller - terminali	Corretto funzionamento di un istanza di controllo continuo (controller), in particolare riguardo problemi di connessione interna I/O e sulle potenze utilizzate (Voltaggio)
0.039166	Programmazione Request Protocol	Condivisione di codice su come attivare correttamente l'interfaccia client
0.042509	Controllo velocità	Importanza nel diminuire le velocità entro tempi e spazi richiesti, cioè farlo diminuire in modo lineare senza danneggiare il braccio robotico col carico
0.043828	Diversificazione problematiche e difficoltà nel gestirle da parte di esperti certificati	Difficoltà di reperire soluzione in modo autonomo e senza costi aggiuntivi a causa di una forte diversificazione delle problematiche da risolvere dovuta alla flessibilità di utilizzo dei bracci robotici collaborativi
0.044042	Movimenti braccio_ScriptTeorico	Definizione di concetti teorici per poter muovere correttamente il braccio robotico, circa l'utilizzo della cinematica inversa e il bilanciamento dei carichi per coppie di forza

0.044183	Condivisione codice risolutivo	Condivisione di codice modulare risolutivo per determinate problematiche molto diffuse nella programmazione di un braccio robotico
0.044855	Programmazione off-line	Importanza di poter programmare il braccio robotico off-line in maniera ordinata e flessibile in modo da snellire il processo di programmazione
0.04642	Applicazione operative	Condivisione di idee inerenti alla possibilità di migliorare un processo aziendale/lavorativo attraverso l'adozione della robotica collaborativa
0.056434	Confronto tra marchi	Confronto tra aziende produttrici di bracci robotici in termini di economicità, compatibilità con tecnologia e software esistenti e facilità di reperimento di pezzi di ricambio
0.057693	Soddisfazione sull'efficacia del forum	Soddisfazione degli utenti sull'efficacia del forum
0.061653	Aggiornamento firmware	Errori di comunicazione tra hardware e software dovuto ad un'obsolescenza rapida del firmware

Tabella 5. Labelling dei Topic inerenti ai requisiti funzionali dei bracci robotici collaborativi

Nella Tabella 5 compare la porzione (cioè la prevalenza media Topica) degli argomenti identificati dal data set di UGC analizzato. Gli argomenti più discussi riguardano gli aggiornamenti dei firmware di comunicazione hardware-software, il confronto tra differenti marchi di prodotto e la soddisfazione degli utenti sull'efficacia dei forum (essendo il data set iniziale composto da commenti fatti da utenti su forum, risulta non troppo significativo per l'analisi). A seguire una serie di tematiche riguardanti aspetti informatici, meccanici e applicativi del prodotto.

Da questi Topic sono stati generati una serie di requisiti funzionali che sono poi stati classificati in cinque cluster argomentativi.

Applicando la metodologia precedente descritta sono stati identificati quattro tematiche argomentative riguardanti il prodotto fisico (Sicurezza, Convenienza, Customizzazione, Operatività, Software) e i servizi ad esso associati.

Di seguito si riporta una descrizione di tali categorie:

- *Sicurezza*: riguarda l'interazione uomo-macchina, in particolare in che modo l'operatore può essere messo in sicurezza da eventuali malfunzionamenti o anomalie. Tra le tematiche più importanti vi sono la necessità di sensoristica avanzata per riconoscimenti di operatori e layout, la possibilità di riduzione delle velocità in determinati contesti collaborativi e segnali visivi che permettano di capire lo stato del COBOT durante il suo processo lavorativo.
- *Convenienza*: riguarda tematiche inerenti all'efficacia e all'efficienza della manutenzione nel breve e nel lungo periodo, i costi da affrontare per l'investimento, il pay back period ad essi associato e l'adattabilità dei componenti del prodotto. In questo contesto emerge delle chiare strategie di lock-in da parte delle aziende produttrici le quali, non esistendo uno standard di settore, difficilmente progettano componenti hardware e software in modo universale.
- *Customizzazione*: riguarda il grado di personalizzazione che può avere il prodotto COBOT, dunque il grado di adattabilità del COBOT alle diverse esigenze del mondo manifatturiero. Avere una vasta possibilità di scelta del prodotto è una condizione necessaria per l'affermazione del prodotto che per definizione deve garantire vantaggi in termini di flessibilità operativa. Un basso grado di personalizzazione richiede, invece, una pianificazione e progettazione ad hoc che necessariamente implica un elevato grado di rigidità per possibili sviluppi strategici aziendali.
- *Operatività*: riguarda in che modo i bracci robotici riescono ad adattarsi in termini di movimenti possibili, spazio occupato durante la movimentazione, velocità di lavorazione, sensibilità e precisione... Questo aspetto incide notevolmente sulle prestazioni del braccio robotico e può mettere in evidenza limiti tecnologici operativi.
- *Software*: riguarda i vari software che interagiscono col braccio robotico, in termini di pacchetti di estensioni per avere una comunicazione più efficace ma anche in termini di ottimizzazione del firmware del prodotto. Problematiche relative ad errate progettazioni software mettono in risalto dei bug che l'utente può percepire sia in fase di progettazione da remoto del lavoro del braccio robotico sia in fase di operatività dello stesso.
- *Servizi*: riguarda il pacchetto di servizi associato al prodotto. Il tema è legato alla rapidità delle prestazioni di manutenzione, servizi post-vendita e servizi di istruzione per gli utenti del prodotto. Un tema che emerge nell'ambito dei servizi è la volontà da parte dell'utente di essere autonomo nell'utilizzo del COBOT in modo da essere immediatamente reattivo ed efficiente nella risoluzione dei problemi. Il concretizzarsi della flessibilità operativa è, in

questo contesto, strettamente legato a strumenti che facilitano l'accesso alla conoscenza in modo libero, come per esempio piattaforme di condivisione di contenuti web inerenti a tale argomento ben strutturate. In questo modo, l'utente può permettersi di aspettare le tempistiche di servizio di assistenza aziendale per problemi prettamente tecnici e risolvere all'istante ed autonomamente problemi riferiti all'adattabilità del COBOT al contesto produttivo.

Nella Tabella 6 si riportano i risultati ottenuti.

CONVENIENZA
Economicità
Durabilità
Compatibilità ricambi
OPERATIVITÀ
Stabilità in esecuzione con qualsiasi carico
Stabilità ad alte vibrazioni
Aumento numero assi
Aumento ampiezza raggio di lavoro
Anticollisione in esecuzione
Sensori di coppia di forza (versatilità operativa)
Movimentazione morbida del braccio
Velocità operativa
Regolazione continua del braccio
Adattamento alla tensione di esercizio (stabilizzatori)
Ridondanza porte di comunicazione
Scalabilità del controller
Latenza di esecuzione
Scorrevolezza dress pack ai movimenti del Cobot
SOFTWARE
Compatibilità continua delle versioni del firmware (es. aggiornamenti automatici)
Garantire programmazione da remoto
Facilità di trasferimento dati

Facilità di attivazione interfaccia client
Diversificazione delle interfacce di comunicazione
Aumentare usabilità tech-pendant
Facilità di programmazione movimentazione braccio
Compatibilità con versioni dei software di supporto
Parallelizzare Thread in esecuzione
Free Drive Mode
Remote Control Mode
Semplicità di definizione degli waypoint
Implementare RPCXML (facilità di interazione con sistemi software)
SICUREZZA
Arresto di emergenza
Luci di monitoraggio stato robot
Laser come barriera di sicurezza
Regolazione velocità di esecuzione in presenza di operatore
CUSTOMIZZAZIONE
Customizzare sensoristica visiva
Diversificazione terminali
Customizzazione Cobot dress pack
SERVIZIO
Offrire codice modulare
Velocità di fornitura del servizio/prodotto
Istruzione operatore tecnico
Istruzione programmatore
Condivisione di modelli strutturali operativi
Interoperabilità con ambienti open source
Documentazione online completa e di immediata comprensione
Documentazione online ben strutturata
Supporto post-vendita garantito

Tabella 6. Requisiti funzionali dei bracci robotici collaborativi

I requisiti sono stati poi ordinati secondo il grado di prevalenza Topica ad essi associato, come mostrato in Figura 18.

Occorre evidenziare che non significa che in base a questo ordinamento alcuni siano meno o più "critici per la qualità" rispetto ad altri. Non esiste una netta correlazione tra importanza del requisito percepita dall'utente e il grado di prevalenza Topica identificato dall'algoritmo STM. La differenza di porzioni dipende sicuramente dal campione di dati estratto ma anche da una serie di fattori quali per esempio gli aggregatori e indicatori utilizzati per l'analisi.

Tuttavia, si è deciso di non considerare questo aspetto. La ragione di tale scelta è basata sul confronto con le varie iterazioni fatte per perfezionare i Topic. Le varie iterazioni, seppur non depurate da ogni fonte di errore, mantenevano all'incirca una disposizione costante: l'importante, per questa ipotesi, è verificare che l'ordinamento delle parole generate dai Topic fosse pressoché identico ad ogni iterazione compiuta. Tale passaggio operativo è stato molto oneroso in termini di tempo ma ha permesso di rilassare l'ipotesi precedente. L'importanza dei requisiti è associata in modo approssimato al grado di prevalenza Topica dei Topic associati ad essi, dunque, al grado di interesse che gli utenti hanno espresso nel complesso. Essendo i requisiti generati dalle keywords che descrivono i vari Topic ed essendo ogni requisito univocamente identificato, è possibile associare il grado di prevalenza Topica ad un requisito.

In questo modo, è stato possibile fare un ordinamento ulteriore della tabella dei requisiti. I requisiti sono stati disposti in ordine decrescente per ogni blocco e analogamente i cluster di requisiti sono stati ordinati in base alla media dei gradi di prevalenza Topica dei requisiti ad essi associati. Inoltre, avendo a disposizione per ogni documento del data set la data di creazione corrispondente, è stato possibile determinare l'evoluzione della proporzione tematica nel tempo degli argomenti, dunque dei requisiti determinati.

In Tabella 7 è costruito un esempio della procedura adoperata, applicata ad un Topic:

<i>Topic</i>	<i>Commento</i>	<i>Anno</i>	<i>Covariato</i>	<i>Media annuale dei covariati</i>
Aggiornamento Firmware	Commento1	2018	0,01873	0,01873
	Commento2	2019	0,01967	0,01929
	Commento3	2019	0,01891	
	Commento4	2020	0,02002	0,02252
	Commento5	2020	0,02502	
	Commento6	2021	0,01244	0,01233
	Commento7	2021	0,01213	
	Commento8	2021	0,01311	
	Commento9	2021	0,00997	
	Commento10	2021	0,03532	

Tabella 7. Esempio della procedura adoperata per identificare l'evolvere della porzione tematica nel tempo, applicata ad un Topic del caso studio sulla robotica collaborativa. Utilizzando la media annuale dei covariati è possibile, plottando i valori, verificare in modo grafico l'evoluzione nel tempo

Nella Figura 17 vengono riportati gli andamenti dei Topic dei requisiti dei robot collaborativi al variare del tempo, su base annuale.

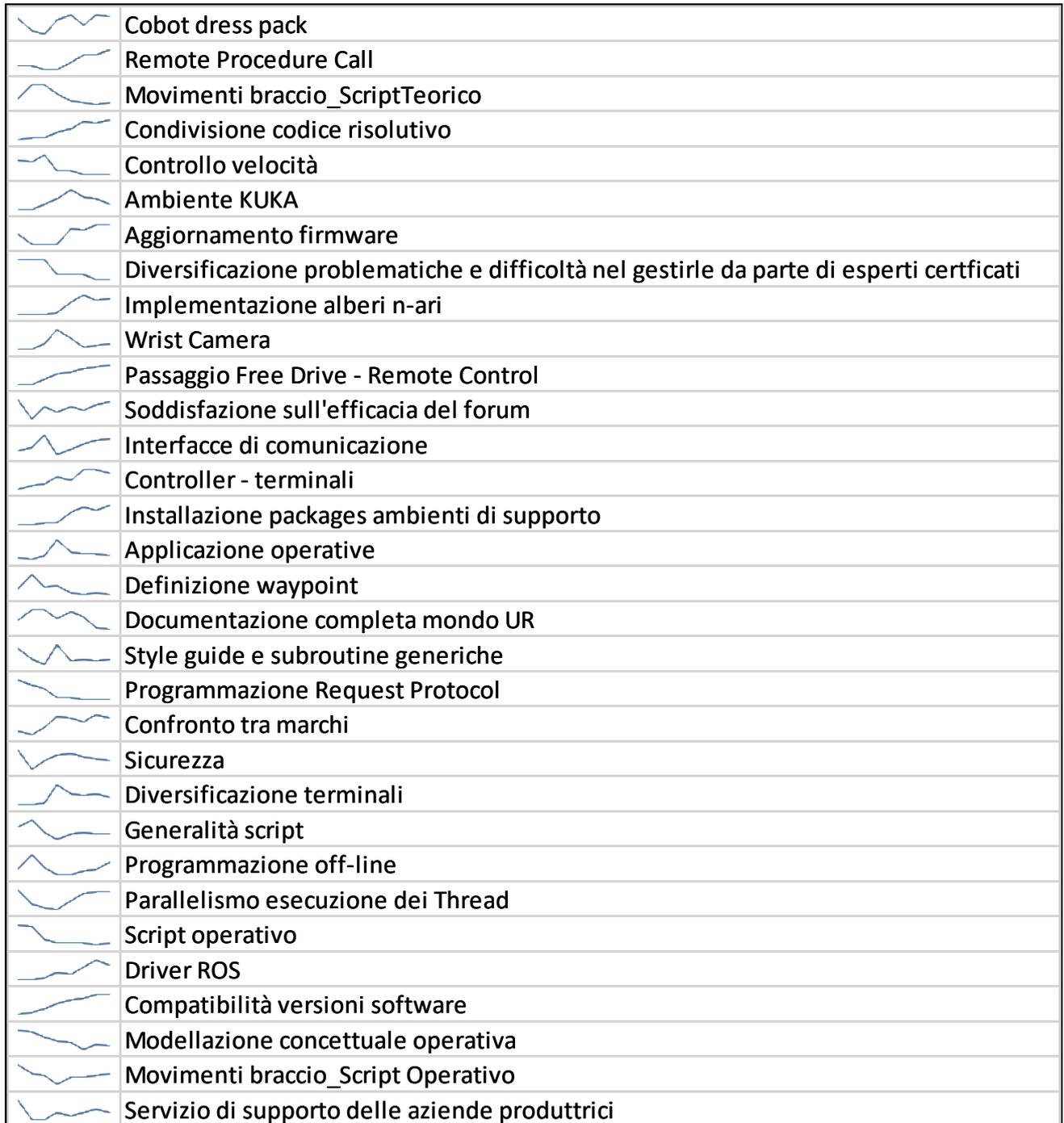


Figura 17. Rappresentazione dell'andamento dei Topic referenziati dai requisiti nel tempo, su base annuale, dal 2013 al 2021

È possibile associare ogni evoluzione temporale dei Topic ai requisiti referenziati da ognuno, come riportato in Figura 18.

