Information Systems 119 (2023) 102284

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

AI-based decision support system for public procurement

Lucia Siciliani [a,*], Vincenzo Taccardi [a], Pierpaolo Basile [a], Marco Di Ciano [b], Pasquale Lops [a]

[a] Department of Computer Science, University of Bari Aldo Moro, Via E. Orabona 4, 70125 Bari, Italy
[b] InnovaPuglia S.p.A., 70010 Valenzano (BA), Italy

ARTICLE INFO

Article history:
Received 8 August 2023
Received in revised form 7 September 2023
Accepted 8 September 2023
Available online 14 September 2023
Recommended by Dennis Shasha

Keywords:
E-procurement
Data analysis
Data visualisation
Natural language processing
Semantic search
Decision support systems

ABSTRACT

Tenders are powerful means of investment of public funds and represent a strategic development resource. Thus, improving the efficiency of procuring entities and developing evaluation models turn out to be essential to facilitate e-procurement procedures. With this contribution, we introduce our research to create a supporting system for the decision-making and monitoring process during the entire course of investments and contracts. This system employs artificial intelligence techniques based on natural language processing, focused on providing instruments for extracting useful information from both structured and unstructured (i.e., text) data. Therefore, we developed a framework based on a web app that provides integrated tools such as a semantic search engine, a summariser, an open information extraction engine in the form of triples (subject–predicate–object) for tender documents, and dashboards for analysing tender data.

1. Introduction

Public procurement, especially when the aim is innovation, represents a powerful means of investment of public funds. Hence, it is crucial to improve two main aspects in the area of transparency and monitoring of the entire investment and procurement cycle. On one hand, the engagement process of the RUPs,[1] procuring entities, administrations, and awarding entities, allowing them to fulfil their tasks in a more effective, efficient and sustainable manner, and on the other hand, to develop assessment schemes that correlate specific logical-temporal sequences of facts and contents that can be traced back to specific anomaly indicators.

Artificial Intelligence technologies and Natural Language Processing (NLP) systems focused on the Italian language represent a new frontier for semantic interpretation, concept extraction, and correlation of texts and documents. This research, leveraging such technologies, aims at developing a system that can interface with existing databases, prepare datasets that are suitable for processing and analysis, execute automatic extraction of relationships between textual entities, perform correlation tests between portions of text even of different lengths (paragraphs vs entire document), then receive queries and return predefined outcomes in web-based format (short report, evidence, reference code, etc.).

Finally, based on a large amount of structured data from past tenders available as public open data, we provide a data exploration and visualisation tool integrated within the framework to enable users to extract valuable information from such resources.

This informative asset aims to support purchasing bodies during the decision-making process by enabling them to respond to pertinent queries based on prior or related cases and enhancing their knowledge about the participants in tenders.

1.1. E-procurement

The digitisation of the procurement processes of public administrations' goods and services (electronic public procurement) is one of the main drivers of the European Commission's policies; the goal, in the medium term, is to digitise the entire procurement process of public administrations in the two phases of pre and post-award, i.e., from the publication of calls for tenders to payment (end-to-end e-procurement).

E-procurement refers to those technologies that can facilitate the acquisition of goods and services by private organisations or public administrations [1]. Thus, e-procurement can improve the processes' efficiency and effectiveness by targeting simplification and automation [1–3]. The heterogeneity of these processes requires the use of various Information and Communications Technologies (ICTs) in order to lead to the transformation of traditional procuring and supplying of goods and services processes into e-processes such as e-tendering, e-awarding, e-auction, and e-sourcing [2,4]. More specifically, e-tendering consists of the application of ICTs for the dissemination and acquisition of procurement information, the announcement of interest in procurement,

* Corresponding author.
E-mail addresses: lucia.siciliani@uniba.it (L. Siciliani), vincenzo.taccardi@uniba.it (V. Taccardi), pierpaolo.basile@uniba.it (P. Basile), m.diciano@innovapuglia.it (M. Di Ciano), pasquale.lops@uniba.it (P. Lops).

[1] Responsabile Unico del Procedimento, *Head Project Manager*.

https://doi.org/10.1016/j.is.2023.102284

the receipt of tender documents, the submission of bids, and the final selection of the procurement bid [5]. Indeed, e-tendering aims to increase productivity in the management of tenders by relieving the bureaucracy and speeding up communication between the parties involved. Essentially it seeks the shift from paper-based methods to ICT-based means of communication. One of the main strengths of e-tendering is the remote accessibility of the system. In this way, the tender manager, bidder, contractor, or customer can freely access the tender management platforms from anywhere in the world without being restricted by location constraints [6].

With the *Open Data Directive*[2] the EU mandates the release of public sector data in free and open formats. The overall objective of the Directive is to continue to strengthen the EU's data economy by increasing the amount of public sector data available for re-use, ensuring fair competition and easy access to public sector information, and enhancing cross-border innovation based on data. Indeed, adopting e-procurement solutions causes a beneficial side effect: the generation of large amounts of digital data and, thus, the opportunity to leverage them to develop innovative IT applications that can improve both government agencies' and bidders' processes.

1.2. Motivation and scenarios

This project has been carried out in collaboration with *Regione Puglia*,[3] and *InnovaPuglia S.p.A.*[4] whose feedback and requirements have been the reference for defining possible use cases and thus the features. Our partners' main objective was to develop a system that enables procurement agents to access a collection of data in an easy-to-use, integrated manner. This could allow gathering pertinent information about tenders, bid documentation, similar cases from the past, or about companies involved in the bidding process to enhance evaluation and decision-making tasks. Therefore, we believe that our platform represents an effective answer to the above scenario since it streamlines access to various information sources and enables searching through structured and unstructured data to gather insightful information. We developed a decision support system to help procurement organisations through the entire investment and contract life cycle. The main features and contributions of our work are the following:

1. Integration of structured and unstructured data information;
2. Semantic search engine for tenders documentation;
3. Search engine based on OIE (Open Information Extraction) techniques;
4. Visualisation and analysis of structured procurement data;
5. Single access point for companies' information gathered from several sources;
6. Set of collusion risk indicators.

2. Background and related work

Decision Support Systems (DSS) are defined as interactive computer-based information systems designed to support solutions on decision problems [7]. To frame our approach, it is helpful to refer to the following DSS classification [8], which splits DSS into six main classes as shown in Table 1.

Table 1
Decision support systems classification.

DSS Type	Description
Model-driven	Based on simple quantitative models. Large databases are not required because they work with the limited information and parameters that decision-makers provide.
Data-driven	Access and manipulate internal corporate data. The development of *data warehouse*, as a subject-oriented, integrated, time-variant, nonvolatile collection of data [9], as well as On-Line Analytical Processing (OLAP) solutions [10], has resulted in the spread of this paradigm. Business Intelligence (BI) was established under this area, incorporating a broad category of applications, technologies, and processes for obtaining, storing, accessing, and analysing data [11].
Group Communications-driven	Utilise network and communication technology. Groupware, video conferencing, and computer-based bulletin boards are among the tools employed.
Document-driven	Provide document retrieval and analysis. Based on large document databases may include scanned documents, hypertext documents, images, sounds and video. A search engine is the primary decision-aiding tool.
Knowledge-driven	Knowledge is built on computerised systems that can store and retrieve knowledge codified as probabilities, rules, and relationships through the application of data mining and Artificial Intelligence (AI) technologies.
Web-based	All of the above can be implemented using Web technologies accessible through a web browser. The server is linked to the user's computer by a network using the Transmission Control Protocol/Internet Protocol (TCP/IP).

Although a few examples of DSS developed for public procurement are available, none of them is intended to aid public agencies throughout the entire tendering process by making holistic use of the bulk of information already in existence. Some concentrate on a particular industry [12–14] or on particular stages of the procurement pipeline, such as bidder selection [15–17], and contractor pre-qualification [18,19]. These are built around a single DSS type, typically model-driven. Rather, in this article,

[2] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN

[3] https://www.regione.puglia.it/

[4] https://www.innova.puglia.it/

we suggest a hybrid DSS that comprises some of the aforementioned types: data-driven, document-driven, knowledge-driven, and web-based.

