

Sistemi Basati sull'Intelligenza Artificiale per il Supporto alle Decisioni negli Appalti Pubblici

Pierpaolo Basile¹, Marco Di Ciano², Nicola Lopane³, Pasquale Lops^{1,*}, Giuseppe Marziale³, Monica Messa², Lucia Siciliani¹, Vincenzo Taccardi¹ and Vincenzo Tamburrano¹

¹Dip. di Informatica - Università di Bari Aldo Moro, Via E. Orabona 4, 70125 Bari, Italy

²InnovaPuglia S.p.A., str. prov. per Casamassima km. 3.000 - 70010 Valenzano (Ba), Italy

³Regione Puglia, Via G. Gentile 52 - 70126 Bari, Italy

Abstract

Gli appalti pubblici rappresentano un potente strumento d'investimento dei fondi pubblici e sono una risorsa strategica per lo sviluppo economico. Risulta quindi fondamentale favorire il miglioramento dell'operatività delle stazioni appaltanti e sviluppare modelli di valutazione che possano agevolare i processi di analisi e monitoraggio nelle varie fasi di esecuzione di una gara pubblica. In questo contributo, presentiamo la nostra ricerca volta alla creazione di un sistema a supporto delle decisioni e delle attività di monitoraggio dell'intero ciclo degli investimenti e degli appalti. Tale sistema prevede l'utilizzo di tecniche di Intelligenza Artificiale basate sull'elaborazione del linguaggio naturale e machine learning, analisi dei dati e visualizzazioni, atte a fornire strumenti che consentano di estrapolare informazioni utili a partire da strutture dati strutturate e non strutturate.

Keywords

Information retrieval, Natural language processing, Machine learning, E-procurement, Data analysis, Business intelligence

1. Introduzione

La Regione Puglia (Servizio Contratti e Programmazione Acquisti) con il supporto tecnico di InnovaPuglia Società in-house e Soggetto Aggregatore della Regione e il contributo scientifico del Dipartimento di Informatica dell'Università degli Studi di Bari ha avviato da alcuni anni lo studio e la sperimentazione di applicazioni di Intelligenza Artificiale a supporto dei processi gestionale-amministrativi e dell'osservatorio regionale dei contratti pubblici (Dlgs 50/2016, art. 29 c.1,2,3,4,4bi) nell'ambito del quadro normativo di settore sull'e-Procurement e gli appalti pubblici.

In particolare, sono stati realizzati due progetti di ricerca e sperimentazione *SLAP* (Sistemi di Intelligenza Artificiale per il Procurement) e *PAI* (Passaporto Appalti Imprese) nel corso dei quali, adottando tecniche di Intelligenza Artificiale come Apprendimento Automatico e Natural Language Processing, è stata realizzata una piattaforma per il monitoraggio del ciclo di investimenti e appalti. Questi progetti hanno preso le mosse dallo

scenario in cui gli acquisti si muovono negli ultimi anni, fatto di mercati sempre più complessi e di rischi sempre nuovi. Le complessità esogene portano gli attori coinvolti a dover perseguire nuovi livelli di efficienza, sostenibilità e salvaguardia. Per farlo è necessario effettuare analisi sempre più puntuali per individuare inefficienze, rischi ed ottimizzare il più possibile i processi interni. Lo scopo dei suddetti progetti è quello di migliorare da un lato le attività dei RUP (Responsabili Unici del Procedimento), delle Stazioni Appaltanti, delle Amministrazioni e degli Enti Aggiudicatori consentendo loro di assolvere a molti degli adempimenti assegnati in maniera più efficace, efficiente e sostenibile e dall'altro, sviluppare schemi di assessment che mettano in correlazione particolari sequenze logico-temporali di fatti e contenuti che possano essere ricondotti a determinati indicatori di anomalia.

2. Architettura

Nei successivi paragrafi descriveremo il risultato dell'attività di ricerca di cui sopra. È stato sviluppato un sistema in grado di fornire supporto nella gestione del ciclo degli appalti, concepito con la finalità di fornire supporto ai RUP. L'architettura è basata su uno schema modulare che consente flessibilità per eventuali modifiche o aggiunte successive.