	CONVENIENZA	0,05643
	Economicità	0,05643
	Durabilità	0,05643
	Compatibilità ricambi	0,05643
	OPERATIVITA	0,03959
	Stabilità in esecuzione con qualsiasi carico	0,04642
	Stabilità ad alte vibrazioni	0,04642
	Aumento numero assi	0,04404
	Aumento ampiezza raggio di lavoro	0,04404
	Anticollisione in esecuzione	0,04404
	Sensori di coppia di forza (versatilità operativa)	0,04404
	Movimentazione morbida del braccio	0,04251
	Velocità operativa	0,04251
	Regolazione continua del braccio	0,04251
	Adattamento alla tensione di esercizio	0,03739
	Ridondanza porte di comunicazione	0,03739
	Scalabilità del controller	0,03739
	Latenza di esecuzione	0,03271
	Scorrevolezza dress pack ai movimenti del Cobot	0,01287
	SOFTWARE	0,03256
	Compatibilità continua delle versioni del firmware	0,06165
	Garantire programmazione da remoto	0,04486
	Facilità di trasferimento dati	0,04486
	Facilità di attivazione interfaccia client	0,03917
	Diversificazione delle interfacce di comunicazione	0,03271
	Aumentare usabilità tech-pendant	0,03005
	Facilità di programmazione movimentazione braccio	0,02961
	Compatibilità con versioni dei software di supporto	0,02883
	Parallelizzare Thread in esecuzione	0,02483
	Free Drive Mode	0,02450
	Remote Control Mode	0,02450
	Semplicità di definizione degli waypoint	0,02018
	Implementare RPCXML	0,01756
	SICUREZZA	0,03165
	Arresto di emergenza	0,03165
	Luci di monitoraggio stato robot	0,03165
	Laser come barriera di sicurezza	0,03165
	Regolazione velocità di esecuzione in presenza di operatore	0,03165
	CUSTOMIZZAZIONE	0,01911
	Customizzare sensoristica visiva	0,02616
	Diversificazione terminali	0,01831
	Customizzazione Cobot dress pack	0,01287
	SERVIZIO	0,02846
	Offrire codice modulare	0,04418
	Velocità di fornitura del servizio\prodotto	0,04383
	Istruzione operatore tecnico	0,03005
	Istruzione programmatore	0,03005
	Condivisione di modelli strutturali operativi	0,02987
	Interoperabilità con ambienti open source	0,02294
	Documentazione online completa e di immediata comprensione	0,02124
	Documentazione online ben strutturata	0,02124
	Supporto post-vendita garantito	0,01275

Figura 18.
Rappresentazione dell'andamento dei requisiti nel tempo, su base annuale, dal 2013 al 2021. Nella prima colonna viene evidenziato graficamente l'andamento temporale, nella seconda colonna sono riportati i requisiti clusterizzati, sulla terza colonna viene riportato il valore medio della prevalenza Topica del requisito

Si possono identificare tre diversi gruppi di comportamento evolutivi:

- *Stazionari*: sono i requisiti che hanno mantenuto invariato il loro interesse per gli utenti nel tempo. Tra questi troviamo la stabilità del braccio robotico in esecuzione, il grado di customizzazione del prodotto, l'istruzione garantita dell'operatore futuro e numerosi requisiti inerenti all'ambiente di programmazione.
- *Decrescenti*: sono i requisiti che presentano una chiara diminuzione della *Topic proportion* nel tempo. Tra questi troviamo temi legati alla sicurezza e alla possibilità di adattamento della velocità operativa.
- *Crescenti*: sono i requisiti la cui proporzione è cresciuta nel tempo. Tra questi troviamo l'economicità del prodotto, la facilità di reperibilità di ricambi, la possibilità da parte dell'operatore di interagire con il robot in modalità free o drive e l'efficienza del servizio post-vendita.
- *Periodici*: sono requisiti che presentano una stagionalità nell'interesse negli anni. Tra questi troviamo requisiti prettamente tecnici come il numero di assi, la sensoristica dei terminali, etc...

Una possibile spiegazione del fenomeno può essere dovuta all'utilizzo del COBOT da parte delle aziende; il prodotto inizialmente presentava tecnologie perfezionabili soprattutto per determinati contesti produttivi, mentre, con l'esponentiale crescita del numero di utilizzatori di tale prodotto potrebbero essere nate nuove esigenze tecniche che mirano ad efficientare quelle attuali, in termini di precisione e adattabilità.

4.2.FUNZIONALITÀ DEL PRODOTTO

Come precedentemente spiegato, i Topic generati dall' algoritmo di text-mining sono 82. Nella Tabella 8 si riportano i risultati ottenuti.

Prevalenza Topica	Nome del Topic	Descrizione del Topic
0.00313	Interfacce di comunicazione	Importanza nella corretta interazione tra le interfacce di comunicazione
0.003683	Chuck Cobot	Modello di robot collaborativo che utilizza l'apprendimento automatico e l'intelligenza artificiale per aiutare i dipendenti a compiere le mansioni più velocemente, rimanere concentrati e lavorare in modo più efficiente
0.003789	Definizione piani e variabili di posizione	Teoria per il corretto funzionamento del braccio robotico
0.00391	Cobot come integratore logica lean	Grazie alla flessibilità dei bracci robotici è possibile abbassare il costo unitario dei processi/prodotti inserendoli nel contesto applicativo
0.004496	Sensoristica di sicurezza	Importanza dello sviluppo della sensoristica per monitoraggio continuo ed efficiente
0.004735	Attivazione interfaccia client	Risoluzione di problematiche inerenti alla corretta attivazione dell'interfaccia client
0.004944	Piattaforme di comunicazione per utenti	Presenza di piattaforme di comunicazione per facilitare la risoluzione di problematiche
0.005644	Latenza di comunicazione	Importanza nel minimizzare i tempi di latenza di comunicazione per sincronizzare al meglio le informazioni
0.005651	Funzioni di script per programmazione	Condivisione di blocchi di codice in riferimento a determinate problematiche

0.005679	Protocollo di comunicazione (encoder)	Efficienza della trasmissione dati dell'encoder, utilizzato per misurare il movimento (velocità e posizione) dei bracci robotici
0.005692	AMR (Autonomous Mobile Robot)	Robot collaborativi che sono in grado di muoversi nello spazio in modo autonomo
0.006039	MOTOMAN con più carico	Modello di braccio robotico collaborativo con un miglioramento prestazionale sulla tenuta del carico
0.006147	Actinav UR	Robot collaborativo che integra visione intelligente e controllo del movimento autonomo in tempo reale per garantire precisione e rendere agile il processo di lavorazione
0.006885	Cambio utensili rapido	Importanza nel riuscire a mascherare il più possibile il tempo di cambio utensili
0.007354	CTO UR visione	Visione del Chief Technical Officer sul futuro dell'azienda Universal Robot
0.007377	Educativo	Utilità dei bracci robotici collaborativi nell'ambito educativo
0.007511	Saldatura	Potenzialità dei bracci robotici collaborativi per l'operazione di saldatura
0.00812	Sostituzione task ripetitivi e alienanti	Utilità dei bracci robotici collaborativi per la sostituzione di mansioni alienanti
0.008136	Stabilità in esecuzione	Importanza nel garantire stabilità e continuità dei movimenti nello spazio attraverso la definizione di punti di riferimento nello spazio (waypoint)
0.008242	Cobot NASA che si autoassemblano	Evoluzione dei bracci robotici collaborativi in ambito aerospaziale

0.008274	COBOT multibraccio	Possibilità di estensioni dei bracci robotici collaborativi con l'aggiunta di più bracci ad uno stesso blocco hardware (multibraccio)
0.008335	Modularità codice disponibile	Disponibilità di codice modulare per la programmazione dei bracci robotici offerta dagli utenti delle piattaforme di settore e dalle aziende produttrici
0.00854	COBOT per risanamento ambientale	Utilità dei bracci robotici collaborativi per mansioni potenzialmente dannose per l'uomo e ad alto impatto ambientale
0.008628	Compatibilità gripper	Importanza nel rendere compatibili universalmente i gripper provenienti da aziende produttrici diverse
0.008646	Sicurezza senza barriere, arresto e velocità bassa	Implementazione di sensoristica apposita per garantire la sicurezza potenzialmente dannose per l'uomo
0.008704	Operatività con ambiente ROS	Compatibilità dei software con l'ambiente ROS
0.008899	Istruzione operatore	Necessità da parte delle aziende produttrici di offrire un servizio di istruzione per i futuri operatori dei bracci robotici
0.008926	Encoder per sensori	Sviluppo delle prestazioni degli encoder
0.009222	ARM per istruire	Possibilità di muovere bracci robotici attraverso schede Arduino. Per la semplicità d'uso è possibile insegnare tali competenze anche nel mondo scolastico
0.009398	Sensibilità operativa	Garantire un'elevata precisione nelle lavorazioni dei bracci robotici anche in presenza di ostacoli e forze esterne nella traiettoria di lavorazione
0.009705	Wrist camera	Importanza delle telecamere da polso per la sensoristica visiva dei bracci robotici

0.00991	Servo-azionamenti per continuare ad operare anche con ostacoli	Utilità della tecnologia del servo-azionamento per poter operare anche in presenza di ostacoli senza modificare i parametri del braccio robotico
0.010174	Pinze a vuoto	Diversificazione delle tipologie di pinze a vuoto per garantire scalabilità e facilità d'uso
0.010353	Concentrazione mercato con acquisizioni di diversi settori	Mercato della robotica collaborativa molto concentrato e legato alle tecnologie di molti altri settori
0.010494	Digitalizzazione	Digitalizzazione come una necessità per poter adottare in modo conveniente le tecnologie collaborative
0.010759	COBOT polimorfo	Sensoristica adottata dai robot collaborativi a disposizione per la generazione di robot polimorfi in grado di auto-assemblarsi e ripararsi
0.011025	Programmazione movimenti spazio	Condivisione di codici per la programmazione del riconoscimento del robot collaborativo nello spazio
0.011118	Cobot per velocizzare le supply chain post-pandemia	Velocizzare i processi aziendali attraverso l'utilizzo di robot collaborativi per aiutare la ricrescita economica delle aziende
0.011175	IREX	È la più grande fiera al mondo di robotica tenuta a Tokyo e risulta molto importante per divulgare la conoscenza del settore
0.011217	Ospedaliero	Utilità dei robot collaborativi nel settore sanitario
0.011267	Sicurezza nel cloud	Necessità di miglioramento della sicurezza informatica per proteggere i dati nel cloud

0.011325	Cobot-drone	Utilità delle sensoristiche visive dei bracci robotici collaborativi per migliorare il settore dei droni
0.011361	kitting per customizzare le esigenze	Possibilità di acquistare kit ad hoc per determinati tipi di lavorazione, quindi, aiutare il consumatore nell'acquisto in modo autonomo della tecnologia collaborativa
0.011364	Sensoristica visiva	Necessità di porre il focus evolutivo sulla sensoristica visiva
0.011805	Picking tech pendant	Diffusione nella gestione del picking da flotte robotiche attraverso un semplice touch pad
0.011865	Funzionalità per ambiente magazzino	Utilità dei robot collaborativi nei magazzini aziendali
0.012415	Moralità e legalità sui COBOT	Scetticismi sull'adozione della tecnologia e timore di una possibile distruzione di lavoro
0.01244	Intelligenza artificiale avanzata per determinare sensoristica	Miglioramento della sensoristica dei bracci robotici collaborativi attraverso l'implementazione di algoritmi di intelligenza artificiale
0.012547	Robotics Innovation Awards	Premio riferito a creazioni originali e di grande impatto nel settore della robotica. Riconosciuti in tutto il mondo, i premi RBR50 sono anche una misura fondamentale della crescita del settore della robotica
0.012595	Carico e sensori di forza	Bilanciamento delle coppie di forza in esecuzione in presenza di carichi elevati o variabili
0.013192	Tipologie gripper	Diversificazione delle tipologie del gripper di lavorazione
0.013873	COBOT assemblaggio componenti	Robotica collaborativa utilizzati in modo efficiente nell'operazione di assemblaggio

0.014026	Fondi di finanziamento per la robotica	Presenza di numerosi fondi di finanziamento dedicati alla robotica, simbolo del grado di importanza che viene attribuita a tali tematiche da parte dei governi e dei mercati
0.014157	Centri industriali cobotizzati	Presenza di intere linee di produzione servite e regolate totalmente da robot collaborativi
0.014287	Adattabilità componenti	Importanza dell'adattabilità dei componenti di marchi differenti
0.014383	Trasporto nel magazzino	Impiego della robotica collaborativa molto diffuso nelle operazioni di movimentazione merce in magazzino
0.014636	Prospettiva in crescita del mercato COBOT	Prospettive di mercato in continua crescita grazie i forti investimenti nelle tecnologie collaborative
0.014849	Economicità e diminuzione del costo per unità	Robot collaborativi molto meno costosi rispetto ai robot tradizionali, con un'incidenza significativa sui costi per unità
0.014984	Boom vendite nei paesi asiatici	L'adozione della robotica collaborativa ha visto la maggior crescita nel mercato cinese
0.015051	Emotività nella tecnologia	Robotica collaborativa in grado di riconoscere l'emotività grazie ad algoritmi complessi di intelligenza artificiale
0.015289	ROI elevato e basso payback period	Cambiamento del mercato del lavoro grazie all'utilizzo della robotica collaborativa che garantisce un ROI elevato ed un Payback Period estremamente basso
0.015444	Sponsorship	Sponsorizzazioni presenti sui BLOG analizzati per invogliare il consumatore ad acquistare tecnologia di un determinato marchio
0.015538	Esibizioni e mostre su funzionalità	Necessità di mostrare le capacità dei robot collaborativi ex-ante l'acquisto di essi per

		rassicurare gli utilizzatori circa sulle loro reali potenzialità
0.015751	Velocità necessaria sviluppo robotica	Volontà di velocizzare la diffusione della robotica collaborativa in quanto portatore di maggiori benefici rispetto alla robotica tradizionale
0.015825	Precisione e ripetibilità	Esecuzione delle lavorazioni dei bracci robotici precisa e ripetibile
0.016189	Smart factory	Adozione della robotica collaborativa per completare la transizione alla Smart Factory
0.016442	Logistica del futuro	Previsioni su scenari possibili dell'evoluzione della logistica attraverso l'utilizzo della robotica collaborativa a guida autonoma e non
0.016533	Miglioramento qualità del lavoro	Benefici apportati dall'adozione di un robot collaborativo alla qualità del lavoro
0.017025	Risk Assesment	Valutazione del rischio dei processi a favore dell'utilizzo della robotica collaborativa
0.017244	Carico e velocità	Regolazione della velocità e bilanciamento del carico in esecuzione
0.01782	Difficoltà nel differenziarsi	Omogeneità delle tecnologie presentate sul mercato
0.018067	Sicurezza nel COBOT	Importanza della tematica della sicurezza per gli operatori del settore della robotica
0.018479	Movimentazione-numero assi	Miglioramento prestazionale per i movimenti dei bracci robotici
0.01865	Cambiamento mercato del lavoro	Previsioni su possibili scenari apportati al mercato del lavoro dall'utilizzo della robotica collaborativa
0.019415	Robobusiness	Sponsorizzazioni di Robobusiness nei post dei BLOG analizzati

0.019462	Diffusione concetto di robotica collaborativa di basso livello	Importanza dell'informazione sulle tecnologie collaborative per operatori non esperti di robotica, in modo da far giungere i loro benefici anche nelle PMI
0.019582	Efficientamento processo produttivo	Miglioramento prestazionale dei processi aziendali dopo l'utilizzo dei robot collaborativi
0.01993	Ergonomia	Benefici apportati dai robot collaborativi all'ergonomia dei processi aziendali
0.021244	HRC-cobot	La robotica collaborativa rappresenta la sintesi della collaborazione uomo-macchina
0.022215	Divulgazione scientifica concentrata di alto livello	Importanza dell'informazione sulle tecnologie collaborative per esperti del settore della robotica
0.022658	Settore della robotica collaborativa in rapido aumento	Sviluppo esponenziale del settore della robotica collaborativa
0.023937	Ecosistemi aziendali	Presenza di ecosistemi aziendali che vincolano tecnologie hardware-software e aumentano la competitività del mercato
0.024173	Spinta alla transizione al 5.0	Spinta dei governi alla transizione verso l'industria 5.0

Tabella 8. Labelling dei Topic inerenti alle funzionalità dei bracci robotici collaborativi

Nella Tabella 8 compare la porzione (cioè la prevalenza media Topica) degli argomenti identificati dal data set di documenti divulgativi analizzato. Come precedentemente detto, i Topic sono stati filtrati in base alla coerenza con la tematica di ricerca affrontata. Gli argomenti di cui si è discusso maggiormente riguardano la spinta alla transizione 5.0 da parte dei governi, alla tendenza delle aziende produttrici di robot collaborativi a concentrarsi in ecosistemi tecnologici e al ripido aumento della domanda di mercato per questo settore. Come si può notare, i primi Topic non sono coerenti con quanto richiesto.