A relevant application for our research is described in [20], where a DSS for fraud detection in public procurement is proposed. The authors defined a series of corruption risk patterns based on data mining and calculated indicators. As highlighted in [21], defining objective criteria is a crucial step in improving the identification of fraudulent behaviour and overcoming the limitations of the majority of indicators currently used, which are based on surveys, audits, perceptions or experiences of corruption among various stakeholders (e.g. general population, firms, experts). Among them, the two that are most frequently used are the Transparency International's Corruption Perceptions Index [22] and the World Bank's Control of Corruption [23]. The main associated issues are determined by biases due to subjective perceptions, which may or may not be related to actual experience and may also be influenced by general sentiment. Additionally, these indicators can result from surveys that are skewed by small and unrepresentative samples of the entire population. [24]. We decided to adapt the indicators presented in [25] since they are based on data whose schema resembles the one that is available to us.

A survey of the most recent research on fraud detection for public procurement is presented in [26]. The paper details the most common approaches based on machine learning algorithms, network analysis, and the growing interest in neural networks.

An asset for an effective DSS is the adoption of data visualisation solutions like, for example, Business Intelligence (BI) dashboards which provide collections of multiple visual components, such as charts, on a single view so that information can be monitored at a glance [27]. Indeed, to enhance information interpretation and amplify human cognition [28], appropriate visual representations in dashboards that use colour, scale, and shape are paired with interactive exploration [29].

A review of the application of visualisation tools applied to Open Government Data (OGD) is proposed in [30,31]. These studies aim to identify the specific government areas served by such tools and the most popular data visualisation techniques. Moreover, they highlight the challenges frequently reported by other researchers and users and evaluate usability.

In these papers [32,33], the authors use visualisation techniques on OGD to make it easier to access and retrieve information. They draw attention to the fact that, despite the government datasets are now widely available, their use remains impeded by the stakeholders' lack of programming skills and data management knowledge. To ease this problem, data visualisation is suggested as a solution for interacting, sharing and understanding data.

In [34], a literature review on the theme of AI applied to the public sector. Authors included within this survey a total of 73 publications. Most publications have been published since 2019. Most of them (42 out of 73) are conceptual/literature reviews addressing AI's benefits and overall effects in the public sector. Finally, [35] is another literature review which is focused on e-procurement.

3. Collusion risk indicators

As introduced in Section 1.2, one request formulated from our partners was to gather information about companies participating in tender bids. In light of this, it would be relevant to derive metrics specific to each company that can detect anomalies or deviations from regulatory and normative standards in procurement activities.

As a result, a portion of our research was devoted to calculating such metrics; therefore, before illustrating the full framework architecture, this section describes the *Collusion Risk Indicators*, which can be calculated from the available Open Data on public procurement. We implemented most of the indicators suggested in [25], though some had to be slightly modified to fit our dataset; otherwise, if a specific type of data was unavailable, we omitted the related indicator. We calculated the following indicators:

Relative tender value. - It is calculated as the ratio of the winning bid to the price estimated in the tender. Since issuers generally expect bid prices to fall below the estimate due to healthy competition, the relative value of the tender can be seen as an indicator of how expensive the tender actually became (assuming, of course, that the initial estimate was not biased). An increase in the relative value of the tender may signal a collusive behaviour, as this may ultimately result in noncompetitive, increased prices. Similarly, values close to the estimated price could indicate market problems.

Variance of bids. - because specific values for all bids submitted during the tender process are not available in our data for individual tenders, a good proxy, correlated with the variance of the bids, is the difference between the highest and lowest bid. Each contractor's expected value (average) is calculated for all tenders won. We can observe from this data that a value of this indicator that is too high or too low can indicate a problem. The presence of intentionally too-high bids could cause the first case, while abnormal rigidness in the market causes the second.

Missing offers. - the absence of bids from a previously active company in a given market could indicate the presence of collusive scenarios. For each tender won by the contractor, it is given by the expected value (average) of the inverse of the number of bids admitted to the tender. A high value could imply a lack of competition.

Superfluous bidders. - the prevalence of incorrect bids indicates the excessively high proportion of bids excluded for administrative reasons, e.g., missing documentation. Competition in procurement markets can be simulated by competitors submitting deliberately incorrect bids. Such artificial, non-competitive bids may contain errors to be excluded, leaving only predetermined companies with considerably high prices as the only valid proposals. Since excluding bids for administrative reasons is widespread in procurement markets, submitting erroneous bids may give the impression of competition, misleading contracting authorities. Although it is natural to make mistakes in the bids submitted, when this ratio is systematically high or is associated with higher prices, the suspected likelihood of collusion between bidders is greater. For each contractor, the expected value (average) between the number of excluded bids and the number of companies that submitted a bid is calculated for all tenders won by that contractor. A high value could be an indication of intentionally submitted erroneous bids.

Concentrated market. - one of the main outcomes of collusive bidding is that the market structure concentrates on a few players rather than having a competitive structure involving several players. A concentrated structure occurs when only one or a few companies win all the tenders while the other bidders are either completely absent or fake their participation. If this occurs in a market that would otherwise be competitive, it is a sign of collusive behaviour. A situation where market concentration is definitely a sign of collusion is when a particular market goes from competitive to concentrated in a short period of time without any apparent alternative explanation (e.g., a change in regulations, technology, or a sharp drop in total demand). It is calculated as the specific company's market share.

Static market structure. - a stable market structure indicates very little variance between market shares. Here we examine whether or not the company's market shares are stable over time. It is calculated as the coefficient of variation; a low value indicates a stable market over time.

Prevalence of consortia. - it indicates that the winning bids were predominantly submitted by a group of companies in consortia. The formation of a consortium can undoubtedly increase efficiency if this allows the specific expertise of the various consortium members to be exploited. However, joint bids reduce the actual number of competing parties, which can decrease the effective competitive pressure, even in a non-collusive configuration. For each contractor, the ratio of the number of bids won by bidding as a consortium to the total number of times it has won a bid is calculated. A high value signals the prevalence of consortium wins and may suggest a lack of competition caused by collusion.

Prevalence of subcontracting. - it indicates the involvement of subcontractors in the handling of the contract. Similar to joint bidding, the use of subcontracts can also have beneficial effects and increase efficiency. However, it is also a convenient way to share profits among colluding sides and can also serve as a means of guaranteeing against possible defeats in the contract award. For each contractor, the ratio of the number of bids won in which subcontracting is allowed to the total number of bids won by that contractor is calculated. A value that is too high may suggest a willingness on the part of the contractor to adhere to collusive arrangements via profit-sharing through subcontracting.

The derivation of such indicators can be interpreted as a *feature engineering* activity on the original dataset that could enhance the application of machine learning, with the aim of detecting suspicious contracts whose allocation may be the result of collusive agreements among firms participating in the tender or pertaining to that market. The main obstacle for the above task is the lack of datasets that record the occurrence of a judicial authority investigation for a given procurement that has proven the collusion among participants.

4. System implementation

In this section, we will discuss the system architecture illustrated in Fig. 1 and introduce the main functionalities implemented within the framework. The system is freely accessible online.[5] The architecture is organised into four main modules:

- Data Collector: collects and integrates data coming from different sources;
- Pre-Processing: extracts the relevant features of the data collected in the previous step and stores it in a database to allow further analysis;
- Tender Analyser: performs specific analysis on the data collected in the previous step exploiting Machine Learning and Natural Language Processing techniques;
- Service Tools: each service tool implements a specific use case.

As we can see from Fig. 1, each module can be further divided into sub-modules. This design choice was done taking into account modularity so that the system is open to changes or future developments. The following sections will thoroughly describe each component, along with the design choices taken for their fulfilment.