La figura 1 mostra lo schema dell'architettura ad alto livello, questa è strutturata in quattro moduli principali ed eventuali plug-in:

- *Data Collector*: il modulo si occupa della raccolta dei dati relativi ai bandi e alle gare. L'estrazione

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy

*Corresponding author.

✉ pierpaolo.basile@uniba.it (P. Basile); m.diciano@innovapuglia.it (M. D. Ciano); n.lopane@regione.puglia.it (N. Lopane); pasquale.lops@uniba.it (P. Lops); g.marziale@regione.puglia.it (G. Marziale); m.messa@innovapuglia.it (M. Messa); lucia.siciliani@uniba.it (L. Siciliani); vincenzo.taccardi@uniba.it (V. Taccardi); vincenzo.tamburrano@uniba.it (V. Tamburrano)

📞 0000-0002-0545-1105 (P. Basile); 0000-0002-6866-9451 (P. Lops); 0000-0002-1438-280X (L. Siciliani)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



- dei dati è modularizzata e scomposta in diversi Plug-Ins, ognuno dei quali si occupa di recuperare l'informazione da una specifica sorgente. Poiché le sorgenti sono per loro natura eterogenee, è necessario includere una componente di Data Integration che si occupi di combinare l'informazione proveniente da ciascuna di esse. Tale componente può verificare che ci sia un corretto overlap tra sorgenti distinte e segnalare eventuali anomalie.
- *Pre-Processing*: il modulo si occupa di sublimare e organizzare le informazioni estratte tramite il modulo di data collection in entità sulle quali possano essere efficacemente effettuati i successivi passi di analisi.
 - *Tender Analyzer*: il modulo ha lo scopo di effettuare le analisi a partire dalle informazioni estratte dalle diverse sorgenti. Data la natura delle informazioni che possono essere associate ad un bando, questo modulo è suddiviso in due diverse componenti: il *Data Analyzer* si occupa di analizzare l'informazione strutturata associata ai bandi ed il *Content Analyzer* che invece, effettua l'analisi dell'informazione non strutturata associata ai bandi.
 - *Service Tools*: date le informazioni e le analisi svolte all'interno del modulo precedente, con i service tools vengono effettuate le operazioni specifiche per la realizzazione di particolari casi d'uso.

3. Applicazioni

Il sistema realizzato è il risultato di due progetti di ricerca ciascuno focalizzato su casi d'uso differenti: SIAP e PAI

Di seguito descriveremo i rispettivi obiettivi e le tecnologie utilizzate per la risoluzione.

3.1. Progetto SIAP

In questo progetto sono proposti i seguenti casi d'uso:

1. *Disambiguità procedurali*: una descrizione simile spesso sottende due tipologie differenti di finalità di acquisto, ad esempio, l'una quale appalto di servizi, l'altra come appalto di fornitura. Attraverso sistemi di interpretazione del linguaggio naturale, dedurre se la procedura di cui trattasi ha come finalità una fornitura (es. stampa di "buoni pasto") o un servizio (es. una nuova metodologia di fruizione di pasti attraverso i "buoni pasto").
2. *Disallineamenti qualitativi*: a volte è difficile ottenere o recuperare prezzi di riferimento in quanto in una gara X fatta dal soggetto Y il disinfettante è stato pagato Z, mentre in una gara M fatta dal soggetto N lo stesso disinfettante è stato

pagato P salvo poi scoprire che le unità di misura di riferimento non erano confrontabili. Testare soluzioni di AI che siano in grado di aprire i singoli fascicoli andando ad estrarre contenuti informativi, lasciando poi al RUP¹ la decisione finale di interpretazione.