Analogamente a quanto fatto per i requisiti funzionali, è stato fatto un ordinamento dei requisiti sulla base del grado di prevalenza Topica associato al requisito. In questo caso, però, in fase di

estrazione dei dati, non è stato possibile estrarre le date di creazione dei documenti selezionati, perciò, non è stato possibile verificare l'andamento delle tendenze evolutive della proporzione tematica nel tempo degli argomenti.

Nella Tabella 9 si riportano i risultati ottenuti.

CONVENIENZA
Economicità
Compatibilità ricambi
OPERATIVITÀ
Aumento numero assi
Aumento ampiezza raggio di lavoro
Stabilità in esecuzione con qualsiasi carico
Anticollisione in esecuzione
Movimentazione morbida del braccio
Velocità operativa
Sensori di coppia di forza (versatilità operativa)
Aumento payload supportabile
Sensibilità operativa
Stabilità in esecuzione in ogni ambiente (waypoint)
Rapidità set-up
Latenza di esecuzione
CUSTOMIZZAZIONE
Diversificazione terminali
Pacchetti personalizzabili per far fronte alle diverse necessità
Customizzare sensoristica visiva
SICUREZZA
Incremento potenzialità sensori con AI
Arresto di emergenza
Laser come barriera di sicurezza
Regolazione velocità di esecuzione in presenza di operatore
SOFTWARE
Semplicità di definizione degli waypoint

Free Drive Mode
Facilità di programmazione movimentazione braccio
Aumentare usabilità tech-pendant
Facilità di attivazione interfaccia client
SERVIZIO
Esposizioni divulgative
Offrire soluzioni applicative complete
Connettività nel cloud
Sicurezza nel cloud
Istruzione programmatore
Istruzione operatore tecnico
Interoperabilità con ambienti open source
Offrire codice modulare
Condivisione di modelli strutturali operativi
Piattaforme ecosistemiche

Tabella 9. Funzionalità dei bracci robotici collaborativi

Per generare la scheda delle funzionalità del prodotto COBOT, riportata nella Tabella 9, è stato necessario filtrare i Topic in base al grado di coerenza che l'etichetta avesse con l'obiettivo dell'analisi in corso. Circa il 60% degli argomenti sono stati scartati, in particolare quelli strettamente collegati a tematiche politiche, previsionali o non strettamente collegate all'ambito manifatturiero.

Dai rimanenti Topic sono poi stati estratti, con le tecniche precedentemente discusse, le funzionalità del prodotto COBOT. Tali funzionalità sono poi state raggruppate nelle cinque categorie precedentemente identificate (Sicurezza, Convenienza, Customizzazione, Operatività, Software, Servizi).

L'algoritmo di clustering in R, in questo caso, è risultato di basso impatto informativo. L'output grafico basato su un grafo delle correlazioni è risultato di difficile comprensione, a causa del numero elevato dei Topic, ma soprattutto non ha permesso di identificare ulteriori macrocategorie di requisiti/funzionalità di prodotto, ma piuttosto i macro-trend del settore che non sono sufficientemente dettagliati e a valore aggiunto per l'attuale analisi.

I Topic scartati, seppur non rilevanti per l'analisi, sono stati utilizzati per comprendere l'attuale scenario di competizione del mercato della robotica collaborativa. La descrizione del contesto di

mercato è stata importante per confermare l'ipotesi di partenza da cui nasce la ricerca: il mercato della robotica collaborativa è in forte crescita ma necessita di differenziazioni di prodotto o servizio. Nei paragrafi 1.1.1 e 1.1.2 si riporta una breve panoramica dello scenario percepito dall'algoritmo STM, integrato con dati più dettagliati di ricerche del settore che ne confermano il contenuto.

4.3.CONFRONTO REQUISITI-FUNZIONALITÀ

In entrambe le analisi è emersa una scheda con circa 50 requisiti/funzionalità.

Avendo due schede ordinate secondo lo stesso criterio, è stato possibile fare un confronto diretto sui requisiti e funzionalità.

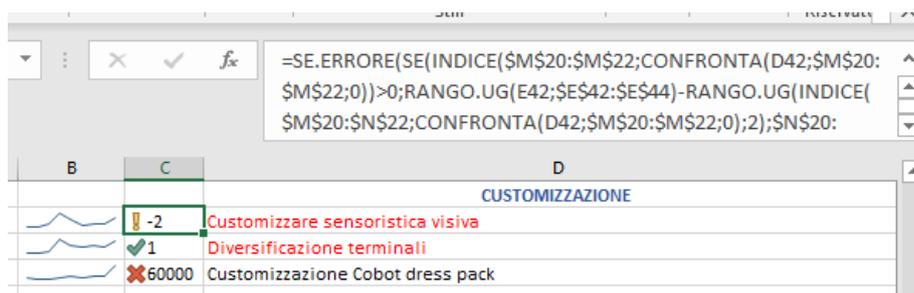


Figura 19. Rappresentazione del layout di analisi della scheda dei requisiti generata

In Figura 19 è riportato il macro-blocco di requisiti chiamato “Customizzazione”. Tale blocco contiene tre tipi di requisiti. Ad ogni requisito è associato nella colonna B la tendenza evolutiva della porzione tematica. Il colore dei requisiti è significativo: rosso, vuol dire che al requisito è associata una funzionalità percepita lato offerta, mentre, nero significa che il requisito non matcha con nessuna funzionalità percepita lato offerta.

Ad ogni requisito, attraverso la funzione riportata nell'immagine è stato possibile fare un confronto in base al grado di prevalenza Topica ad essi associato. Il confronto avviene tra la domanda e l'offerta del prodotto. La formula visibile nell'immagine permette di identificare il matching dei requisiti e valutarne la posizione assoluta all'interno dei rispettivi blocchi lato domanda e lato offerta, restituendo le posizioni di differenza di tale confronto. Nell'immagine si vedono le tre casistiche possibili, rappresentate dalla simbologia nella colonna C:

- “Customizzazione sensoristica visiva” rispetto all’ordine generato nelle funzionalità del prodotto ha un grado di interesse maggiore per gli utenti, rappresentato dal numero 2 che indica le posizioni di differenza da tale ordinamento. È un segnale di allarme per le aziende produttrici che possono porre un focus maggiore su tali requisiti.
- “Diversificazione terminali” ha una spunta verde. Il grado di interesse dell’offerta è maggiore o uguale a quello della domanda. In questo caso, è possibile continuare a mantenere un livello adeguato di tale requisito, senza trascurarlo.
- “Customizzazione Cobot dress pack” non matcha. Il numero 60000 è un upper bound e non ha valore di differenza di posizione. Il requisito è considerato importante ma non viene percepito come un Topic di interesse nell’offerta o, in modo alternativo, non ha riscontri evidenti tale da essere considerato dall’algoritmo STM. Queste tipologie di requisiti possono offrire un vantaggio di differenziazioni se considerati nella strategia aziendale.

Nell’Allegato 7 e nell’Allegato 8 sono riportate rispettivamente la scheda complete dei requisiti e delle funzionalità analizzate.

Complessivamente risulta evidente come la rapidità dello sviluppo tecnologico stia incrementando le potenziali dei bracci robotico con sensoristiche avanzate governate da algoritmi di intelligenza artificiale. Gli utenti del prodotto COBOT percepiscono positivamente tale tendenza tecnologica, ma sembrano evidenziare l’importanza del gestire in modo semplice e veloce gli ambienti ad esso associati: tra queste si evidenziano gli ambienti di sviluppo e progettazione, l’interoperabilità tra ambienti di programmazione e l’interoperabilità dei protocolli di comunicazione con i macchinari già esistenti nel manifatturiero. Emerge quindi la volontà da parte delle aziende di generare ambienti resilienti, robusti industrialmente in modo da poter creare processi adattabili e flessibili. Il lato servizi sembra, invece, essere migliorabile. Dal campione analizzato risulta evidente l’importanza della rapidità dei servizi e l’adattabilità dei software con ambienti open source. Le aziende produttrici del settore sono attualmente racchiuse nel loro ecosistema tecnologico che pone grossi vincoli per i nuovi utenti del prodotto, soprattutto in termini di creazione di conoscenze, competenze e abilità ad-hoc per tali tecnologie.

Inoltre, l’assistenza post-vendita (per esempio per manutenzioni predittive o per servizi di consulenza, ormai sempre più specifici a causa della diversificazione delle problematiche) può comportare dispendi di tempo enormi se non gestita in modo rapido.

Non esistendo attualmente uno standard di settore, il miglioramento di tali servizi, può portare ad un vantaggio competitivo o, almeno, ad un punto di differenziazione dai concorrenti. Considerando l'azienda leader del settore della robotica collaborativa, Universal Robot, è possibile confermare tali assunzioni. L'azienda ha ottenuto un successo enorme negli ultimi anni grazie a due sostanziali modifiche nella fornitura di servizi:

- Compatibilità dei software connessi all'ambiente UR con software open source già utilizzati nell'automazione manifatturiera (ad esempio ROS, Python, Java, RDK...)
- Utilizzo di piattaforme online ben organizzate in cui gli utenti possono condividere esperienze, soluzioni e componenti. Lo scambio di informazioni così strutturato risulta semplice e veloce e rende l'utente auto-sufficiente qualora necessitasse di soluzioni personalizzate in tempi rapidi. In questo modo risulta anche velocizzato il processo divulgativo delle conoscenze di tali tecnologie

Tale caso studio, conferma i risultati dell'analisi svolta.

4.4. CONFRONTO DELLA SOVRAPPOSIZIONE DEI TOPIC

È possibile verificare la sovrapposizione dei requisiti in modo da permettere di comprendere in che modo gli argomenti trattati vadano effettivamente a corrispondere in termini di proporzioni di contenuto.

Come precedentemente detto, esistono delle definite relazioni tra requisiti e Topic, in particolare un Topic può generare più requisiti e un requisito è associato ad un unico Topic. Grazie a tali relazioni è stato possibile seguire gli approcci che vengono successivamente descritti.

Per valutare il grado di sovrapposizione, si è voluto considerare i quattro cluster di requisiti adottati per l'analisi (Sicurezza, Convenienza, Customizzazione, Operatività, Software) e considerarli come degli agglomerati di Topic. L'obiettivo è quello di comprendere in che modo le percezioni degli utenti vanno a sovrapporsi con quelle delle aziende produttrici.

Sono stati adottati tre diversi approcci per valutare tali sovrapposizioni, come riportato di seguito.

4.4.1. Approccio del matching tra requisiti e funzionalità

L'approccio basato sul matching tra requisiti e funzionalità è il più semplice ed immediato da adottare.

La procedura è basata sul rapporto percentuale del matching dei requisiti:

$$\text{Indice matching} = \frac{\text{numero requisiti corrispondenti ad una funzionalità}}{\text{numero requisiti totali}} \quad (4.1)$$

La Formula 4.1 può essere applicata sia per valutare i requisiti corrispondenti alle funzionalità sia per valutare le funzionalità non corrispondenti a dei requisiti. Il punto di vista di questi due modi di vedere l'indice rappresenta i punti di vista rispettivi del mercato e dell'offerta di mercato. Nel primo caso, un alto indice di corrispondenza lato domanda può essere significativo per comprendere in che modo è possibile soddisfare l'utente ed eventualmente costruire delle strategie di mercato ad hoc per soddisfare tali esigenze ed ottenere differenziazione.

Nel secondo caso, un alto indice può significare una buona comprensione da parte dell'offerta di mercato delle esigenze dei clienti.

Esempio riferito al cluster CUSTOMIZZAZIONE:

Requisiti funzionali:

- I. Customizzare sensoristica visiva*
- II. Customizzazione Cobot dress pack*
- III. Diversificazione terminali*

Funzionalità:

- I. Customizzare sensoristica visiva*
- II. Diversificazione terminali*
- III. Pacchetti personalizzabili per far fronte alle diverse necessità*

Riferendoci ai requisiti funzionali si trova che due di essi matchano nel lato delle funzionalità, in particolare "Customizzare sensoristica visiva" e "Diversificazione terminali".

Con tali informazioni, è possibile applicare la Formula 4.1 nel seguente modo:

$$\text{Indice matching} = \frac{2}{3} = 0.67$$

Di seguito si riportano i risultati ottenuti nel caso studio:

<i>CLUSTER</i>	<i>MATCHING LATO DOMANDA</i>	<i>NON MATCHING LATO OFFERTA</i>
CONVENIENZA	0.67	0.00
OPERATIVITA	0.57	0.33
SOFTWARE	0.38	0.00
SICUREZZA	0.75	0.25
CUSTOMIZZAZIONE	0.67	0.33
SERVIZI	0.56	0.50

Tabella 10. Indice di matching dei requisiti (domanda) e funzionalità (offerta)

Dal precedente scenario si possono analizzare tre scenari:

- L'analisi del cluster CONVENIENZA ha un matching soddisfacente dei requisiti. Le funzionalità del prodotto sembrano essere percepite positivamente dal mercato. Tuttavia, essendo una parte dei requisiti scoperta (33% dei requisiti non trovano una corrispondenza nelle funzionalità percepite dal mercato come significative per il prodotto COBOT) sembra esserci della miopia da parte delle aziende produttrici su una fetta di requisiti. Tale porzione di requisiti può essere un buono spunto per offrire differenziazione alle aziende.
- L'analisi del cluster SOFTWARE ha un matching migliorabile dei requisiti; tuttavia, il mercato sembra cogliere solamente una parte dei requisiti. In questo caso, ci sono ampie possibilità di miglioramento da parte delle aziende. La ragione è giustificabile dal fatto che attualmente non esistono standard di mercato nel settore della robotica collaborativa e gli utenti risentono molto della compatibilità tra software diversi.
- Le analisi dei cluster SICUREZZA e CUSTOMIZZAZIONE sono interessanti perché rappresentano percentuali tra loro complementari. In questo caso, è possibile approfondire i requisiti non matching e vedere se le funzionalità non corrispondenti possono in qualche modo incidere su tali requisiti. Nel caso del cluster SICUREZZA, per esempio, il requisito relativo alla sensoristica artificiale può portare significativi miglioramenti a tutti i requisiti relativi a tale cluster. Rappresenta, quindi, un buon allineamento lato domanda e lato offerta.

L'approccio attuale, prescinde dall'analisi dei Topic e risulta essere molto puntuale. I cluster relativi agli argomenti di discussione vengono discretizzati in un numero di requisiti e funzioni. Se esiste un matching allora l'argomento è in parte corrisposto altrimenti è da considerare non soddisfatto. Tale analisi evidentemente può risultare approssimativa. Gli argomenti trattati sono da percepire come ricchi di sfumature concettuali, e sotto tale ipotesi si basano i seguenti approcci proposti.