[5] http://www.semanticframework.it:8080/SearchMetadata/

4.1. Data collector module

The *Data Collector* module collects data about tender notices that may be retrieved from different databases. The datasets included are the following:

- ANAC[6](Autorità Nazionale Anti Corruzione - *National Authority Against Corruption*), which is the Italian national anti-corruption authority, and stores the essential information on Italian public procurement contained in the National Database of Public Contracts from 2007 to present, in the form of structured data;
- EmPULIA,[7] is an online platform for public administration tenders in the Apulian Region, and it stores documentation inherent to the tenders issued by Apulian contracting agencies.

Due to the heterogeneity of the data sources, data extraction is divided into plug-ins, making it easier to add new sources or modify those that are already included.

In particular, the first plug-in extracts the data available on the ANAC website, which is available as open data. All the Open Data on the ANAC platform is structured into several tables and can be obtained by querying its CKAN[8] APIs. The download application was developed in Java with the help of the Springboot framework. This procedure allowed us to collect a total of 7,017,055 tender records in about 4 h.

For what concerns our other data source, i.e., EmPulia, unfortunately, the platform did not make available some dedicated APIs, so with the authorisation of InnovaPuglia, which manages the website, we developed a crawling system that was capable of automatically extracting the data contained within the platform. This aspect is to take into great consideration since, despite the efforts made until now to have open access to information regarding the public sector, many platforms still lack proper access points.

Within the same module, we have the Data Integration submodule. This specific component addresses the aforementioned heterogeneity of the data extracted by the different data sources and merges overlapping data where possible.

4.2. Pre-processing module

The *Pre-Processing* module takes the data collected in the previous phase and memorises it using databases or indexes. It is important to select the relevant features necessary to build a unified view of a tender allowing the subsequent analysis tasks. Although numerous tools for textual analysis are available in the literature, given our domain of interest, it is necessary to focus on techniques that allow us to process texts written in different languages and deal with a very specific vocabulary.

The unstructured data we processed is currently extracted from the EmPulia platform. We want to process the documents' metadata and the attachments related to a tender. The number of such attachments is not fixed: it varies based on how many are loaded onto the system by the RUP or the delegates dealing with platform interaction. The unstructured data extracted from EmPulia will be stored separately and indexed using the CIG (Codice Identificativo Gara - Tender Identifier Code) of the call to which they refer. The data is available in the following formats: doc/docx, pdf, p7 m. Unfortunately, several documents contain images of the scanned pages of the documents instead of the text itself, making them unreadable by an automated system. For this

[6] https://dati.anticorruzione.it/opendata

[7] http://www.empulia.it

[8] https://ckan.org/

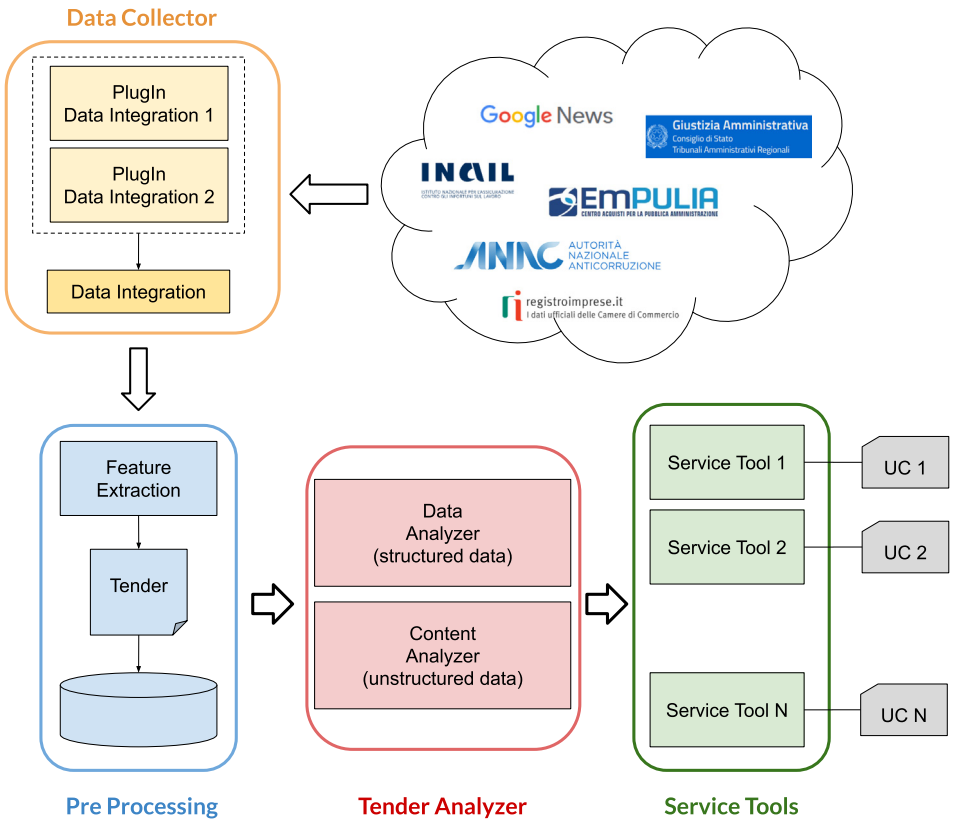

Fig. 1. High-level framework architecture.

reason, it is necessary to convert these files to a format containing only text. This goal is achieved using layout extraction and Optical Character Recognition techniques.

To extract the text, we employ the 5th version of Tesseract OCR,[9] released on November 2021. Tesseract OCR is available in several languages, which can be easily downloaded.[10] In particular, we use the `ita.traineddata` Italian model.

To store structured data, an SQL database was implemented. The data obtained from ANAC was in the form of CSV or JSON files split into several tables, each focused on a specific subject. They are indeed linked by keys, allowing the user to join and extract a set of desired columns from across the tables. The database (DB) was created by importing those datasets using the same source schema because it was already reliable and effective for structuring our DB. The following step involved defining calculated columns and queries as required for calculating collusion indicators (Section 3).

4.3. Tender analyser module

The *Tender Analyser* carries out the analyses of the data memorised in the Pre-Processing step. This module is divided into two distinct sub-components: the Data Analyser and the Content Analyser. The first one handles the structured data related to tenders like codes (e.g., CPVs), dates, and quantities; the latter, instead, elaborates unstructured data, such as any attachment that is part of the tender notification (e.g., equipment supply, the main construction contract, specifications).

The core of the Content Analyser is the Natural Language Processing component (NLP pipeline), which feeds both the Search Engine and the OpenIE components. The current pipeline is an evolution of a previous framework [36] that we developed for the public administration of Apulia Region. The pipeline performs several text-processing steps for English and Italian:

- Sentence Detection: splits a text into sentences by exploiting punctuation characters that mark the end of a sentence.
- Tokenization: splits the text into tokens. Each token is a word.
- Part-of-Speech (POS) tagging: identifies the grammatical role of each word: noun, verbs, adjective, adverb, punctuation, preposition, and so on.
- Lemmatization: provides the lemma for each word. The lemma is the basic form of a word, for example, the singular form of a noun or the infinitive form of a verb, as shown at the beginning of a dictionary entry.
- Chunking: divides a text into syntactically correlated parts of words, like noun or verb groups, but it specifies neither their internal structure nor their role in the main sentence.
- Phrase Extraction: is able to find n-grams (sequence of words) that identify a single concept. Examples of n-grams are "Information Retrieval", "Document Management", and "Public Administration".
- Random Indexing [37]: constructs a WordSpace by analysing a collection of documents. A WordSpace is a geometrical space in which words are represented as points. If two words are close in the WordSpace, they are semantically related. RI allows to perform retrieval and semantic search by exploiting a dense representation of both queries and documents.