3. *Indagini preliminari di mercato*: Il RUP ha spesso necessità di accedere ad informazioni in aree di fornitura e per tipologie di merceologie non sempre di sua pertinenza e competenza per raccogliere informazioni e colmare il gap informativo tra amministrazione e mercato. Mediante sistemi di Intelligenza Artificiale identificare le variabili principali di una fornitura (se parliamo di PC sarebbero identificate alcune variabili rilevanti quali processore, ram, batteria, etc.) che consentano al RUP di adeguare l'asimmetria informativa normalmente presente in chi si appropria ad una procedura di gara soprattutto in ambiti nei quali la tecnologia e le soluzioni cambiano con rapidità.

3.1.1. Risoluzione dello Scenario

Per rispondere alle richieste formulate in questo progetto di ricerca, sono stati utilizzati i dati testuali dei documenti ottenuti dal portale EmPULIA², nonché i dati sui bandi in forma strutturata messi a disposizione da ANAC³ come open data. Per i dati in formato non strutturato le attività di analisi sono state condotte attraverso l'utilizzo di tecniche basate sul Natural Language Processing (NLP) [1], sull'analisi semantica (Semantic analysis), Natural Language Understanding e Generation [2], tecniche di Open Information Extraction [3], [4] e Machine Learning.

Nel dettaglio, rispettivamente ad ogni caso d'uso individuato sono state proposte la seguenti applicazioni:

1. *Classificatore CPV*: Questa componente permette all'utente di inserire un testo e di ottenere una lista di codici CPV⁴ (Common Procurement Vocabulary) ad esso associati. L'approccio utilizzato si basa sull'utilizzo di un classificatore addestrato sui bandi di gara disponibili.
2. *Open Information Extraction*: mediante tecniche di Natural Language Processing e Information Retrieval, è possibile analizzare la documentazione relativa ad un bando e fornire al RUP supporto nel ritrovamento ed analisi di tali dati. In particolare potranno essere rintracciati i documenti rilevanti e loro porzioni di interesse per aiutare il RUP a recuperare e gestire più facilmente tali informazioni.

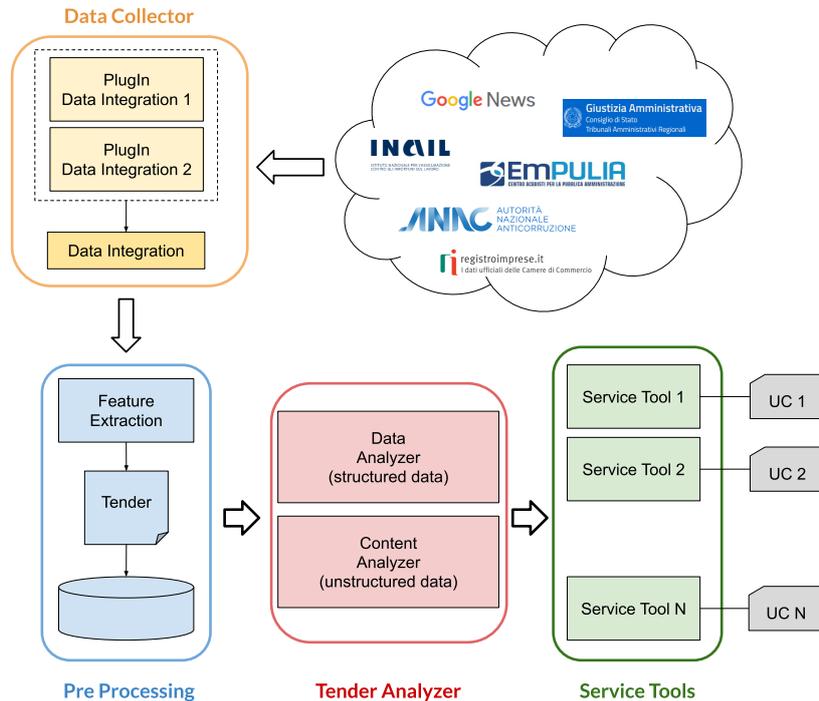
¹Responsabile Unico del Procedimento

²<http://www.empulia.it/>

³<https://dati.anticorruzione.it/#/home>

⁴<https://simap.ted.europa.eu/it/cpv>

Figure 1: Architettura del sistema ad alto livello.