4.4.2. Approccio con indice di Jaccard

L'indice Jaccard, noto anche come Intersection over Union e coefficiente di somiglianza Jaccard, è una statistica utilizzata per misurare la somiglianza e la diversità degli insiemi di campioni. Si definisce come il rapporto tra l'intersezione di due insiemi finiti e la loro unione.

Dati due insiemi A e B, di cui sono noti gli elementi al loro interno, la formula per il calcolo dell'indice è la seguente:

$$J(A, B) \stackrel{\text{def}}{=} \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (4.2)$$

L'indice è compreso tra i valori 0 e 1; tanto maggiore è il suo valore, tanto più i due insiemi risultano tra loro sovrapposti.

In modo inverso, è possibile calcolare la distanza di Jaccard per calcolare il grado di dissomiglianza tra i due insiemi.

Dati due insiemi A e B, di cui sono noti gli elementi al loro interno, la formula per il calcolo dell'indice è la seguente:

$$D_J(A, B) \stackrel{\text{def}}{=} 1 - \frac{|A \cap B|}{|A \cup B|} = 1 - \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (4.3)$$

Le implicazioni della Formula 4.2 e della Formula 4.3 sono considerevoli in ogni ambito e permettono di rendere gli algoritmi capaci di comprendere le differenze tra due insiemi distinti. Per tale ragione, tale indice, nelle sue varie forme, viene utilizzato negli algoritmi di intelligenza artificiale, in particolare per quello che riguarda l'apprendimento automatico.

Nel contesto in analisi, l'indice di Jaccard può essere utile per capire il grado di sovrapposizione tra requisiti funzionali e funzionalità.

Il confronto può essere fatto tra cluster di requisiti tra loro corrispondenti dal punto di vista della domanda e dell'offerta. In questo modo, ci si può aspettare di valutare in che modo le argomentazioni degli utilizzatori dei robot collaborativi vengano percepite nel loro insieme dalle aziende produttrici del settore. Per la riuscita di tale operazione è necessario avere a disposizione:

- un dataset di UGC e un dataset divulgativo, in modo da riuscire a disaccoppiare requisiti funzionali e funzionalità
- i Topic simboleggiano al meglio ogni cluster di requisiti; in particolare sapendo che ogni requisito è referenziato da un solo Topic, sarà possibile estrarre i Topic referenziati dai requisiti di un singolo cluster
- il set di documenti più pertinenti per ogni Topic, estratto attraverso una percentuale minima di prevalenza Topica o un numero fisso di documenti, in modo da generare insiemi equipotenti. Per definizione, due insiemi sono equipotenti se presentano lo stesso numero di elementi, non necessariamente uguali tra loro

Nella Tabella 11 si riporta un esempio di una parte del dataset preparato per essere manipolato correttamente dal codice riportato nell'Allegato 2; per semplicità i documenti relativi alle funzionalità sono stati riportati parzialmente. Si riportano due insiemi di documenti equipotenti in base al numero fisso di documenti riportati in ognuno (cinque documenti per ogni insieme).

Covariato	Documento
<i>Funzionalità</i>	
0,98151	A challenge of collaborative functions in industrial automation is making sure that a person can enter a work cell safely with a minimum of reduced function...
0,98027	One of the selling points of collaborative robots is that they are safer to operate around people than conventional industrial automation...
0,89001	SICK and Universal Robots (UR) announced a new safety solution for the safeguarding of accessible cobot applications for the URe Series of cobots...
0,88573	When combined, the sBot safety system and URcap from SICK increase productivity with immediate automated restart and profit from an easy configuration of the safety system directly from the UR teach pendant...

0,84925	The newest ultra-compact safety laser scanner, the nanoScan3. The nanoScan3 combines smart safety functions with measurement data quality for accurate and reliable localization...
<i>Requisiti funzionali</i>	
0,82954	There is no way to smoothen the transition when transitioning between Normal and Reduced mode and vice versa. During certification of the Safety system, it was a demand from the notified certification body, that the transition must be completed within a given timeframe in order to conform with the safety standards.
0,77086	When an emergency stop occurs during a program, this means the emergency stop button has been pressed. It can also come from an I/O you previously set up to perform an emergency stop. In any case, when an emergency stop occurs, someone must go and assess what happened to trigger the emergency stop. An unwanted and potentially dangerous situation was happening, thus requiring an operator to come in order to restart the robot application.
0,76174	If you want to dynamically change speed of robot depending on distance between robot and operator, yes then a safety scanner or safety mat can be used. If you just want to pause the robot when an operator is approaching, a light barrier can be used to trigger safeguard. This is often the more affordable option.
0,69542	ok so I'm basically setting up a safety configuration and adding a safety plane where when it crosses it turns the robot into reduced mode.
0,67536	Hallo, we like to limit the speed of a fanuc robot with a safety scanner. we have the option dcs. is there another possibility than reduce the override and monitoring the speed with dcs after a time?thank you for your responsechris

Tabella 11. Esempio di una porzione del dataset con due insiemi equipotenti di documenti per il calcolo dell'Indice di Jaccard, riferito al caso studio della robotica collaborativa, in particolare, a due Topic del cluster Sicurezza per i requisiti funzionali e le funzionalità di prodotto

Si consideri il seguente esempio di calcolo riferito alla Tabella 11:

Esempio

Dati i due insiemi Requisiti funzionali e funzionalità d prodotto, identifichiamo tutte le parole componenti tali insiemi (per semplicità si riportano solamente dieci parole, come mostrato nella Tabella 12). Inoltre, si suppone di considerare la presenza dei dati come binaria (1 c'è almeno una corrispondenza, 0 altrimenti, come riportato nella Tabella 12).

<i>Funzionalità</i>	1	1	1	0	1	0	0	1	1	1
<i>Requisiti funzionali</i>	0	1	0	1	0	1	1	1	1	1
<i>Keywords</i>	<i>Collaborative</i>	<i>Safety</i>	<i>Laser</i>	<i>Mode</i>	<i>Solution</i>	<i>You</i>	<i>Set</i>	<i>The</i>	<i>And</i>	<i>On</i>

Tabella 12. Esempio di calcolo del coefficiente di Jaccard relativo ai documenti riportati in Tabella 10

Dati due insiemi A e B, la formula per dati binari è possibile definirla come segue (Orlando, 2009):

$$J(A, B) = \frac{a}{a + b + c} \quad (4.4)$$

dove:

- *a* rappresenta il numero totale di attributi dove A e B hanno entrambi valore 1.
- *b* rappresenta il numero totale di attributi dove l'attributo di A è 0 e l'attributo di B è 1.
- *c* rappresenta il numero totale di attributi dove l'attributo di A è 1 e l'attributo di B è 0.

Applicando la Formula 4.4 dell'Indice di Jaccard per dati binari e utilizzando i dati nella Tabella 12 si ottiene:

$$J(A, B) = \frac{4}{4 + 3 + 3} = 0,4$$

Per procedere con l'analisi è stato utilizzato l'ambiente R. È stato generato un codice che permettesse di prendere in input i dataset dei documenti più pertinenti dei Topic selezionati divisi per cluster e per requisito/funzionalità.

Il codice è stato generato partendo dalla definizione matematica dell'indice di Jaccard, senza l'utilizzo di librerie aggiuntive in R. La ragione di tale scelta deriva dalla volontà di generare un codice estremamente compatto e di facile intuizione, in modo che potesse essere comprensibile ad ogni tipo di utente. L'utilizzo di tale codice, essendo basato sulla definizione dell'indice, è applicabile in ogni contesto e permette di valutare la sovrapposizione tra due insiemi contenenti documenti testuali.

Nell'allegato 2 si riporta il codice utilizzato.

Dopo aver preparato la documentazione in modo processabile dall'algoritmo, è stato utilizzato il codice generato per l'analisi. I codici per la predisposizione dell'ambiente di lavoro sono riportati nell'allegato 3.

Dall'analisi è emerso che in ogni cluster considerato non vi fosse alcun tipo di analogia concettuale, riportando l'indice di Jaccard ad un valore approssimabile allo 0.

L'algoritmo, seppur sofisticato, non ha rilevato alcun tipo di sovrapposizione tra i requisiti. La ragione di tale risultato è da ricercare alla differenza tra le basi di dati analizzate, sia nei contenuti sia nella loro forma. Il disaccoppiamento dei requisiti dalle funzionalità richiede l'analisi di due ambienti di lavoro ben distinti soprattutto nella forma.

Nel dataset con gli UGC la documentazione proviene da forum, dunque, i documenti testuali sono commenti contenenti il formalismo dei forum approssimabile in tre fasi: saluto, corpo di testo e ringraziamenti. In tali commenti c'è un linguaggio molto specifico, talvolta contenente frammenti di codici operativi scritti. Il mix tra linguaggio naturale e linguaggi di programmazione si discosta totalmente dal formalismo divulgativo e promozionale del secondo dataset.

A testimonianza di quanto detto, confrontando i due database, l'indice di Jaccard risultante è approssimabile allo zero (per l'esattezza $1 \cdot 10^{-5}$).

In questo modo, la tecnica dell'utilizzo dell'indice di Jaccard risulta inappropriata a tale contesto.

4.4.3. Approccio basato sulla frequenza delle keywords

L'approccio si basa sulla frequenza delle keywords significative per ogni requisito. L'idea di fondo di tale metodo è che se una parola si presenta in quantità proporzionali nella scheda dei requisiti e nella scheda delle funzionalità, allora la sovrapposizione sarà massima.

Per la riuscita di tale operazione è necessario avere a disposizione:

- un dataset di UGC e un dataset divulgativo, in modo da riuscire a disaccoppiare requisiti funzionali e funzionalità
- i Topic simboleggiano al meglio ogni cluster di requisiti; in particolare sapendo che ogni requisito è referenziato da un solo Topic, sarà possibile estrarre i Topic referenziati dai requisiti di un singolo cluster
- il set di documenti più pertinenti per ogni Topic, estratto attraverso una percentuale minima di prevalenza Topica o un numero fisso di documenti in modo da generare insiemi

equipotenti. Per definizione, due insiemi sono equipotenti se presentano lo stesso numero di elementi, non necessariamente uguali tra loro

Il metodo può essere sintetizzato nei seguenti passi:

1. Definire un contatore di keywords per ogni set di documenti riferiti ad un determinato Topic
2. Calcolare la frequenza delle keywords in analisi sia nella scheda dei requisiti funzionali sia nella scheda delle funzionalità di prodotto
3. Definire un ordinamento delle keywords, in base alla loro frequenza
4. Verificare quanto l'ordinamento sia simile in termini di posizione e di rapporto percentuale delle frequenze attraverso la Formula 4.5:

$$\text{Indice frequenza keywords} = \frac{\text{numero keywords validate}}{\text{totale keywords analizzate}} \quad (4.5)$$

In questo modo, è possibile vincolare l'analisi sulle parole più pertinenti per ciascun Topic. La metodologia proposta è molto onerosa in termini di tempo in quanto prevede una laboriosa manipolazione dei dati al fine di poterli analizzare in modo corretto. L'ambiente utilizzato è Excel. I codici utilizzati vengono riportati nell'allegato 3.

L'obiettivo di tale analisi è intersecare tra di loro i Topic di ogni cluster di argomenti.

<i>TOPIC LATO DOMANDA</i>	<i>TOPIC LATO OFFERTA</i>
Wrist Camera	Tipologie gripper
Diversificazione terminali	kitting per customizzare le esigenze
Cobot dress pack	Wrist camera

Tabella 13. Esempio dell'ordinamento dei Topic referenziati da requisiti e funzionalità all'interno del

Nella Tabella 13 vengono riportati i Topic del cluster CUSTOMIZZAZIONE, referenziati dai requisiti e dalle funzionalità, ordinati secondo la massima prevalenza Topica; con tale informazione è possibile verificare in che modo avvenga la sovrapposizione dei contenuti.

Il metodo finora adottato è esattamente quello applicato per la preparazione dei vari dataset per il calcolo dell'indice di Jaccard. La novità di questo approccio è finalizzata a distaccare i formalismi semantici dei dataset originali e sincronizzare l'analisi ai requisiti trovati.

Per far questo, sono state selezionate le keywords significative per ogni Topic relativo ai requisiti di prodotto. Ogni insieme di keyword non necessariamente è indipendente, dunque occorre lavorare con l'unione degli insiemi considerati, rimuovendo cioè le intersezioni. Definendo un contatore è poi possibile valutare la frequenza delle keywords sui documenti referenziati da ogni Topic, definendo poi due scale d'ordine (lato domanda e offerta).

In questo modo, sono state confrontate le scale mantenendo una tolleranza posizionale del 10% circa. La ragione di tale numero è da ricercare nella diversità dei campioni analizzati, sia per quel che riguarda il contenuto informativo sia per il numero di referenze Topiche associate. Ad ogni modo, la percentuale rappresentata simboleggia il livello di tolleranza che poniamo all'analisi: tolleranze troppo basse rischiano di non considerare le fonti di diversità dei dataset d'origini, d'altro canto tolleranze troppo alte non rendono significativa l'analisi.

Si veda il seguente esempio semplificato del caso studio.

Esempio riferito al cluster CUSTOMIZZAZIONE:

Topic associati ai requisiti funzionali:

- I. Cobot dress pack*
- II. Diversificazione terminali*
- III. Wrist Camera*

Topic associati alle funzionalità:

- I. kitting per customizzare le esigenze*
- II. Tipologie gripper*
- III. Wrist camera*

In riferimento ai precedenti Topic, per semplificare la trattazione, vengono riportate solamente le dieci keyword selezionate dal criterio Highest Probability:

Keywords relative ai Topic dei requisiti funzionali:

- I. Cobot dress pack: feb, back, anoth, insid, null, noth, replac, put, timeout, mechan*
- II. Diversificazione terminali: gripper, robotiq, finger, coupl, adapt, univers, use, sensor, kit, plate*
- III. Wrist Camera: part, camera, use, place, pick, object, locat, detect, vision, calibr*

Keywords relative ai Topic delle funzionalità:

- I. kitting per customizzare le esigenze: pallet, packag, product, robot, box, line, pack, oper, handl, task*

II. *Tipologie gripper: gripper, robotiq, forc, robot, hand, finger, use, new, adapt, sensor*

III. *Wrist camera: camera, vision, system, object, part, pick, use, imag, wrist, need*

Usando l'unione delle keywords, rimuovendo eventuali duplicati nei due insiemi requisiti-funzionalità, è possibile generare i seguenti insiemi:

I. *Keywords requisiti funzionali: feb, back, anoth, insid, null, noth, replac, put, timeout, mechan, gripper, robotiq, finger, coupl,, adapt, univers, use, sensor, kit, plate, part, camera, use, place, pick, object, locat, detect, vision, calibr*

II. *Keywords funzionalità: pallet, packag, product, robot, box, line, pack, oper, handl, task, gripper, robotiq, forc, robot, hand, finger, use, new, adapt, sensor, camera, vision, system, object, part, pick, use, imag, wrist, need*

Occorre quindi selezionare per ciascun Topic un campione di argomenti associati più pertinenti, per esempio selezionando i primi 100 documenti (commenti dei forum o post dei blog) in base al maggior grado di coerenza al Topic associato.

Avendo a disposizione le keywords dei requisiti funzionali, è possibile identificare un ordine in base alla frequenza di tali parole nel campione di documenti selezionato generando un contatore, come riportato nell'Allegato 3.