The OpenIE component is devoted to extracting structured triples from text. We rely on an open-domain approach since we do not know the predefined relations set as in the classic relation extraction task. We combine two approaches: the former extracts

9 https://github.com/tesseract-ocr/tesseract
10 https://github.com/tesseract-ocr/tessdata

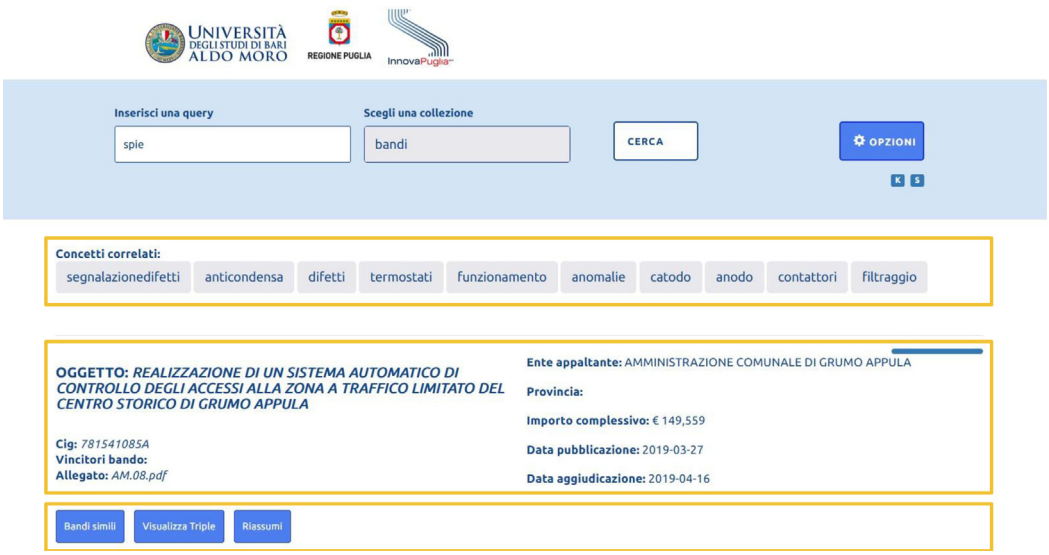

Fig. 2. Smart Search — An example of result for the query: "indicator lights".

triples using an unsupervised method [38] based on syntactic rules, while the latter filters triples in relevant and not relevant using a supervised system [39].

For example, given the following piece of text extracted from a tender's attachment: *"The legal subject delegated to carry out the inspection cannot be commissioned by more than one competitor. The contracting station issues a proof certificate of inspection"*., the NLP pipeline is able to identify two sentences. For each sentence, the list of tokens and lemmas is extracted. The chunking module is able to identify noun phrases such as *"contracting station"* or verb phrases like *"carry out"*. The phrase extraction can automatically identify relevant concepts such as *"certificate of inspection"*. Finally, the triple extraction module derives the following triple: *"The contracting station"* as the subject, *"issues"* as the predicate, and *"a proof certificate of inspection"* as the object.

4.4. Service tools

The *Service Tools*, as the last module of the architecture, is designed as a collection of applications. They will be tied to a well-defined set of use cases and carry out specific tasks to satisfy the requirements in light of the data and analysis supplied by the earlier modules.

Finally, given the high-level schema overview, the following is a list of the key functionalities developed to implement the use cases requested by our partners:

1. Semantic search engine for tenders and their documentation:
   (a) Summarising tool;
   (b) Semantically related concept suggestion tool;
2. Search engine based on Open Information Extraction;
3. Company's information aggregator;
4. Data analysis and visualisation dashboards for tenders data:
   (a) Collusion risk indicators.

The features listed above will be described in detail in the following sub-sections.

4.4.1. Semantic search engine

The search engine (Fig. 2) can retrieve documents based on the user's query. More specifically, the search engine ranks documents by calculating a relevance score between the user's query and the document. The relevance score can be calculated using different approaches. In the current version of our platform, we implement two methodologies:

Classical search: this approach is based on the Vector Space Model [40], in particular the BM25 model [41]. This method can only retrieve documents that contain at least one of the keywords provided in the user's query;

Semantic search: this approach can map both documents and queries as points of the same WordSpace built by Random Indexing and used to represent words [42]. After this mapping, it is possible to calculate the cosine similarity between each document and the user's query and then rank the documents according to their similarity. This approach allows finding documents that do not contain the keyword provided in the query. The distributional space, which is the key component of the semantic search, is built by the NLP pipeline through Random Indexing (RI).

All documents are indexed using the BM25 model implemented by *Apache Lucene*.[11] We allow the user to choose which kind of search the system will perform by selecting one or both. When both types are selected, the results of each search model are combined in a unique results set [42,43]. Whichever type of search is selected, the results of the queries will be tenders or tenders' attachments that have a high relevance score to the initial request.

One interesting feature the system offers is the possibility to search for documents or notices similar to another one obtained as a result of an initial query. This kind of search is performed by measuring the cosine similarity between the dense vector representation of documents. Given a document vector of a document, we rank all the other documents according to the similarity between document vectors. The related concepts section represents another functionality offered within the search engine. Given the list of keywords specified by the user, being able to identify other related concepts can be particularly important. Related concepts are computed using the same approach of document similarity but using the dense vector representation of terms. In fact, this allows helping the user to satisfy his information need by extending the search with words very similar to those they inserted but which they did not think of. The related concepts'

[11] https://lucene.apache.org/core/

Fig. 3. Smart search — Tender winner tab.

functionality is particularly useful when a RUP seeks information in a field outside their expertise. For example, inserting as query *"computer memory"*, the system returns as related concepts the following phrases: *"portable computers"*, *"processor"*, *"gb"*, *"ram"*, *"software"*, *"electronics"*, *"quad"*, *"memorisation"*, *"microelectronics"*. If we consider a case where the RUP has no knowledge of computer components, this kind of functionality can help bridge the vocabulary gap.

A sample of search results is shown in Fig. 2. Three frames are highlighted and denoted on the interface by yellow boxes, each with a distinct aim. In the first one, a collection of *correlated concepts* shows to the user other pertinent inquiries to enhance and broaden the capabilities of information retrieval. The second frame contains the most relevant information about the returned results for the query. The fields on the left side are clickable and connected to specific functions:

- Oggetto (Object) and *CIG*: return the tender object along with identification code. By clicking on them, it is possible to access a panel with more available contents divided into tabs (Fig. 3):
  - *Dettagli* (Details): contains additional selected tender info;
  - *Vincitori Bando* (Winner of the tender): contains tender winner data. From this tab, by clicking on the VAT[12] field is possible to access the Company Data Aggregator (see Section 4.4.3);
  - *Documenti* (Documents) contains whole tender documentation, if available;
  - *Ulteriori Informazioni* (Further Information) leads to a new page (Fig. 4) that returns all of the available tender data stored in the database (see Section 4.2);
- *Vincitori* (Winners): clicking on it gives direct access to the Company Data Aggregator;
- *Allegato* (Attachment): it displays the retrieved document whose content is most pertinent to the query.

Finally, in the third frame, there are three buttons that enable the user to access additional features: (1) retrieve other similar tenders, (2) get triples extracted from the text (Section 4.4.2), and (3) a document summariser based on compositionality of word embeddings [44].

4.4.2. Triples search engine

The triple search engine allows the user to navigate among the triples extracted from the attachments of the tender notices taken from EmPulia. The user can insert their query in the research field and click the search button to start the retrieval procedure. The retrieval model is already based on Lucene-BM25 and indexes each triple as a document composed of three fields: subject, object and predicate. By default, the triples search engine looks for triples where the keywords inserted by the user appear as subject, predicate, or object. Nevertheless, it is possible to change this configuration by opening the options menu: this allows searching in a single field or any combination of the three. Here, the user can limit the number of results to a specific value. The triple search engine allows the retrieval of relevant information from the text in the form of relations.

In addition to the standalone component, the functionalities of the triple search engine are also integrated within the Semantic Search engine (Fig. 5): The user can see all of the triples (soggetto - predicato - oggetto) that belong to that specific document, as well as the singular phrase (frase) that comprises them, in the result list. This option is only possible if the result is a tender whose information is available on EmPulia since triples can only be extracted from documents attached to the tender.