Questa ricerca potrà essere effettuata sulla base della terminologia comunemente utilizzata per indicare le informazioni relative alle quantità e ai costi delle forniture. Ad esempio: “Unità di misura”, “Lotto”, “Pezzo”, “pz”, “Quantità”, “Q.tà”, “Importi base d’asta”, etc.

3. *Semantic Search*: un motore di ricerca in grado di ricevere interrogazioni e restituire, sulla base di queste ultime, i documenti più rilevanti non solo sulla base di misure di co-occorrenza tra i termini che compaiono nella richiesta (query) e quelli presenti all’interno dei documenti, ma anche sfruttando la similarità semantica. Inoltre permette la creazione di riassunti in grado di condensare, con la granularità desiderata, le informazioni presenti all’interno di testi caratterizzati da dimensioni considerevoli.

3.2. Progetto PAI

Lo scopo di questo progetto di ricerca è di rispondere al seguente caso d’uso:

1. *“Passaporto” operatore economico*: per il RUP è possibile e tutto sommato facile accedere a documenti pubblici di gare bandite da altri enti/isti-

tuzioni, qualora lo stesso fosse interessato a informazioni relative a forniture similari. Ciò che risulta difficile reperire è il grado di affidabilità dei fornitori. Sarebbe molto utile riuscire a gestire e analizzare informazioni e documenti (comunque disponibili in altri database o in rete) su eventuali procedure relative a ricorsi, inchieste, sentenze o altro, anche disponibile su notizie di stampa, che aiutino il responsabile del procedimento a capire meglio ad esempio come si siano concluse le altre procedure provando ad assegnare quindi un gradiente di affidabilità alla potenziale fornitura sul tema di interesse di gara.

L’obiettivo è quindi quello di costruire un «Passaporto» per le imprese/operatori economici nell’ambito degli appalti pubblici che consenta alla stazione appaltante e al RUP di accedere a detto documento per una presa visione in tempo reale, rimandando ad un separato processo di attestazione le verifiche.

3.2.1. Risoluzione dello Scenario

In questo ambito di ricerca sono stati utilizzati i seguenti dati: (1) gli *open data ANAC*, che raccolgono i dati sui contratti pubblici in Italia, (2) i dati forniti dal *Registro*

*Imprese*⁵ il portale con i dati e i documenti ufficiali della Camera di Commercio, (3) le notizie di stampa ottenute da *Google News* e (4) le informazioni su eventuali contenziosi amministrativi reperibili sul portale *Giustizia Amministrativa*⁶.

La soluzione proposta prevede di recuperare le informazioni riguardanti un'impresa oggetto di interesse, attingendo ad un portafoglio di dataset. Queste sono utili a definire un profilo aziendale che permetta all'operatore impegnato nell'iter di un appalto di valutare la congruità di un'azienda rispetto ad una possibile gara. Nel dettaglio il caso d'uso proposto è stato risolto come sotto:

Business Intelligence: sfruttando i dati storici disponibili nel dataset ANAC sono state implementate delle dashboard che, a partire dall'ingente mole di dati disponibili, permettono di estrarre ed analizzare informazioni sull'attività storica di un'azienda, nel settore degli appalti pubblici. Inoltre, considerato quanto proposto in [5] ed adattandolo al dataset a nostra disposizione, sono stati elaborati degli indicatori di collusione che, inseriti nelle dashboard di cui sopra, arricchiscono il patrimonio informativo permettendo di condensare indici di qualità della concorrenza per le gare a cui ha partecipato l'azienda oggetto di interesse. Gli indicatori proposti sono: *Valore relativo del bando*, *Intervallo delle offerte*, *Offerte mancanti*, *Prevalenza di domande erronee*, *Prevalenza dei consorzi*, *Prevalenza dei subappalti*, *Mercato concentrato*, *Struttura del mercato stabile*.

Con gli stessi presupposti è stata anche implementata una dashboard che permette di ricavare informazioni attinenti un mercato che può essere definito, secondo diversi livelli di dettaglio, per luogo, data, settore commerciale (CPV) o stazione appaltante. Ciò consente il monitoraggio dell'attività delle stazioni appaltanti.