Qualora vi fosse un interesse nel porre maggiore focus sulle funzionalità di prodotto percepite dal mercato, è possibile svolgere l'analisi nel seguente modo ma considerando il set di keywords relative alle funzionalità di prodotto.

Nella Tabella 14 vengono riportate le frequenze calcolate.

<i>Keywords</i>	<i>Frequenza Keyword nei documenti dei requisiti funzionali</i>	<i>Frequenza Keyword nei documenti delle funzionalità</i>
<i>part</i>	<i>1011</i>	<i>293</i>
<i>camera</i>	<i>492</i>	<i>195</i>
<i>use</i>	<i>1155</i>	<i>209</i>
<i>place</i>	<i>355</i>	<i>67</i>
<i>pick</i>	<i>472</i>	<i>98</i>
<i>object</i>	<i>331</i>	<i>95</i>
<i>locat</i>	<i>106</i>	<i>82</i>

<i>detect</i>	113	136
<i>vision</i>	541	58
<i>calibr</i>	0	0
<i>gripper</i>	954	310
<i>robotiq</i>	363	168
<i>finger</i>	204	94
<i>coupl</i>	17	96
<i>adapt</i>	181	46
<i>univers</i>	659	34
<i>sensor</i>	340	59
<i>kit</i>	128	54
<i>plate</i>	4	31
<i>anoth</i>	126	45
<i>back</i>	94	85
<i>insid</i>	56	50
<i>put</i>	68	35
<i>noth</i>	5	8
<i>replac</i>	50	39
<i>feb</i>	4	4
<i>null</i>	0	72
<i>timeout</i>	0	34
<i>mechan</i>	0	0

Tabella 14. Frequenze assolute delle keywords calcolate nei set di documenti inerenti ai requisiti funzionali e alle funzionalità

Attraverso la definizione delle frequenze assolute, è possibile creare un ordinamento delle keywords. La differenza posizionale delle keywords nei due insiemi è corretta da una tolleranza del 10% rispetto al totale delle keywords, dunque:

$$Tolleranza\ posizionale = \text{numero keywords} * 10\% = 30 * 0,1 = 3$$

Keywords che differiscono per meno di tre posizioni sono considerate valide per l'analisi, dunque, 8 keywords come riportato nella Tabella 15.

<i>Keywords</i>	<i>Ordinamento secondo il set di documenti dei requisiti funzionali</i>	<i>Ordinamento secondo il set di documenti delle funzionalità</i>	<i>Differenza posizionale</i>	<i>Keywords rilevanti</i>
<i>part</i>	2	2	293	1
<i>camera</i>	6	4	195	1
<i>use</i>	1	3	209	1
<i>place</i>	9	14	67	0
<i>pick</i>	7	7	98	1
<i>object</i>	11	9	95	1
<i>locat</i>	23	12	82	0
<i>detect</i>	22	6	136	0
<i>vision</i>	5	16	58	0
<i>calibr</i>	45	28	0	0
<i>gripper</i>	3	1	310	1
<i>robotiq</i>	8	5	168	0
<i>finger</i>	15	10	94	0
<i>coupl</i>	40	8	96	0
<i>adapt</i>	17	19	46	1
<i>univers</i>	4	23	34	0
<i>sensor</i>	10	15	59	0
<i>kit</i>	20	17	54	0
<i>plate</i>	42	25	31	0
<i>anoth</i>	21	20	45	1
<i>back</i>	27	11	85	0
<i>insid</i>	34	18	50	0
<i>put</i>	30	22	35	0
<i>noth</i>	41	26	8	0
<i>replac</i>	36	21	39	0
<i>feb</i>	42	27	4	0
<i>null</i>	48	13	72	0
<i>timeout</i>	45	23	34	0

<i>mechan</i>	45	28	0	0
---------------	----	----	---	---

Tabella 15. Ordinamento delle keywords in base alle relative frequenze assolute calcolate nei set di documenti inerenti ai requisiti funzionali e alle funzionalità

Applicando la Formula 4.5 è possibile identificare l'Indice di frequenza:

$$\text{Indice frequenza keywords} = \frac{8}{30} = 0,27$$

La sovrapposizione dei requisiti e delle funzionalità risulta, secondo questo metodo, del 27%.

Dall'analisi del caso studio emergono i seguenti risultati:

<i>CLUSTER</i>	<i>Indice frequenza keywords</i>
CONVENIENZA	0.63
OPERATIVITA	0.37
SOFTWARE	0.59
SICUREZZA	0.85
CUSTOMIZZAZIONE	0.32
SERVIZI	0.46

Tabella 16. Indice di frequenza delle keyword dei macro-cluster

Dalla Tabella 16 emerge una sovrapposizione maggiore della tematica della sicurezza, uno dei tratti caratterizzanti del prodotto COBOT.

Il tema della customizzazione e dei requisiti operativi sembrano, invece, non essere perfettamente sovrapposti. La motivazione può essere ricercata sia nelle diversità del dataset, sia nel prodotto stesso. Nel primo caso i tecnicismi degli utenti nei loro commenti non vengono percepiti nel dataset divulgativo, nel secondo caso, le applicazioni dei COBOT riguardano ambiti estremamente variegati. In quest'ultimo caso occorre, infatti, considerare che le esigenze di mercato sono estremamente variegati e danno origine ad una molteplice customizzazione delle problematiche che possono emergere nelle fasi operative; da questo punto di vista una documentazione divulgativa non può entrare nel dettaglio di ogni problematica.

La soluzione per far fronte a tali considerazioni può essere quella di ampliare i margini di tolleranza per l'analisi di tali cluster di requisiti funzionali, con la conseguente creazione di un risultato non omogeneo che può deviare l'interpretazione dell'analisi da parte dell'analista.

4.4.4. Confronto tra metodi

Mettendo a confronto i risultati dei due approcci, emergono i seguenti risultati, riportati nella Tabella 17.

	<i>MATCHING LATO DOMANDA</i>	<i>NON MATCHING LATO OFFERTA</i>	<i>Indice frequenza keywords</i>
CONVENIENZA	0.67	0.00	0.63
OPERATIVITA	0.57	0.33	0.37
SOFTWARE	0.38	0.00	0.59
SICUREZZA	0.75	0.25	0.85
CUSTOMIZZAZIONE	0.67	0.33	0.32
SERVIZI	0.56	0.50	0.46

Tabella 17. Confronto tra indice di matching e frequenza delle keywords

I due metodi proposti possono essere interpretati in modo complementare.

La discretizzazione degli argomenti trattati nei cluster del primo metodo risponde sul grado di efficacia della tecnologia percepita tra mercato e offerta. L'approccio con le keywords mette invece in risalto la comprensione dei vari cluster dei requisiti, dunque, il modo in cui vengono percepiti lato domanda e offerta.

Nella Tabella 17 si riporta un confronto dei risultati ottenuti con l'applicazione dei due metodi.

Contestualizzando i risultati a quanto detto nei precedenti paragrafi, il tema della sicurezza sembra quello maggiormente corrisposto e ben implementato dalle aziende.

Successivamente l'economicità e la durabilità del prodotto sono altri fattori chiave che vengono percepiti in modo simile da entrambi i punti di vista.

Il tema dei servizi sembra potenzialmente migliorabile e a testimonianza di ciò, come detto nei precedenti capitoli, il leader del settore Universal Robot sta rivoluzionando i servizi dei robot collaborativi ottenendo sempre più successo nel mercato.

Infine, interessante è l'analisi dei Topic legati ai software. Seppure il prodotto non risulti pienamente soddisfacente dal punto di vista del software, a causa dell'assenza di uno standard che porta a incompatibilità tra ambienti di lavoro, la corrispondenza tra le tematiche toccate nei Topic risulta sufficientemente sovrapposta. La ragione di tale aspetto sembra far emergere una mancanza di efficacia da parte delle aziende nel proporre soluzioni a tematiche percepite in modo sufficientemente simile al mercato e la causa può essere ricercata nell'alta concorrenza del settore a causa di strategie di lock-in dovuto proprio a incompatibilità con le varie tecnologie del settore. Anche in questo caso, riferendoci al leader del settore Universal Robot, è possibile notare l'apertura delle tecnologie UR ad ambienti open source in modo da poter garantire una migliore efficienza a livello software e permettere all'utente di poter usare competenze e abilità già sviluppate precedentemente (si pensi ad esempio all'importanza di poter riutilizzare codici modulari, usare ambienti di simulazione già conosciute dall'utente, etc...).

5. SVILUPPI FUTURI

Le metodologie spiegate in questo documento richiedono uno sforzo computazionale supportato da un calcolatore con prestazioni medie, ma richiedono un dispendio di tempo importante per l'analista, in particolare per quel che riguarda la generazione dei requisiti. Comprendere i Topic ed estrarne il loro contenuto informativo collegato alle keywords è un processo che richiede una comprensione avanzata della tematica affrontata e può essere un grosso limite di tale approccio.

L'obiettivo di questo capitolo è proporre delle metodologie generalizzabili, dunque che prescindono dal contesto di applicazione, ma automatizzabili. L'automatizzazione richiederà sempre da parte dell'analista una parziale comprensione della tematica in esame, ma dovrà porre le sue valutazioni sulla scelta e comprensione di requisiti funzionali già presenti come output di un modello. L'approccio che si è voluto seguire segue procedure informatico-matematiche implementabili in adatti software open source (ad esempio R, Python...).

5.1. APPROCCIO CON MARKOV CHAIN

L'attuale metodologia proposta proviene dall'ambito letterario, in particolare per la generazione di testi a senso compiuto partendo da un data set iniziale. L'approccio si basa sul concetto matematico elaborato da Andrej Andreevič Markov, noto come *le catene di Markov*.

I processi di Markov sono tali per cui l'informazione ottenibile dall'osservazione della storia del processo fino al presente può essere utile a fare inferenza sullo stato futuro, dunque, un processo è Markoviano se, conoscendo (esattamente) lo stato presente, gli stati passati e futuri sono prevedibili.

In termini matematici, un processo di Markov a tempi T finiti e stati E discreti gode della seguente proprietà (Alfieri, 2021):

$$\begin{aligned}
 &\text{un processo } \{X_{t_i}\} \text{ } i = 0, \dots, n \text{ con } t_0 < t_1 < \dots < t_n \\
 &\text{è detto di Markov (rispetto a } P(\cdot | I)) \text{ se,} \\
 &\text{per ogni } k \in \{1, \dots, n - 1\}, e_0, e_1, \dots, e_{k+1} \in E, \\
 &\text{vale } P(X_{t_{k+1}} = e_{k+1} | I \cap \{X_{t_k} = e_k\} \cap X_{t_{k-1}} = e_{k-1} \cap \dots \cap \{X_{t_0} = e_0\}) = \\
 &= P(X_{t_{k+1}} = e_{k+1} | I \cap \{X_{t_k} = e_k\})
 \end{aligned}$$

L'utilità delle catene di Markov è attualmente molto diffusa in ambito ingegneristico e linguistico; in particolare, con la sua definizione è possibile generare un algoritmo che prevedere attraverso l'utilizzo di variabili stocastiche la parola più strettamente correlata ad una data parola in input. In questo modo, l'algoritmo va a definire una matrice delle correlazioni per ogni keyword associata ad un relativo Topic. Tale matrice viene poi associata alla *Step transition matrix*, dunque alla matrice che permette di definire i cammini percorsi futuri in base agli stadi passati.

Considerando l'output del modello in R a disposizione, si riporta la procedura algoritmica da seguire elaborata nel lavoro di tesi:

Passi da seguire	Esempio applicativo
<p>1. <i>Seleziona i documenti per ogni Topic</i></p> <p>Si può adottare un valore soglia del grado di coerenza di tale argomento con il Topic oppure definire un numero costante di documenti da considerare per ogni Topic. La differenza risulta nell'avere rispettivamente data set di lunghezza variabile o costante.</p>	<p>Seleziono i primi 100 commenti del Topic "Sicurezza" ordinati per prevalenza Topica, secondo quanto detto a pagina 46 di questo documento.</p>
<p>2. <i>Seleziona il numero di keywords descrittive per ogni Topic</i></p> <p>Il numero di keyword descrittive può essere modificato come parametro</p>	<p>Scelgo in modo arbitrario e non definitivo 5 keywords da estrarre su questo Topic.</p>

dell'algoritmo in R e deve essere scelto in modo razionale dall'analista

3. *Crea un numero di dataset pari al numero di Topic a disposizione*

Occorre associare ad ogni data set il corrispettivo insieme di documenti e parole in modo da permettere il corretto funzionamento dell'algoritmo. Tale fase può non essere rispettata qualora l'analista si accorga di Topic estremamente sconnessi dall'obiettivo di ricerca; in questo modo si può risparmiare tempo computazionale.

Automatizzo i primi due punti per ogni Topic a disposizione che considero rilevante per l'analisi.

4. *Pre-processamento dei dati*

È possibile rimuovere parole superflue che possono rendere l'analisi non robusta in termini di risultati ottenuti.

Identifico le seguenti keywords "*robot, nbsp, stop, speed, safeti*".

Decido di migliorare l'estrazione delle keywords eliminando parole non troppo pertinenti o ridondanti quali "*nbsp, robot*".

Rifaccio l'estrazione ed estraggo "*light, signal, stop, speed, safeti*".

5. *Fit del modello*

Attraverso l'uso di funzioni predefinite per l'applicazione dell'approccio delle Markov chain, è possibile prendere i documenti come sequenze di parole e

L'algoritmo genera delle matrici di probabilità (steps matrix) in riferimento ad ogni coppia/terna/... di parole.

valutarne la probabilità di generazione di una parola data quelle adiacenti. In questa fase è utile valutare il grado di precisione dell'approssimazione, in particolare è possibile costruire le *steps matrix* riferite ad 1, 2, 3...parole precedenti alla parola corrente. Aumentando tale numero aumenta la precisione dell'algoritmo ma anche il tempo di calcolo computazionale.

6. *Scegliere i parametri per generare il set di frasi*

Questa fase è il core per la generazione dei requisiti funzionali. È possibile precisare il numero di parole da concatenare e la parola da cui iniziare tale generazione. Tale fase deve essere costruita in modo iterativo al fine di poter generare un set di frasi che nascono dal set di keywords salvate per ogni Topic. Qualora sia ritenuto necessario, è possibile modificare il numero di frasi generate da ogni keyword.

7. *Controllo e validazione dei risultati*

Le frasi generate devono essere controllate per verificare il grado di pertinenza e correttezza logica. Questa fase deve essere affrontata a discrezione dell'analista e dipende dai risultati ottenuti fino a quel momento.

Utilizzo l'algoritmo per generare, ad esempio, coppie di parole. L'algoritmo basandosi sulle matrici di probabilità di ogni parola, può generare un output come segue:

"safeti light"
"safeti stop"
"safeti stop"
"light signal"

È possibile giudicare le coppie di parole in output in base alla loro facilità di comprensione. In questo caso si nota facilmente che queste parole possono generare i seguenti requisiti:

"Arresto di emergenza"
"Luci di monitoraggio stato robot"
"Laser come barriera di sicurezza".

Se questa fase non viene superata occorre ripartire dal punto 4.

Aumentando il numero di keywords (al passo 2) l'analisi dei requisiti risulta più approfondita e dettagliata.

Con questa metodologia è possibile generare requisiti attraverso l'utilizzo di variabili stocastiche, in questo modo l'output dell'algoritmo prescinde dal concatenamento grammaticale dei documenti in input.

Questo permette a tale algoritmo di generare un contenuto informativo preciso anche quando due parole distano tra loro di più posizioni, aumentando quindi l'affidabilità dell'output.