These functionalities offer users a new way to explore data, which can be more efficient and help them find novel information in a wide number of domain-specific documents. Moreover, applying OpenIE to tender notices represents a first step in transforming the huge amount of unstructured data available in the public sector into a structured format. This could help to automatically extract databases or knowledge graphs containing information about tenders, making handling procedures more efficient and helping RUPs in their daily activities. Moreover, in this form, it would be easier to double-check the information about each tender avoiding errors or highlighting anomalies which can indicate the presence of collusive behaviours.

4.4.3. Company data aggregator

This framework component enables the user to collect various useful facts about a company by retrieving data from a set of sources. All of the collected data is exposed in the user interface (UI), creating a sort of "one-stop shop" for accessing forms of diverse information that lets users have a more thorough understanding of a company. The list of resources used is provided below:

- *Registro Imprese* (Business Registry): is a company record that contains the data for all enterprises with headquarters or local units in the Italian territory. It contains all of the essential information about companies (name, articles of association, management, headquarters, etc.) as well as all subsequent events that affect them after registration (e.g., changes in the articles of association and company officers, transfer of registered office, liquidation, bankruptcy proceedings, and so on). The Companies Register, therefore, gives a detailed picture of each company's legal state. It provides an API to access its database;
- *News*: this API gathers news articles mentioning the selected company in the Italian media over the past five years;
- *Giustizia Amministrativa* (Administrative Justice): this website collects all administrative justice-related lawsuits and makes them accessible to users. Although the data are freely available, the supplier does not offer an API to automatically

[12] https://en.wikipedia.org/wiki/VAT_identification_number

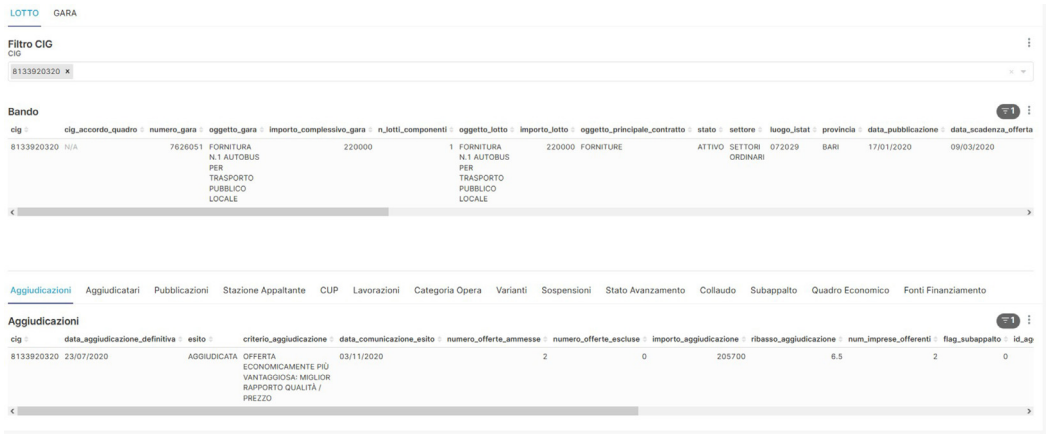

Fig. 4. Smart search — Tender data page.

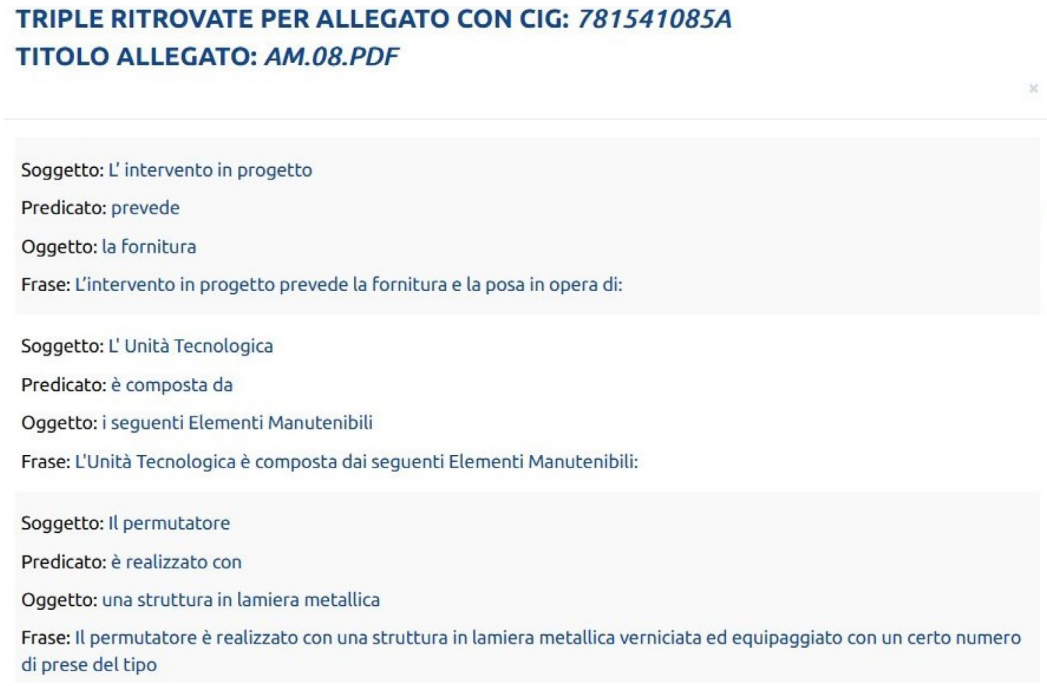

Fig. 5. Smart search — Example of triple extraction from a document.

retrieve them, so we came up with a workaround by setting up a bot that mimics user behaviours on the website to submit the necessary query first and then collect the responses returned on the same page;

- *Analytic* (Analisi): this resource is linked with the Data Analysis and Visualisation tool. More details are in Section 4.4.4.

This component can be accessed in one of two ways: either as a standalone service from the system home page, which exposes to the user a form to enter the requested query or through the search engine results by clicking on the dedicated field (the contractor name), which automatically passes to the module the relevant query extracted from the results: Giustizia Amministrativa and the News module use the company name as the default query, Registro Imprese uses the VAT number.

4.4.4. Data analysis and visualisation

This service tool is connected to the one described in Section 4.4.3 and has the goal of visualising and analysing structured data available in the ANAC dataset. For example, it has the scope to provide an analysis of the historical records of a company in the area of public procurement.

Public Open Data typically have a major drawback: the way they are distributed makes it difficult for non-technical end-users to access and understand vast amounts of data. Indeed, most data are provided as raw datasets, which creates a technical barrier that restricts their use or even their access [32].

Hence, the main focus is to provide a simple data access point that enables the user to skip all the preliminary procedures, including the collection of appropriate data, loading and reading them with proper tools, searching for pertinent information, and finally visualising the findings. Due to these factors, we opted for using a BI software solution, namely the open source *Apache Superset*[13] application. One of its remarkable features is that the app front-end provides a web-based UI with all of the back-end components installed on a server, for example, as a docker container. This allowed us to smoothly integrate it within our

[13] https://superset.apache.org/

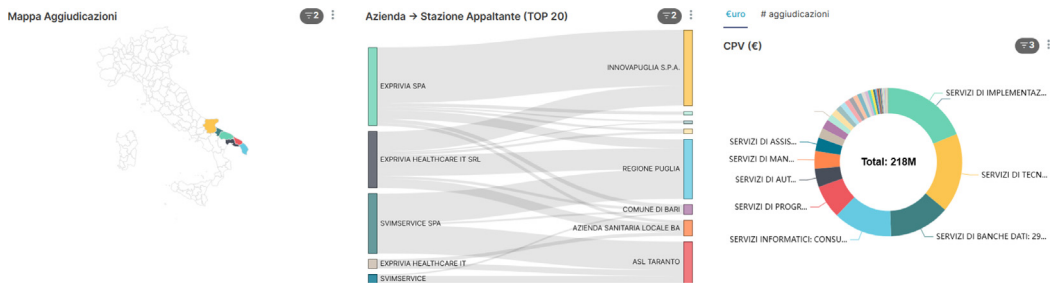

Fig. 6. Company's Dashboard — Map, Sankey diagram, CPV pie chart.

framework; in order to access the functionality, we just had to link its service URL inside the platform.