Parallelamente a questa soluzione, per arricchire il profilo informativo dell'azienda è stata realizzata una sezione del sistema che concentra in un'unica interfaccia i risultati ottenuti, interrogando diverse banche dati, ad esempio, mediante API REST: *Registro Imprese* e *Google News*. Per accedere ad un'ulteriore fonte di dati, individuata nel portale *Giustizia Amministrativa*, non essendo disponibile un'API, è stato necessario implementare un servizio di crawling che sfruttando l'interfaccia di ricerca disponibile sul sito, in forma automatizzata sottomette la query desiderata e ne restituisce i risultati ottenuti.

4. Conclusioni e Sviluppi Futuri

Il progetto di ricerca fin qui esposto si propone di investigare le possibilità offerte dalle tecnologie nel campo

dell'intelligenza artificiale per offrire agli attori impegnati nella gestione degli appalti pubblici una serie di strumenti utili ad agevolarne il lavoro sia nella fase di engagement che di assessment.

Il framework proposto è in grado di operare sia su dati strutturati che dati testuali con un'architettura modulare e scomposta in diversi Plug-Ins, ognuno dei quali si occupa di recuperare l'informazione da una specifica sorgente. Ciò permette di aggiungere in maniera più semplice nuove sorgenti o modificare quelle già esistenti. Allo stesso modo il modulo Tender Analyzer conserva la natura modulare al fine di poter sviluppare applicativi specifici ad ogni fonte di dati che si vuole utilizzare.

Questa capacità di aggiornamento ed espansione lascia aperta la possibilità di sviluppi ed aggiunte futuri, qualora la stazione appaltante proponga nuove esigenze e richieda funzionalità aggiuntive sia su dataset già utilizzati nonché su sorgenti di nuovo interesse. Le politiche di digitalizzazione sempre più incisive a livello centrale e periferico e la disponibilità crescente di dati in formato digitale aprono a scenari di utilizzo ed integrazione di strumenti per l'analisi e l'elaborazione assistita dall'Intelligenza Artificiale.

References

- [1] D. Jurasky, J. H. Martin, *Speech and language processing: An introduction to natural language processing*, Computational Linguistics and Speech Recognition. Prentice Hall, New Jersey (2000).
- [2] G. Rossiello, P. Basile, G. Semeraro, Centroid-based text summarization through compositionality of word embeddings, in: G. Giannakopoulos, E. Lloret, J. M. Conroy, J. Steinberger, M. Litvak, P. A. Rankel, B. Favre (Eds.), *Proceedings of the Workshop on Summarization and Summary Evaluation Across Source Types and Genres, MultiLing@EACL 2017*, Valencia, Spain, April 3, 2017, Association for Computational Linguistics, 2017, pp. 12–21. URL: <https://doi.org/10.18653/v1/w17-1003>. doi:10.18653/v1/w17-1003.
- [3] P. Cassotti, L. Siciliani, P. Basile, M. de Gemmis, P. Lops, Extracting relations from italian wikipedia using unsupervised information extraction, in: V. W. Anelli, T. D. Noia, N. Ferro, F. Narducci (Eds.), *Proceedings of the 11th Italian Information Retrieval Workshop 2021*, Bari, Italy, September 13-15, 2021, volume 2947 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021.
- [4] L. Siciliani, P. Cassotti, P. Basile, M. de Gemmis, P. Lops, G. Semeraro, Extracting relations from italian wikipedia using self-training, in: E. Fersini, M. Passarotti, V. Patti (Eds.), *Proceedings of the Eighth Italian Conference on Computational Lin-*

⁵<https://www.registroimprese.it/>

⁶<https://www.giustizia-amministrativa.it/>

guistics, CLiC-it 2021, Milan, Italy, January 26-28, 2022, volume 3033 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021.

- [5] B. Tóth, M. Fazekas, Á. Czibik, I. J. Tóth, Toolkit for detecting collusive bidding in public procurement. with examples from hungary., Report number: CRC-WP/2014:02 (2014).