Dall'esempio, emerge in modo significativo come, a partire da un dataset iniziale, sia possibile creare un set di frasi in modo autonomo basandoci sulle correlazioni di esse nel dataset iniziale. Tale procedura è applicabile al contesto in analisi e richiede come precedentemente detto una pulizia accurata dei dati in modo da rendere significative per l'analisi le parole più correlate alle keywords.

L'output permette al progettista di avere un documento con più requisiti per ogni Topic che dovranno poi essere manualmente verificati, validati e scritti in maniera adatta al contesto settoriale, con una terminologia adeguata, ben definita e comprensibile facilmente da un progettista esterno.

Nell'Allegato 5 è possibile approfondire tale metodologia con un esempio operativo e con il relativo codice in R.

5.2. APPROCCIO CON REGULAR EXPRESSION

In informatica, un'espressione regolare è una sequenza di caratteri che forma un modello di ricerca, principalmente per l'uso nella corrispondenza del modello con le stringhe. I modelli sono spesso una combinazione di abbreviazioni di testo e metacaratteri. Le espressioni regolari vengono utilizzate per cercare oggetti, eseguire estrazioni o operazioni di ricerca/sostituzione.

Questo approccio risulta più semplice rispetto al precedente. Il modello si basa sull'utilizzo della correlazione tra parole e sulla frequenza nei documenti di tali espressioni regolari generate da tali

correlazioni. Il risultato è deterministico in quanto basato sulla maggiore correlazione tra una parola con l'intero vocabolario disponibile.

Considerando l'output del modello in R a disposizione, si riporta la procedura algoritmica da seguire:

Passi da seguire	Esempio applicativo
<p>1. <i>Seleziona i documenti per ogni Topic</i> Si può adottare un valore soglia del grado di coerenza di tale argomento con il Topic oppure definire un numero costante di documenti da considerare per ogni Topic. La differenza risulta nell'averne rispettivamente data set di lunghezza variabile o costante.</p>	<p>Seleziono i primi 100 commenti del Topic "Sicurezza" ordinati per prevalenza Topica, secondo quanto detto a pagina 46 di questo documento.</p>
<p>2. <i>Seleziona il numero di keywords descrittive per ogni Topic</i> Il numero di keyword descrittive può essere modificato come parametro dell'algoritmo in R e deve essere scelto in modo razionale dall'analista</p>	<p>Scelgo in modo arbitrario e non definitivo 5 keywords da estrarre su questo Topic.</p>
<p>3. <i>Crea un numero di dataset pari al numero di Topic a disposizione</i> Occorre associare ad ogni data set il corrispettivo insieme di documenti e parole in modo da permettere il corretto funzionamento dell'algoritmo. Tale fase può non essere rispettata qualora l'analista si accorga di Topic estremamente sconnessi dall'obiettivo</p>	<p>Automatizzo i primi due punti per ogni Topic a disposizione che considero rilevante per l'analisi.</p>

di ricerca; in questo modo si può risparmiare tempo computazionale.

4. *Pre-processamento dei dati*

È possibile rimuovere parole superflue che possono rendere l'analisi non robusta in termini di risultati ottenuti.

Identifico le seguenti keywords "*robot, nbsp, stop, speed, safeti*".

Decido di migliorare l'estrazione delle keywords eliminando parole non troppo pertinenti o ridondanti quali "*nbsp, robot*".

Rifaccio l'estrazione ed estraggo "*light, signal, stop, speed, safeti*".

5. *Preparazione delle frasi regolari*

Occorre determinare un set di espressioni regolari che andranno poi ad essere la base per il successivo passo. Tali combinazioni si possono generare dalla combinazione di più keywords di un dato Topic. Utilizzando un iteratore. Il consiglio è di utilizzare coppie di parole in modo da non complicare la fase successiva.

Date le keywords identificate genero

preferibilmente coppie di parole, possibilmente generando ogni combinazione possibile in modo da rendere l'analisi il più precisa possibile.

Ottingo le seguenti coppie, per semplicità non vengono riportate tutte le 25 combinazioni:

"safeti light"

"safeti stop"

"light signal"

"light stop"

"light safeti"

"speed stop"

"speed safeti"

"safeti speed"

Etc...

6. *Utilizzo della ricerca per Regular Expressions*

Occorre prendere le coppie di parole generate nella precedente fase e

Identificando ogni coppia di parole con la corrispettiva nel testo attraverso il metodo *gregex*, calcolo la frequenza assoluta della presenza delle parole nei vari documenti:

<p>identificarle per corrispondenza approssimativa all'interno del set di documenti predisposti per il dato Topic.</p> <p>Le espressioni regolari trovate rappresentano i requisiti. La sintassi sarà sintetica e basata sulla semplice frequenza delle volte in cui tali parole compaiono vicine.</p>	<p>"safeti light" - 10</p> <p>"safeti stop" - 8</p> <p>"light signal" - 5</p> <p>"light stop" - 0</p> <p>"light safeti" - 0</p> <p>"speed stop" - 0</p> <p>"speed safeti" - 2</p> <p>"safeti speed" - 8</p>
<p><i>7. Controllo e validazione dei risultati</i></p> <p>Le frasi generate devono essere controllate per verificare il grado di pertinenza e correttezza logica. Questa fase deve essere affrontata a discrezione dell'analista e dipende dai risultati ottenuti fino a quel momento. Se questa fase non viene superata occorre ripartire dal punto 4.</p>	<p>Basandosi sulle frequenze, in questo caso si nota facilmente che queste parole possono generare i seguenti requisiti:</p> <p>"Arresto di emergenza"</p> <p>"Luci di monitoraggio stato robot"</p> <p>"Laser come barriera di sicurezza".</p> <p>Aumentando il numero di keywords (al passo 2) l'analisi dei requisiti risulta più approfondita e dettagliata.</p>

Il modello, seppur più semplice del precedente, ha un limite importante: correla l'output alla predisposizione grammaticale iniziale del documento, dunque, può non percepire correttamente le reali correlazioni tra le parole nei documenti.

Nell'Allegato 6 è possibile approfondire tale metodologia con un esempio operativo e con il relativo codice in R.

6. CONCLUSIONI

In questo documento è stato mostrato un approccio che a partire da UGC non strutturati ha permesso di definire una scheda di requisiti che è stata poi analizzata.

Nella Tabella 18 si riporta un confronto tra il metodo proposto in questo documento e la Casa della Qualità (HOQ), attualmente utilizzata per la generazione dei requisiti funzionali.

CASA DELLA QUALITÀ	NUOVO APPROCCIO METODOLOGICO
<u>VANTAGGI</u>	<u>VANTAGGI</u>
Occasione di confronto tra collaboratori di diversi reparti	Occasione di confronto tra collaboratori di diversi reparti
Semplice ed immediata da comprendere	Semplice ed immediato da comprendere
	Risultati più oggettivi grazie all'utilizzo dell'intelligenza artificiale
	Basso impiego di risorse per la preparazione
	Minore lead time per la definizione dei requisiti
<u>SVANTAGGI</u>	<u>SVANTAGGI</u>
Risultati basati su giudizi soggettivi	Tempistica dipendente dalla tecnologia adoperata
Preparazione onerosa in termini di risorse	Risultato dipendente dal database iniziale
Preparazione onerosa in termini di tempo	

Tabella 18. Confronto tra l'attuale metodo di definizione dei requisiti (Casa della Qualità) e quello proposto in questo lavoro di tesi

Dall'analisi nel mondo della robotica collaborativa, sono emersi in tutto circa 100 requisiti funzionali percepiti in modo diverso dall'ottica di mercato e delle aziende produttrici.

Il prodotto COBOT è attualmente poco diffuso nei mercati, ma in rapida espansione, perciò comprendere le esigenze di mercato in modo semplice e rapido è d'aiuto per l'ingegnerizzazione

del prodotto su cui basare delle strategie competitive vincenti.

Il metodo propone spunti innovati in quanto riesce a correlare le tecnologie di text-mining, finalizzate allo studio dei Big Data, agli Small Data, un tema attualmente poco conosciuto ma che risulta molto utile per la comprensione dei trend di settore.

Nonostante il metodo produca idee interessanti dal punto di vista metodologico, con risultati che permettono di offrire punti di vista svariati del contesto competitivo del settore dei COBOT, è importante sottolineare come la robustezza di tale metodo sia fortemente correlata al campione di dati in input. Attualmente il Topic modeling deve affrontare direttamente queste limitazioni. Tuttavia, lo studio garantisce una base di partenza su cui impostare punti di discussione per comprendere le esigenze di mercato in modo più rapido.

Le implicazioni di tale metodo possono avere riscontri positivi per i seguenti soggetti:

- **Accademici:** emerge la necessità di sviluppo di metodologie più rapide ed efficienti per la definizione dei requisiti funzionali dei prodotti. L'applicazione reale permette di comprendere che combinando tecnologie all'attuale contesto di disponibilità dei dati, si può giungere rapidamente ad una base di partenza in modo rapido e senza l'utilizzo di troppe risorse. Nel capitolo "Sviluppi futuri" vengono proposte delle metodologie che sono implementabili per l'automatizzazione del processo di definizione dei requisiti: in questo modo, vengono indicate delle vie da percorrere per efficientare ulteriormente l'approccio in termini di tempo e sforzo richiesto dall'analista.
- **Aziende:** la rapidità dei cambiamenti tecnologici mette le aziende sul piano di velocizzare ogni tipo di processo aziendale. Attualmente, comprendere le esigenze di mercato è riconducibile ai modelli di elaborazione dei Big Data, che non permettono di rendere evidenti le specifiche di prodotto. Utilizzando il concetto di Small data, le aziende riescono a percepire i determinanti della qualità dei prodotti in modo più rapido, diminuendo il lead time di progettazione del prodotto e riuscendo a comprendere esigenze specifiche, differenzianti per il mercato e non percepibili dalle attuali analisi già utilizzate.
- **Progettisti:** la possibilità di avere a disposizione un sistema di automatizzazione dei requisiti, è una base di partenza che permette di porre maggiore focus nei risultati anziché al processo di acquisizione dei dati dal mercato. Effettuando più volte l'analisi proposta con parametri differenti, l'algoritmo di Topic modeling può mettere in luce scenari differenti; un ottimo spunto per capire i determinanti della qualità del prodotto e le interazioni che possono esserci tra questi.

7. ALLEGATI

ALLEGATO 1

Di seguito si riportano alcuni codici relativi alla fase di data cleaning

UNISCE I FILE GENERATI DALL'STM IN UN UNICO FILE EXCEL (A1.1)

L'algoritmo di Topic modelling restituisce in output una serie di documenti con estensione .txt e .csv nella medesima cartella in cui viene salvato il codice. occorre salvare i file .csv come "a/b" + "numero progressivo dell'iterazione" (es. a1, b1, a2, b2...) e il file .txt come "c" + "numero progressivo dell'iterazione" (es. c1,c2...) nella medesima cartella. Nel fogliol del file excel in cui viene implementato il codice occorre scrivere nelle celle, a partire dalla cella a1, i riferimenti delle celle stesse (a1, b1, c1, a2...). In questo modo, ogni volta che viene restituito in output dall'algoritmo di text mining dei documenti, si riesce ad evitare il processo di importazione dei dati sul file excel che risulta essere molto dispendioso in termini di tempo. L'algoritmo è in grado di rilevare in modo automatico quali file sono già stati importati, dunque, non occorre fare alcuna modifica al codice per ogni iterazione eseguita.

Option Explicit

```
Private Sub CommandButton1_Click()
```

```
Dim n As Integer
```

```
Dim owb As Workbook
```

```
Dim i As Integer
```

```
Dim j As Integer
```

```
Dim sheet As Worksheet
```

```
j = 1
```

```
i = 1
```

```
n = 10
```

```
For i = 1 To n
```

```
    For j = 1 To 3
```

```

If Not (ThisWorkbook.Worksheets(1).Cells(j, i).Value = "") Then
If j = 3 Then

Dim TextFile As Integer
Dim FilePath As String
Dim FileContent As String

FilePath = "C: \inserire il percorso file\" + ThisWorkbook.Worksh
eets(1).Cells(j, i) + ".txt"

TextFile = FreeFile
Open FilePath For Input As TextFile

FileContent = Input(LOF(TextFile), TextFile)

Set sheet = ThisWorkbook.Sheets.Add(After:=ThisWorkbook.Worksheet
s(ThisWorkbook.Worksheets.Count))

ThisWorkbook.Worksheets(j + 1 + 3 * (i - 1)).Range("A2") = FileCo
ntent

ThisWorkbook.Worksheets(j + 1 + 3 * (i - 1)).Name = ThisWorkbook.
Worksheets(1).Cells(j, i)

Close TextFile

ThisWorkbook.Worksheets(1).Cells(j, i).Clear

Else

Set owb = Workbooks.Open("C: \inserire il percorso file\" + ThisW
orkbook.Worksheets(1).Cells(j, i) + ".csv")

owb.Worksheets(1).Range("A1:CG3000").Copy

Set sheet = ThisWorkbook.Sheets.Add(After:=ThisWorkbook.Worksheet
s(ThisWorkbook.Worksheets.Count))

ThisWorkbook.Worksheets(j + 1 + 3 * (i - 1)).Range("A2").PasteSpe
cial xlPasteValues

ThisWorkbook.Worksheets(j + 1 + 3 * (i - 1)).Name = ThisWorkbook.
Worksheets(1).Cells(j, i)

owb.Close False

ThisWorkbook.Worksheets(1).Cells(j, i).Clear

End If

End If

Next j

Next i

End Sub

```

**FORMATTA IL FILE EXCEL IN MODO DA RENDERLO PRONTO PER L'ANALISI,
SINTETIZZANDO GLI INDICATORI PRINCIPALI E PREDISPONENDO L'AMBIENTE
DI LAVORO IN MODO FLESSIBILE PER L'ANALISI (A1.2)**

Seguendo la predisposizione che viene effettuata dall'output del precedente algoritmo, è possibile utilizzare questo codice per calcolare in automatico la prevalenza Topica di ogni Topic, associare ogni commento alla relativa riga di metadati ed estrarre in automatico i primi 10 documenti più pertinenti per tutti i Topic, ordinati sulla base dei relativi covariati riferiti ad ogni Topic.