After establishing a connection, through the software connector, between Superset and the database implemented in the pre-processing module (Section 4.2), we defined a number of visualisations that are useful for information mining. These are accessible via dashboards, each of them tailored to a specific area of interest. Two are the main dashboards: one about companies' records related to previously awarded tenders and another about a specific market that can be defined based on geographic areas (region, province, etc.), date and specific business sectors identified by *CPV* [14] (Common Procurement Vocabulary) codes. Such codes define a unique classification system for public procurement aimed at standardising the references used by contracting authorities and entities to describe the subject matter of procurement contracts. The basic vocabulary is built on a tree structure that includes codes of up to nine numbers that are associated with text that specifies the supplies, works, or services covered by the contract.

There are two options to access these dashboards: as a standalone service that the user may utilise to obtain information on demand or from a link available in the search engine results (Section 4.4.1) from the tender winner's field, which is then, preselected in the dashboard filter. A more detailed description of the dashboard content is provided.

For the **company data** dashboards, we propose the following charts:

1. Filter modules that allow the user to select the desired company by corporate name or VAT number, as well as another that limits the gathered data on date and location with increasing granularity, such as region and province;
2. A table contains all of the available companies in the DB whose names begin with the same string selected in the name filter, or even if the names do not match exactly, have the same VAT codes;
3. Two counters measuring the total number of contracts awarded and the total amounts for them;
4. A map showing the locations of the won tenders (Fig. 6);
5. A Sankey flow chart connecting the company to the most relevant purchasing agencies that issued the awarded bids (Fig. 6);
6. A pie chart displaying the percentage of contracts categorised by the CPV based on the number of tenders or the total value (Fig. 6).
7. A time-series chart of the contracts awarded, with the corresponding values, and categorised according to the awarding criteria;
8. A pie chart displaying the main competitors in the same company's business areas, ranked by market share, as well as an associated table detailing the markets with respective winners;
9. A bullet chart with the values of each collusion indicator (Section 3), and a spider chart with a global overview of all indicators (Fig. 7);
10. A table containing the most relevant information about all of the company's won tenders. By clicking on the CIG in the corresponding field, it is possible to access another dashboard that contains all the exhaustive data gated from the database relative to the selected tender. In the same way, clicking on the company's fiscal code field lets the user retrieve information about the firm by accessing the system's section that collects data from a variety of sources like news, administrative-juridical issues, and the business register, as described in the corresponding paragraph 4.4.3.

The **market** data dashboard is comprised of the following charts:

1. A set of filters to define the market configuration: date, region, province, procuring authority, the business sector (CPV), award procedures, as well as a table with the entire CPV hierarchy;
2. A pie chart depicting the largest market contractors in terms of the number of contracts and total amount, arranged by tabs;
3. A time-series chart of the last 50 contracts awarded;
4. Divided into three tabs as many pie charts illustrating the number of tenders with respect to the following subjects: contractor selection criterion, awarding criterion, and primary contract issue (Fig. 8);
5. A tile chart that divides the market by each CPV and its subclass, arranged by the total amount with respect to the CPV;
6. A graph chart linking each procurement agency active in the targeted market with its key suppliers, split by the number of assigned tenders or total granted amount (Fig. 9);
7. A table containing the most relevant information about all of the tenders awarded in that market. In the same way, as in the company's information dashboard, additional details regarding tenders and firms may be retrieved via the links in the CIG and fiscal code columns.

5. Evaluation

In order to evaluate our decision support system, we set up an online assessment survey. We gathered 33 participants from Public Administration employees aged between 20 and 70 years, equally distributed by gender. The chosen sample considers final users of the proposed DSS with high familiarity with the domain of public tenders (RUPs, Project Managers, e-procurement specialists, etc.). We provided the participants in advance with tutorials in the form of videos directly accessible in a dedicated section on the system website. We then asked them to perform some tasks and finally evaluate the system.

14 https://simap.ted.europa.eu/web/simap/cpv

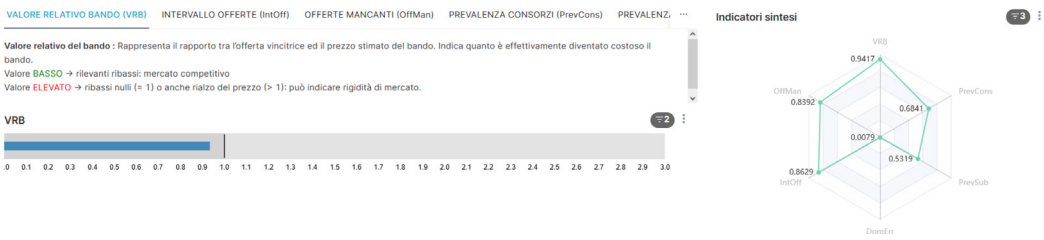

Fig. 7. Company's Dashboard, collusion indicators.

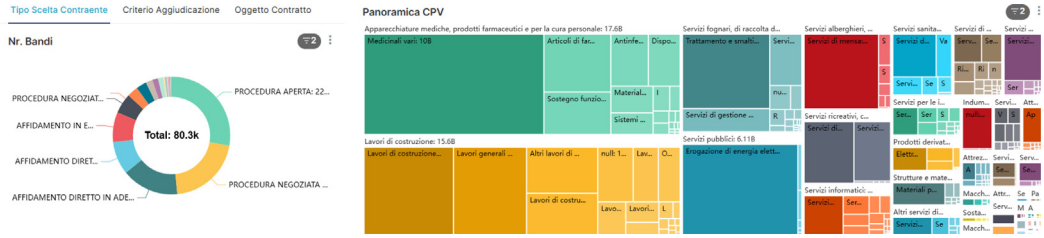

Fig. 8. Market Dashboard, Contract types pie charts, Tile diagram.

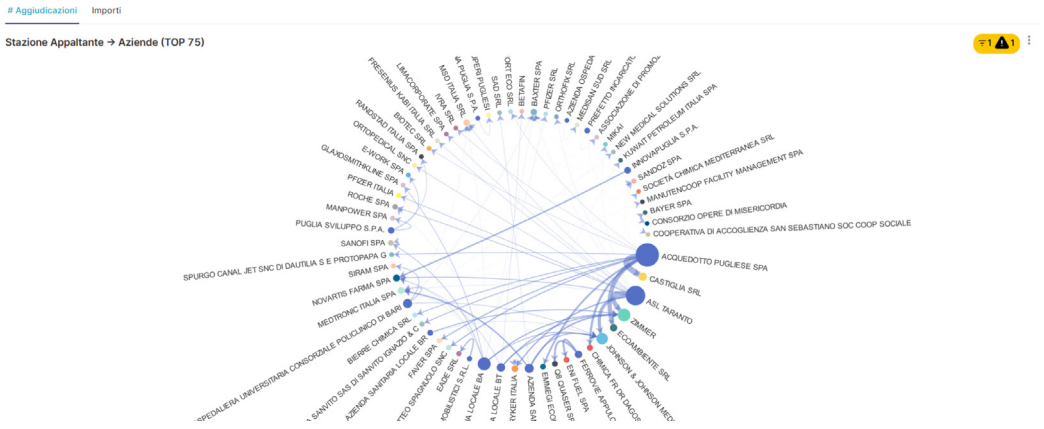

Fig. 9. Market Dashboard — Graph chart.

The survey presents a total of 19 questions divided into 5 groups as displayed in Tables 2, 3, and 4. The first group consists of 5 questions concerning participants' information (Fig. 10). The second group is composed of 10 questions that are structured into a questionnaire to assess our platform's usability, and it provides five different response options on a scale from one to five (1 — strongly disagree, 5 — strongly agree). We used these answers to compute the *System Usability Score*, a measure of a user's perception of the system's usability. To obtain feedback from participants about which of the implemented features could mostly support their work, we added 2 more questions in the third group, asking them to select from a list of the available functionalities which one is most and least useful. Moreover, we also included an optional open-ended question to gather detailed feedback on the system, participants' experiences, and their suggestions for improvements. The last question allows us to compute the *Net Promoter Score* [45], a metric employed to evaluate the likelihood the users would recommend a product, a service, or software.