L'algoritmo è in grado di rilevare in modo automatico quali file sono già stati importati, dunque, non occorre fare alcuna modifica al codice per ogni iterazione eseguita.

```
Private Sub CommandButton2_Click()

    Dim i, j As Integer
    Dim foglio As Integer
    Dim righe As Integer
    Dim colonne As Integer
    Dim Rng As Range
    Dim k As Integer
    Dim m As Integer
    Dim numero As Integer

    For j = 1 To 10
        If Not (ThisWorkbook.Worksheets(1).Cells(3, j).Value = "") Then
            numero = 1 + 3 * (j - 2)
            Exit For
        End If
    Next j

    For foglio = numero To ThisWorkbook.Worksheets.Count Step 3

        righe = ThisWorkbook.Worksheets(foglio + 2).Range("A3", ThisWorkb
ook.Worksheets(foglio + 2).Range("A3").End(xlDown)).Rows.Count
```

```

        colonne = ThisWorkbook.Worksheets(foglio + 2).Range("B4", ThisWorkbook.Worksheets(foglio + 2).Range("B4").End(xlToRight)).Columns.Count

        For j = 2 To colonne + 1

            Set Rng = Worksheets(foglio + 2).Range(Worksheets(foglio + 2).Cells(3, j), Worksheets(foglio + 2).Cells(righe + 2, j))

            Worksheets(foglio + 2).Cells(1, j) = Application.Average(Rng)

        Next j

        Worksheets(foglio + 2).Cells(1, 1) = "Medie:"

        m = 2

        'For j = 1 To colonne
        ' For i = 1 To righe
        '     ThisWorkbook.Worksheets(foglio + 2).Cells(i + 2, 2 + colonne + j) = Application.WorksheetFunction.Rank(Worksheets(foglio + 2).Cells(i + 2, j + 1), Worksheets(foglio + 2).Range(Worksheets(foglio + 2).Cells(3, j + 1), Worksheets(foglio + 2).Cells(righe + 2, j + 1)))
        'Next i
        'Next j

        For j = 1 To colonne
        For k = 1 To 5
        For i = 1 To righe

            If Application.WorksheetFunction.Rank(Worksheets(foglio + 2).Cells(i + 2, j + 1), Worksheets(foglio + 2).Range(Worksheets(foglio + 2).Cells(3, j + 1), Worksheets(foglio + 2).Cells(righe + 2, j + 1))) = k Then

                ThisWorkbook.Worksheets(foglio + 3).Cells(m, 14) = ThisWorkbook.Worksheets(foglio + 1).Cells(i + 2, 6).Value

                m = m + 1

            End If

        Next i
        Next k
        Next j

```

Next foglio

End Sub

SPOSTA IN COLONNA TUTTO IL DATASET DELLO SCRAPING (A1.3)

L'output di Octoparse salvato in file .csv può essere estratto in modo differente, per esempio estraendo il contenuto per paragrafi. In questo caso Octoparse salva i diversi paragrafi in riga, non riconoscendo l'origine di tale paragrafo. Con questo algoritmo è possibile distinguere i paragrafi in colonne e ricreare i documenti di testo originali partendo dall'output della fase di scraping.

```
x = 2
```

```
y = 70
```

```
For j = 3 To 58 Step 3
```

```
    n = j / 3 + 1
```

```
    ActiveSheet.Range(Cells(x, j + 1), Cells(y, j + 3)).Copy
```

```
    ActiveSheet.Paste Destination:=Worksheets("Sheet1").Range(Cells((n - 1) * y + 1, 1), Cells(y * n + 1, 3))
```

```
Next j
```

```
Range(Cells(x, 4), Cells(y, 59)).Delete
```

UNISCE LE COLONNE DEL DATASET DELLO SCARPING IN UN' UNICA CELLA

(A1.4)

L'output di Octoparse salvato in file .csv può essere estratto in modo differente, per esempio estraendo il contenuto per paragrafi. Dopo l'applicazione dell'algoritmo 3, è possibile applicare il seguente codice e raggruppare il contenuto dei documenti in un'unica cella e ricreare i documenti di testo originali.

```
x = 2
```

```
y = 499
```

```
Z = 11
```

```
Dim q As String
```

```
For i = x To y
```

```

q = ""
  For j = 3 To Z
  If Not (Cells(i, j) = "") Then
  Cells(i, Z + 1) = q + " " + Cells(i, j)
  End If
  q = Cells(i, Z + 1)
  Next j
Next i

```

FORMATTAZIONE DATE (A1.5)

Applicando il seguente codice è possibile modificare le date con il formalismo italiano. In alternativa, può essere usato per determinare il numero progressivo del mese referenziato dalla stringa di testo modificando la condizione dopo il Then di ogni If.

```

Private Sub CommandButton1_Click()

For i = 2 To 5900
  If Cells(i, 4) = "Jan" Or Cells(i, 4) = "January" Then
    Cells(i, 4) = "Gen"
  End If

  If Cells(i, 4) = "Feb" Or Cells(i, 4) = "February" Then
    Cells(i, 4) = "Feb"
  End If

  If Cells(i, 4) = "Mar" Or Cells(i, 4) = "March" Then
    Cells(i, 4) = "Mar"
  End If

  If Cells(i, 4) = "Apr" Or Cells(i, 4) = "Aprile" Then
    Cells(i, 4) = "Apr"
  End If

  If Cells(i, 4) = "May" Or Cells(i, 4) = "May" Then
    Cells(i, 4) = "Mag"
  End If

  If Cells(i, 4) = "Jun" Or Cells(i, 4) = "June" Then

```

```

        Cells(i, 4) = "Giu"
    End If
    If Cells(i, 4) = "Jul" Or Cells(i, 4) = "July" Then
        Cells(i, 4) = "Lug"
    End If
    If Cells(i, 4) = "Aug" Or Cells(i, 4) = "August" Then
        Cells(i, 4) = "Ago"
    End If
    If Cells(i, 4) = "Sep" Or Cells(i, 4) = "September" Then
        Cells(i, 4) = "Set"
    End If
    If Cells(i, 4) = "Oct" Or Cells(i, 4) = "October" Then
        Cells(i, 4) = "Ott"
    End If
    If Cells(i, 4) = "Nov" Or Cells(i, 4) = "November" Then
        Cells(i, 4) = "Nov"
    End If
    If Cells(i, 4) = "Dec" Or Cells(i, 4) = "December" Then
        Cells(i, 4) = "Dic"
    End If

Next i
End Sub

```

Si riportano di seguito funzioni Excel utilizzate per velocizzare il data cleaning; ccome detto nel documento, la formattazione è avvenuta ad hoc per ogni singolo file generato dalla fase di scraping.

La personalizzazione del data cleaning di oltre 40 file differenti ha richiesto circa la definizione di circa 100 funzioni differenti l'una dall'altra.

Le seguenti formule sono state selezionate in quanto le più significative per la formattazione:

- Generare un filtro dei dati testuali, escludendo valori duplicati ex-post l'applicazione del precedente algoritmo. La formula se applicata permette di restituire il set di dati definitivo correggendo gli errori computati dal precedente algoritmo.

=SE(STRINGA.ESTRAI(A1;1;6)=STRINGA.ESTRAI(B1;1;6);SE(E(LUNGHEZZA(A1)-LUNGHEZZA(B1)>11;O(C1=""
";C1=""));B1;SE(E(STRINGA.ESTRAI(A1;1;6)=STRINGA.ESTRAI(C1;1;6);LUNGHEZZA(A1)-
LUNGHEZZA(C1)>11;LUNGHEZZA(B1)>LUNGHEZZA(C1));C1;B1));SE(STRINGA.ESTRAI(A1;1;6)=STRINGA.ESTR
AI(C1;1;6);SE(LUNGHEZZA(A1)-LUNGHEZZA(C1)>11;C1;A1);A1))

- Rimuovere rispettivamente parte iniziale e finale di un documento testuale

=SINISTRA(DESTRA(B2;5);4)
= DESTRA (SINISTRA (B2;5);4)

- Eliminare all'occorrenza stringhe di testo superflue

=DESTRA(A12;LUNGHEZZA(A12)-TROVA("#";A12)-3)

- Estrarre stringhe di testo e formattarle in modo da generare una data riconoscibile dal programma

=ANNULLA.SPAZI(STRINGA.ESTRAI(B8;TROVA(", ";B8)+1;5))
=SE.ERRORE(SE(ANNO(J5)>2021;J5;J5);
DATA(DESTRA(D5;4);SINISTRA(D5;2);DESTRA(SINISTRA(D5;5);2)))

- ...

ALLEGATO 2

Di seguito si riporta l'implementazione dell'indice di Jaccard.

L'implementazione è avvenuta sulla definizione matematica dell'indice, senza l'utilizzo di librerie, per tale ragione il codice può essere utilizzato per il matching tra qualsiasi dataset di documenti testuali.

#input dataset

```
md1<-read.csv("BLOG.csv",sep=" ",header = FALSE)
```

```
md2<-read.csv("FORUM.csv",sep=" ",header = FALSE)
```

#output indice di Jaccard

```
I <- length(intersect(md1$V1,md2$V1))
```

```
S <- I / (length(md1$V1)+length(md2$V1)-I)
```

ALLEGATO 3

Di seguito si riportano i codici usati per predisporre l'ambiente di lavoro all'analisi della sovrapposizione dei requisiti col metodo della frequenza delle keywords.

FILTRA I DOCUMENTI DI OGNI TOPIC RESTITUENDO LA LISTA DEI 100

COMMENTI CON UN MAGGIOR GRADO DI COERENZA TOPICA (A3.1)

Dopo avere formattato i file Excel con il codice A1.1 e A1.2, è possibile estrarre in un unico file i primi 100 documenti più pertinenti per tutti i Topic, ordinati sulla base dei relativi covariati riferiti ad ogni Topic.

L'ambiente di lavoro viene poi formattato in modo automatico per essere di immediata comprensione e flessibilità per l'analisi.

Option Explicit

```
Private Sub CommandButton2_Click()
```

```
    Dim i, j, l, k, m As Integer
```

```
        l = 9
```

```
        colonne = 34
```

```
        righe = 16574
```

```
For j = 1 To colonne
```

```
    m = 3
```

```
    l = l + 1
```

```
For k = 1 To 100
```

```
For i = 1 To righe
```

```
If Application.WorksheetFunction.Rank(Worksheets(2).Cells(i + 2, j + 1),  
Worksheets(2).Range(Worksheets(2).Cells(3, j + 1), Worksheets(2).Cells(righe + 2, j + 1))) = k Then
```

```
    ThisWorkbook.Worksheets(1).Cells(m, l) = ThisWorkbook.Worksheets(2).Cells(i + 2, 34).Value
```

```

        m = m + 1

End If

    Next i

    Next k

    Next j

End Sub

```

- Il contatore delle keywords è stato generato attraverso la seguente formula, in ambiente Excel:

```

=(LUNGHEZZA(ANNULLA.SPAZI(MINUSC(AAA$2)))-
LUNGHEZZA(SOSTITUISCI(MINUSC(AAA$2);$B11;"")))/LUNGHEZZA($B11)

```

Tale formula permette di valutare la lunghezza della stringa di testo e paragonarla a tale lunghezza privata della keyword in esame, ove presente. Con questo approccio, il numero di keywords è uguale al rapporto tra i caratteri annullati e la lunghezza della keyword stessa.

Eventuali frazioni di multiplo sono da imputare ad un'errata predisposizione delle keywords, a causa dello stemming effettuato dall'stm, nel documento: la formula è stata costruita in modo da poter annullare tale errore con una semplice approssimazione all'intero più vicino.

- Il ranking delle keywords è stato creato con la seguente formula:

```

=RANGO.UG(GU4;$GU$4:$GU$43;0)

```

Per ulteriori formule utilizzate si rimanda alla documentazione del lavoro di tesi.

Di seguito si riportano i codici usati per predisporre l'ambiente di lavoro all'analisi della sovrapposizione dei requisiti col metodo dell'indice di Jaccard.

IMPORTA UN FOGLIO IN UN FILE CSV DELIMITATO DA VIRGOLE (A3.2)

Il codice trasforma in automatico un file qualsiasi in un file con estensione .csv. Avendo svolto tutte le manipolazioni dei dati e le analisi quantitative in Excel, questo codice è risultato utile per trasformare i dataset di documenti da .xml a .csv per passarli successivamente in input agli algoritmi di intelligenza artificiale in R.

```

Sub Importa()

```

```

Dim foglio As Worksheet

```

```
Dim percorso As String

'Application.DisplayAlerts = False

For Each foglio In ActiveWorkbook.Worksheets

    percorso = " C: \inserire il percorso file\"

    ActiveWorkbook.Sheets(foglio.Index).Copy

    ActiveWorkbook.SaveAs Filename:= percorso & foglio.Name, FileFormat:=
xlCSV, CreateBackup:=True

    ActiveWorkbook.Close

Next foglio

'Application.DisplayAlerts = True

End Sub
```

ALLEGATO 4

Di seguito si riportano i link utili a cui rimanda la documentazione di input per l'algoritmo di Topic modeling, riferite rispettivamente ai requisiti di prodotto e funzionalità di prodotto.

REQUISITI DI PRODOTTO:

FORUM – REQUISITI DI PRODOTTO
https://www.zacobria.com/
https://www.robotforum.com/
https://www.societyofrobots.com/
https://robotics.stackexchange.com/
https://dof.robotiq.com/
https://github.com/
https://robodk.com/
https://forum.universalrobots.com/
https://www.universalrobots.com/
https://www.practicalmachinist.com/

FUNZIONALITÀ DI PRODOTTO:

BLOG – FUNZIONALITÀ DI PRODOTTO
https://www.therobotreport.com/
https://www.roboticsbusinessreview.com/
https://www.cobottrends.com/
https://www.weforum.org/
https://www.alumotion.eu/
https://www.roboglobal.com/
https://www.zdnet.com/
https://spectrum.ieee.org/searchContent?q=coBOT&query=collaborative+robot&max=10
https://www.universal-robots.com/
https://www.automate.org/
https://robotiq.com/it/
https://robohub.org/
https://www.reeco.co.uk/
https://www.techbriefs.com/
https://www.tm-robot.com/
https://www.doosanrobotics.com/
https://www.tenenga.it/
https://singularityhub.com/
https://www.kuka.com/
http://preciseautomation.com/
http://www.nexman.it/
https://www.zacobria.com/

ALLEGATO 5

Di seguito si riporta un esempio applicativo preso dalla letteratura di RStudio (<https://rpubs.com/nabiilahardini/text-generator>).

```
# data wrangling
library(tidyverse)

# text processing
library(tidytext)
library(textclean)
library(tokenizers)

# markov chain
library(markovchain)

# read data
tlp <- read.delim("the_little_prince.txt",
                 col.names = "text")

tlp_clean <- tlp %>%
  slice(-1) %>% # remove first line (version info)
  filter(!str_detect(text, "[/:]"), # remove lines with certain character
        !str_detect(text, "Chapter")) # remove lines with certain string
tlp_clean <- tlp_clean %>%
  mutate(text = tolower(text) %>% # tolower sentences
         replace_contraction() %>% # expand contraction
         replace_white() %>% # replace double white space into single
space
         str_remove_all(pattern = "lpp_1943.") %>% # remove pattern
         str_remove_all(pattern = "[0-9]") %>% # remove numbers
         str_remove_all(pattern = "[()]" ) %>% # remove specific punctua
tion
         str_remove_all(pattern = "--") %>%
         str_replace_all(pattern = " - ", replacement = "--") %>% # re
place pattern
         str_replace_all(pattern = "n't", replacement = "not") %>%
         str_remove(pattern = "[.]") %>% # remove first matched pattern
```

```

        str_remove(pattern = " ")
# glimpse data; first 10 sentences
head(tlp_clean, 10)
# split words from sentences
text_tlp <- tlp_clean %>%
  pull(text) %>%
  strsplit(" ") %>%
  unlist()
text_tlp %>% head(27)
fit_markov <- markovchainFit(text_tlp)
create_me <- function(num = 5, first_word = "i", n = 2) {
for (i in 1:num) {
  set.seed(i+5)
  markovchainSequence(n = n, # generate 2 additional random words
                      markovchain = fit_markov$estimate,
                      t0 = tolower(first_word), include.t0 = T) %>%
# joint words
  paste(collapse = " ") %>% # join generated words with space
# create proper sentence form
  str_replace_all(pattern = " ,", replacement = ",") %>%
  str_replace_all(pattern = " .", replacement = ".") %>%
  str_replace_all(pattern = " !", replacement = "!") %>%
  str_to_sentence() %>% # start every sentences with capitalization
  print()
}
}
create_me(num = 5, first_word = "i", n = 3)

```

A partire dalle seguenti keywords...

```

#> [1] "once"          "when"           "i"              "was"           "six"
#> [6] "years"          "old"            "i"              "saw"           "a"
#> [11] "magnificent"   "picture"        "in"             "a"             "book"
#> [16] ",,"            "called"         "true"           "stories"       "from"

```

```
#> [21] "nature"      ", "          "about"      "the"        "primeva  
1"  
#> [26] "forest"      "."
```

...è possibile generare le seguenti frasi:

```
#> [1] "I saw the flower"  
#> [1] "I were some use"  
#> [1] "I do not show"  
#> [1] "I was a flower"  
#> [1] "I could fly to"
```

ALLEGATO 6

Di seguito si riporta un esempio applicativo preso dalla letteratura del web.