According to the standard ISO 9241-11, usability can be measured in terms of system effectiveness, system efficiency, and system satisfaction. Created by John Brooke in 1986, the *System Usability Score* [46] (SUS) proved to be intuitive and solid over hundreds of studies and nowadays, the SUS is widely used to measure the usability of websites and applications [47]. The survey consists of 10 questions (see Table 3) and 5 rating options, the 5-point Likert Scale [48]. The SUS score for each survey participant is computed as follows:

$$SUS = (((\sum odd\_items) - 5) + (25 - \sum even\_items)) * 2.5$$

and can assume values between [0, 100].
Odd_items (Q.2.1, Q.2.3, Q.2.5, Q.2.7, Q.2.9) and *even_items* (Q.2.2, Q.2.4, Q.2.6, Q.2.8, Q.2.10) are the scores assigned literally to odd and even numbered questions in the questionnaire.

As illustrated in Fig. 11, the individual SUS scores range between 25 and 100. Thus, averaged on the number of participants, the SUS is equal to 77.1, considerably above the margin of the acceptable range, which the guidelines [49] state to be 68.

Fig. 12 shows the distribution of the responses to each of the ten questions. Considering the odd answers, it can be affirmed that users found the proposed tool convenient and straightforward. This observation is backed up by the responses to even-numbered questions, negative-toned by definition, which predominantly gather around low values (1–2).

For an exhaustive evaluation, in addition to a numerical estimate, we asked the participants which features of the search engine they considered the most useful and which the less one. Finally, we asked what we should improve through an open-ended question.

The results shown in Fig. 13 provide remarkable insight from a user's perspective. At first glance, we notice that the most useful

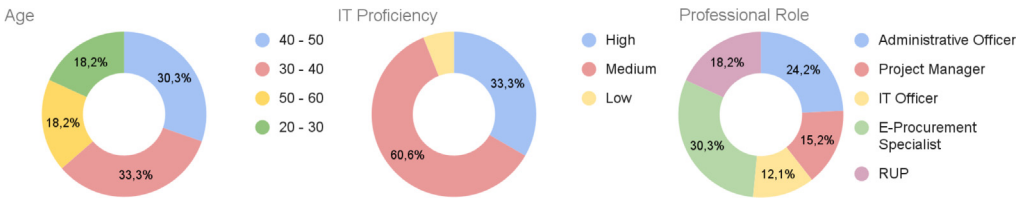

Fig. 10. Evaluation survey — Age, IT proficiency, and Professional role of participants.

Table 2
Questions related to the collection of personal information about the user.

ID	Question	Type of answer
Q.1.1	What is your age?	Open-ended
Q.1.2	What is your educational qualification?	
Q.1.3	What is your job?	
Q.1.4	How many years have you been in this profession?	
Q.1.5	What is your IT proficiency?	

Table 3
Questions for the SUS questionnaire.

ID	Question	Type of answer
Q.2.1	I think that I would like to use this system frequently.	5 point likert scale
Q.2.2	I found the system unnecessarily complex.	
Q.2.3	I thought the system was easy to use.	
Q.2.4	I think that I would need the support of a technical person to be able to use this system.	
Q.2.5	I found the various functions in this system were well integrated.	
Q.2.6	I thought there was too much inconsistency in this system.	
Q.2.7	I would imagine that most people would learn to use this system very quickly.	
Q.2.8	I found the system very cumbersome to use.	
Q.2.9	I felt very confident using the system.	
Q.2.10	I needed to learn a lot of things before I could get going with this system.	

Table 4
Final questions. Questions 3.1, 3.2, and 4.1 are aimed at collecting detailed feedback from the user about the system, while the final question is necessary to compute the NPS score.

ID	Question	Type of answer
Q.3.1	Based on your needs, which feature do you find most interesting/useful?	List of options
Q.3.2	Based on your needs, which feature do you find less interesting/useful?	
Q.4.1	How could we improve our website?	Open-ended
Q.5.1	How likely is it that you would recommend this system to a friend or colleague?	Score from 1 to 10

tools are almost evenly divided between the Semantic search and various types of dashboards. This supports our decision to include in the DSS the ability to gather information from both structured and unstructured data. We believe that such a dichotomy is related to the various participants' duties and roles at work.

Meanwhile, the second question provided a clear response as to which tools people find least helpful. In fact, 57.6% of study respondents believe *Indicators Dashboard*[15] is unhelpful. Given the magnitude of the outcome, we decided to drop this component of the architecture entirely.

[15] This dashboard provided an exhaustive distribution of the values estimated for each collusion indicator in certain market selection.

Table 5
Answers to the open question "How could we improve our system?".

How could we improve our system?
I would enhance the market dashboard filter section by including a choice to filter by contracting type and to filter CPV using its code rather than just textual description.

Table 5 contains the responses to Q.4.1 intended to provide feedback and suggestions to improve the system. Unluckily, only one user added such an answer, formulating the request for some improvements in the filtering section in the Market Dashboard. The DSS has received such minor upgrades.

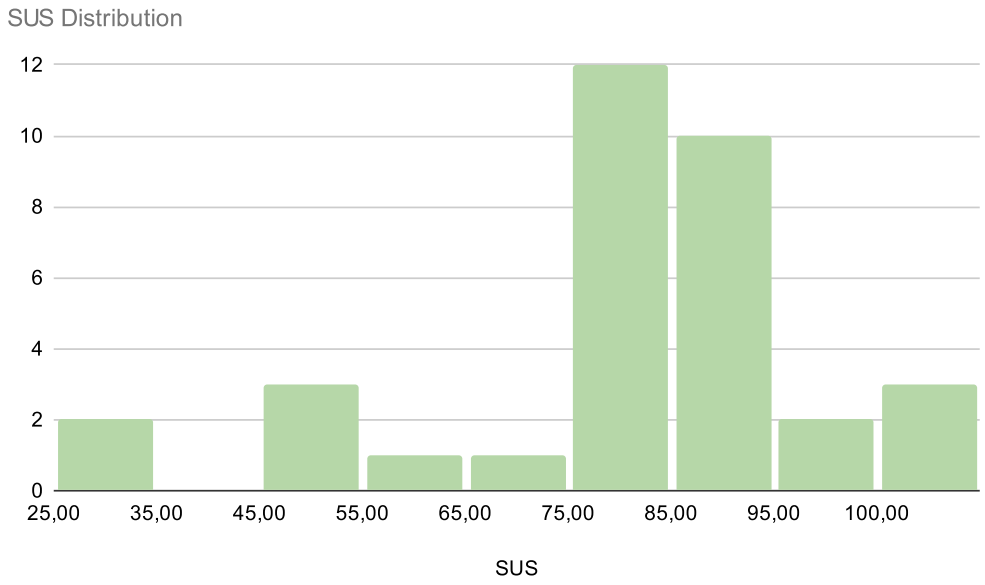

Fig. 11. SUS — Distribution per user.