Il seguente esempio utilizza due approcci differenti per trovare la frequenza relativa ad ogni combinazione di keywords, basati però sullo stesso costrutto logico.

```
# example using regexpr and gregexpr
species <- c("Onthophagus_vacca", "Copriss_hispanus", "Carabus_hispanus_hispanus")
regexpr("hisp", species)

#regexpr return the position in the string of the first occurrence of the
pattern as well as the length of the pattern matched
## [1] -1 8 9
## attr(,"match.length")
## [1] -1 4 4
## attr(,"useBytes")
## [1] TRUE
gregexpr("hisp", species)

#gregexpr will return the position of all matched occurrence of the pattern
## [[1]]
## [1] -1
## attr(,"match.length")
## [1] -1
## attr(,"useBytes")
## [1] TRUE
##
## [[2]]
## [1] 8
```

```
## attr(,"match.length")
## [1] 4
## attr(,"useBytes")
## [1] TRUE
##
## [[3]]
## [1] 9 18
## attr(,"match.length")
## [1] 4 4
## attr(,"useBytes")
## [1] TRUE
regexec("hisp", species)
```

#similar as regexpr but with a different output formatting

```
## [[1]]
## [1] -1
## attr(,"match.length")
## [1] -1
##
## [[2]]
## [1] 8
## attr(,"match.length")
## [1] 4
##
## [[3]]
## [1] 9
## attr(,"match.length")
## [1] 4
```

In entrambi i casi l'utilizzo di *regexpr* e *gregex* restituiscono il numero di occorrenze trovate per una data stringa: l'unica differenza risulta nell'output della formattazione.

Nel primo caso troviamo la posizione delle uguaglianze all'interno del vettore contenente le stringhe di testo (il documento del Topic), rendendo più semplice le fasi di analisi e ricerca di eventuali concetti logici da parte del progettista.

Dall'altra parte, viene restituito il numero di occorrenze totali, rendendo la fase di selezione dei requisiti più rapida.

La scelta è assolutamente irrilevante per la selezione dei requisiti, ma permette di efficientare il processo di analisi in due fasi diverse, per tale ragione, dovrebbe essere considerato fin da subito nella scelta.

ALLEGATO 7

Nella pagina seguente si riporta la scheda dei requisiti del prodotto COBOT con le relative tendenze evolutive e gli ordini di importanza percepiti dai clienti e dal mercato.

Nella prima colonna si riporta l'evoluzione temporale su base annuale dei requisiti e nella seconda colonna si riporta il confronto domanda-offerta tra i gradi di importanza attribuiti a tali requisiti, come spiegato nel paragrafo 4.3. In rosso si evidenziano le i requisiti che sono percepiti dall'offerta di mercato, mentre, nella quarta colonna il grado di rilevanza che viene dato dai consumatori allo specifico bisogno.

		CONVENIENZA	0,05643
	✓0	Economicità	0,05643
	✗60000	Durabilità	0,05643
	!-1	Compatibilità ricambi	0,05643
		OPERATIVITA	0,03959
	✓0	Stabilità in esecuzione con qualsiasi carico	0,04642
	✗60000	Stabilità ad alte vibrazioni	0,04642
	✓2	Aumento numero assi	0,04404
	✓2	Aumento ampiezza raggio di lavoro	0,04404
	!-1	Anticollisione in esecuzione	0,04404
	!-4	Sensori di coppia di forza (versatilità operativa)	0,04404
	✓3	Movimentazione morbida del braccio	0,04251
	✓3	Velocità operativa	0,04251
	✗60000	Regolazione continua del braccio	0,04251
	✗60000	Adattamento alla tensione di esercizio	0,03739
	✗60000	Ridondanza porte di comunicazione	0,03739
	✗60000	Scalabilità del controller	0,03739
	✓1	Latenza di esecuzione	0,03271
	✗60000	Scorrevolezza dress pack ai movimenti del Cobot	0,01287
		SOFTWARE	0,03256
	✗60000	Compatibilità continua delle versioni del firmware	0,06165
	✗60000	Garantire programmazione da remoto	0,04486
	✗60000	Facilità di trasferimento dati	0,04486
	!-1	Facilità di attivazione interfaccia client	0,03917
	✗60000	Diversificazione delle interfacce di comunicazione	0,03271
	✓3	Aumentare usabilità tech-pendant	0,03005
	✓4	Facilità di programmazione movimentazione braccio	0,02961
	✗60000	Compatibilità con versioni dei software di supporto	0,02883
	✗60000	Parallelizzare Thread in esecuzione	0,02483
	✓8	Free Drive Mode	0,02450
	✗60000	Remote Control Mode	0,02450
	✓11	Semplicità di definizione degli waypoint	0,02018
	✗60000	Implementare RPCXML	0,01756
		SICUREZZA	0,03165
	!-1	Arresto di emergenza	0,03165
	✗60000	Luci di monitoraggio stato robot	0,03165
	!-1	Laser come barriera di sicurezza	0,03165
	!-1	Regolazione velocità di esecuzione in presenza di operatore	0,03165
		CUSTOMIZZAZIONE	0,01911
	!-2	Customizzare sensoristica visiva	0,02616
	✓1	Diversificazione terminali	0,01831
	✗60000	Customizzazione Cobot dress pack	0,01287
		SERVIZIO	0,02846
	!-7	Offrire codice modulare	0,04418
	✗60000	Velocità di fornitura del servizio\prodotto	0,04383
	!-3	Istruzione operatore tecnico	0,03005
	!-2	Istruzione programmatore	0,03005
	!-4	Condivisione di modelli strutturali operativi	0,02987
	!-1	Interoperabilità con ambienti open source	0,02294
	✗60000	Documentazione online completa e di immediata comprensione	0,02124
	✗60000	Documentazione online ben strutturata	0,02124
	✗60000	Supporto post-vendita garantito	0,01275

Tabella 19.
Scheda dei requisiti
del prodotto COBOT

ALLEGATO 8

Nella pagina seguente si riporta la scheda delle funzionalità di prodotto percepite dal mercato, dunque dalle aziende produttrici. In rosso si evidenziano le funzionalità che sono corrisposte da un bisogno del cliente, mentre, nella seconda colonna il grado di rilevanza che viene dato dalle aziende produttrici a tali funzionalità.

CONVENIENZA	0,01457
Economicità	0,01485
Compatibilità ricambi	0,01429
OPERATIVITA	0,01354
Aumento numero assi	0,01848
Aumento ampiezza raggio di lavoro	0,01848
Stabilità in esecuzione con qualsiasi carico	0,01848
Anticollisione in esecuzione	0,01724
Movimentazione morbida del braccio	0,01724
Velocità operativa	0,01724
Sensori di coppia di forza (versatilità operativa)	0,01259
Aumento payload supportabile	0,01259
Sensibilità operativa	0,00940
Stabilità in esecuzione in ogni ambiente (waypoint)	0,00814
Rapidità set-up	0,00688
Latenza di esecuzione	0,00564
CUSTOMIZZAZIONE	0,01142
Diversificazione terminali	0,01319
Pacchetti personalizzabili per far fronte alle diverse necessità	0,01136
Customizzare sensoristica visiva	0,00970
SICUREZZA	0,00959
Incremento potenzialità sensori con AI	0,01244
Arresto di emergenza	0,00865
Laser come barriera di sicurezza	0,00865
Regolazione velocità di esecuzione in presenza di operatore	0,00865
SOFTWARE	0,00704
Semplicità di definizione degli waypoint	0,01103
Free Drive Mode	0,00814
Facilità di programmazione movimentazione braccio	0,00565
Aumentare usabilità tech-pendant	0,00565
Facilità di attivazione interfaccia client	0,00473
SERVIZIO	0,01005
Esposizioni divulgative	0,01807
Offrire soluzioni applicative complete	0,01416
Connettività nel cloud	0,01127
Sicurezza nel cloud	0,01127
Istruzione programmatore	0,00922
Istruzione operatore tecnico	0,00890
Interoperabilità con ambienti open source	0,00870
Offrire codice modulare	0,00834
Condivisione di modelli strutturali operativi	0,00565
Piattaforme ecosistemiche	0,00494

Tabella 20.
Scheda delle
funzionalità del
prodotto COBOT

8. BIBLIOGRAFIA

- Akao, Y., Mazur, G. H., (2003), "The leading edge in QFD: past, present and future", International Journal of Quality & Reliability Management, volume 20 issue 1 pag. 20 - 35
- Alfieri, A., (2021), slides del capitolo "Random Variate Generation" del corso di Simulazione dei Sistemi Gestionali, Politecnico di Torino, Torino
- Bird, S., Klein, E., Loper, E., (2009), "Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit", 1st edition, published by O'Reilly Media
- De Sirisuriya, S. C. M. (2015) 'A Comparative Study on Web Scraping', in Proceedings of 8th International Research Conference of KDU. General Sir John Kotelawala Defence University, pag. 135–140
- Erdil, N. O., Arani, O. M., (2018), "Quality function deployment: more than a design tool", International Journal of Quality and Service Sciences, volume 11, issue 2
- Franceschini, F., Galetto, M., Maisano, D. A., Mastrogiacomo, L., (2020), "Ingegneria della qualità - Applicazioni ed esercizi", Quarta edizione, C.L.U.T. Editrice, pag. 13, 185 – 188, Torino
- Giardinelli, A. (2018), "Analisi di fattibilità del prodotto finalizzata all'applicazione di Robot Collaborativi nel General Assembly in FCA", EPrints , pag. 80-118
- Govers, P. M., (2001), "QFD not just a tool but a way of quality management", International Journal of Production Economics, vol. 69, issue 2, pag. 151-159
- Grandi, A., (2017), "Gestione dei progetti d'innovazione", McGraw Hill Education, pag. 542-561, Bolog
- Gupta, V., Lehal, G. S., (2009), "A Survey of Text Mining Techniques and Applications", Journal of Emerging Technologies in Web Intelligence, pag. 60-76
- Hu, M., (2014), "Small Data Surveillance v. Big Data Cybersurveillance", Pepperdine University, volume 42, issue 4, rev.773
- Ionescu, T. B., (2020), "Leveraging Graphical User Interface Automation for Generic Robot Programming", Robotics, 10, 3, EISSN 2218-6581, published by MDPI
- ISO/IEC. 13407 Human-Centred Design Processes for Interactive Systems, ISO/IEC 13407: 1999 (E), 1999
- ISO 10218-1, Robots and robotic devices — Safety requirements for industrial robots — Part 1: Robots, ISO 10218-1:2011, 2011
- ISO 10218-2, Robots and robotic devices — Safety requirements for industrial robots — Part 2: Robot systems and integration, ISO 10218-2:2011, 2011
- Jung, H.W., (1998), "Optimizing value and cost in requirements analysis", IEEE Software, vol. 15, no. 4, pp. 74-78
- Krumm, J., Davies, N., Narayanaswami, C., (2008), "User Generated Content", IEEE Pervasive Computing, volume 7, issue 4, pag. 10-11
- Marres, N., Weltevrede, E., (2013), "SCRAPING THE SOCIAL?", Journal of Culture Economy, volume 6, issue 3, pag.313-335

- Hearst, M. A., (1999), "Untangling Text Data Mining", Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, University of Maryland, June 20-26
- Mastrogiacomo, L., Barravecchia, F., Franceschini, F., Marimon, F., (2021), "Mining quality determinants of product-service systems from user-generated contents", Journal of the American Society for Quality, volume 33, issue 3, pag. 425 - 442
- Mazur Glenn (2014), "Methods and tools of modern QFD", webcast of American Society for Quality
- Mei, Q., Shen, X., Zhai, C. X., (2007), "Automatic labeling of multinomial Topic models", 23rd International Conference on Computational Linguistics Proceedings of the Conference Volume 2, pag.490-499
- Mokaram, S., Aitken, J. M., Martinez-Hernandez, U., Eimontaite, I., Cameron, D., Rolph, J., Gwilt, I., McAree, O., Law, J., (2017) , "A ROS-integrated API for the KUKA LBR iiwa collaborative robot", IFAC-PapersOnLine, volume 50, issue 1, pag. 15859-15864
- Orlando, S., (2009), slides del capitolo "Clustering" del corso Data e Web Mining, Università Ca' Foscari di Venezia, Venezia
- Polillo, R., (2010), "FACILE DA USARE - Una moderna introduzione alla ingegneria dell'usabilità", Edizioni Apogeo, Mondadori, Capitolo 6-7
- Porru, N., (2018), "Big Data e Data Analytics come supporto ai processi decisionali: definizione e analisi dei vantaggi competitivi per le aziende", EPrints 3
- Roberts, M. E., Stewart, B.M., Tingley, D., (2019)"stm: R Package for Structural Topic Models", Journal of Statistical Software, volume 91, issue 2, pag 1-40
- Roberts, M. E., Stewart, B.M., Tingley, D., Benoit, K., (2020), "Estimation of the Structural Topic Model", Journal of Statistic Software, pag.21-55
- Rocchi, D., (2020), "La rivoluzione dei Big Data", Università Ca' Foscari Venezia, Venezia
- Saurkar, R., Anand, V., Kedar, G., Pathare and Shweta, Gode, A., (2018), "An Overview On Web Scraping Techniques And Tools", International Journal on Future Revolution in Computer Science & Communication Engineering, pag. 363-367
- Scott, J. and Baldridge, J. (2013) 'A recursive estimate for the predictive likelihood in a Topic model', Journal of Machine Learning Research, Vol. 31, pag. 527-535
- Vanderborght, B., - European Commission, (2019), "Unlocking the potential of industrial human–robot collaboration - A vision on industrial collaborative robots for economy and society", Publications Office of the European Union, pag. 6-30
- Vicentini, F., (2017), "Robotica collaborativa", collana Tecnologie Industriali, Tecniche Nuove, Milano
- Wyrwoll, C., (2014), "User Generated Content", Social Media, Springer Fachmedien Wiesbaden, pag. 11-45

9. SITOGRAFIA

<https://blog.robotiq.com/what-does-collaborative-robot-mean>

<https://it.semrush.com/>

<https://rpubs.com/nabiilahardini/text-generator>

<https://secure.asq.org/perl/msg.pl?prvurl=http://asq.org/2014/12/methods-and-tools-of-modern-qfd-webcast-pptslides.pdf>

<https://www.automate.org/news/sepro-group-and-universal-robots-announce-new-cobot-partnership>

<https://www.bastiansolutions.com/blog/collaborative-robots-part-2-benefits-and-expanding-capabilities/>

<https://www.blog-management.it/2018/09/14/case-study-sneakers-lego-small-data/>

<https://www.cobottrends.com/china-us-to-drive-cobot-sales-post-covid/>

<https://www.olapic.com/resources/consumer-trust-usage-attitudes/>

https://www.roboticsbusinessreview.com/rbr/how_will_president_trump_affect_robotics_industry/

<https://www.therobotreport.com/cobot-market-grow-8b-2030-report-finds/>

<https://www.universal-robots.com/blog/tags/human-cobot-collaboration/>

<https://www.universal-robots.com/blog/the-benefits-of-an-ecosystem/>