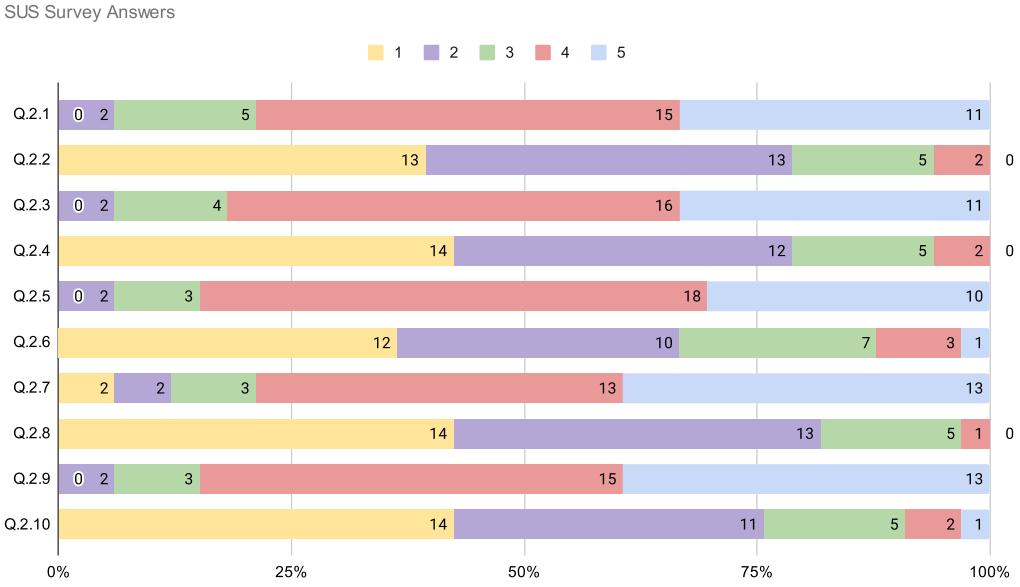

Fig. 12. SUS — Responses per questions.

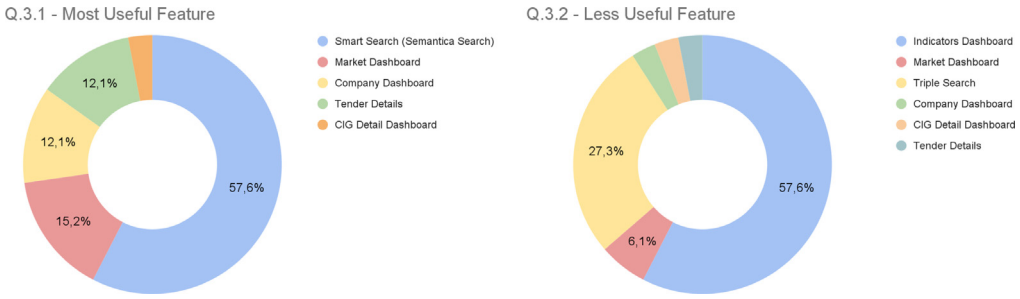

Fig. 13. Users Opinion for the most or least useful system feature.

Fig. 14 depicts the distribution of the ratings supplied by each user to the Q.5.1 question required to compute *Net Promoter Score*; on a scale from 0 to 10, the 6.1% of the participants answered 7, 33.3% 8, 30.3% 9 and 15.2% 10. The idea behind the NPS is to divide the users into *promoters*, *passives* and *detractors* of the item, based on their answer: users providing ratings between 10 and 9 are considered to be promoters, between 8 and 7 are passives and finally, from 6 to 0 are detractors. The NPS is computed as follows:

$$NPS = \%Promoters - \%Detractors$$

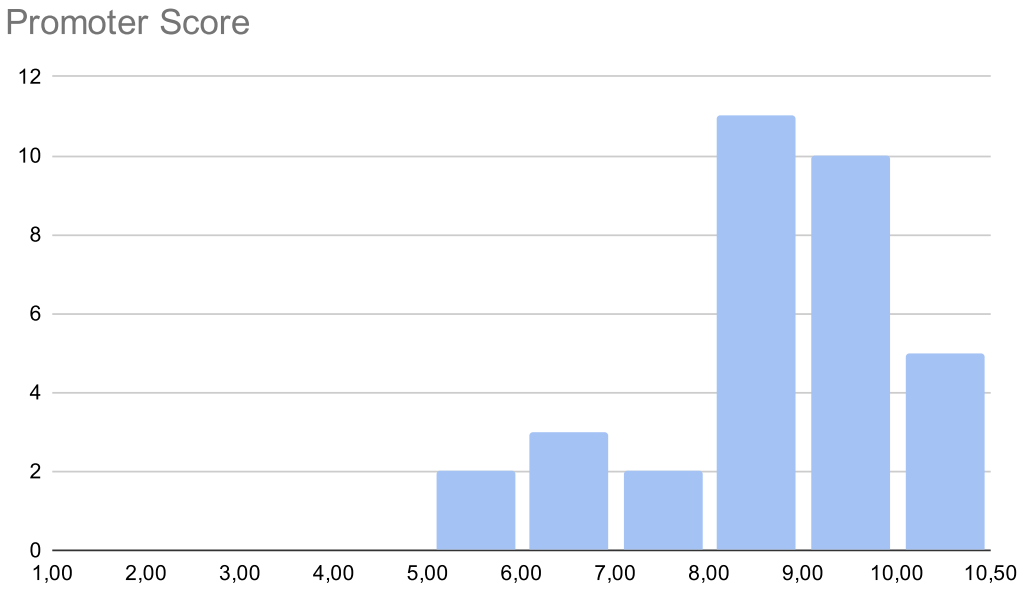

Fig. 14. Promoter score distribution.

and can assume values included in the interval $[-100, +100]$; despite the fact that the computation occurs between percentages, the NPS is actually expressed as a decimal value. General guidelines established by Bain&Co.,[16] inventors of the NPS state that any positive, non-zero score of the NPS is considered "good" since it means that there are more promoters than detractors; however, any score above 20 is considered encouraging, whereas 50 is excellent and above 80 first-rate. Our system scores 30.3, having 45.5% of promoters and 15.2% of detractors.

5.1. Evaluation of collusion risk indicators

We conducted a separate study to evaluate how much the collusion risk indicators can help RUPs support their activities and, more specifically, detect potential anomalies and deceitful practices. For this study, we involved 11 RUPs which are in charge of performing anti-collusion assessments over tenders. First of all, we gathered all the RUPs for a meeting where we explained in detail the tool and how each collusion risk indicator was computed. Next, we asked the RUPs to use the Company's Dashboard to perform their daily activities. Given the topic sensitivity, we are not able to directly access and share data about collusive companies. After 2 h, we then asked the RUPs to answer a questionnaire. We decided to take into account the following four constructs defined in [50,51]:

- Perceived value: represents an individual's overall evaluation of the costs and benefits of adopting the tool, as determined by their attitude towards the change;
- Switching benefit: represents an individual's perception of the benefit that will derive from the adoption of the new tool;
- Switching cost: represents an individual's perception of the costs and efforts required to switch or integrate the new tool;
- Self-efficacy for change: represents the individual's perception of their ability to easily adapt to the new tool.

Given the questions related to each construct, we adapted them to our scenario. The survey consists of 14 questions with the 7-point Likert Scale (1 — Strongly Disagree, 7 — Strongly Agree). The final list of questions is shown in Tables 6 and 7.

Table 8 shows the main statistics related to the results of the survey in terms of mean standard deviation and variance. The values show how the responses were very favourable, with the three positive constructs (i.e. PVL, SWB, and SFC) with a mean value over 5 and the negative construct (i.e. the SWC) with a mean of 3.3. The variance and the standard deviation highlight that RUPs share their opinion about the collusion risk indicators. In Table 9, we reported the Pearson correlation matrix of the scores obtained by each construct which shows a strong relationship among all the constructs. All the correlations are statistically significant and this proves the effectiveness of our solution since:

- Perceived value has a strong positive correlation with Switching Benefits;
- Self-efficacy for change has a positive correlation on the perceived value;
- Switching Benefits have a positive correlation on Self-efficacy for change;
- Switching Costs have a negative correlation with all the other constructs.

6. Conclusions and future works

This research presented a Decision Support System capable of assisting and facilitating the work of personnel participating in public procurement procedures. Our proposal concept is to supply users with a variety of instruments that can ease access to various types of information beneficial to their job and facilitate procedures at various stages of the procurement pipeline: call for tender definition, bidder selection, contract award as well as for monitoring the whole operations in the specific area of interest.

To serve this purpose, we have implemented a system that can process data of different types, both structured and unstructured, deriving from the vast amount of resources available within public agencies, e.g. documentation or tabular data from past tenders. On top of them, we built the system illustrated above, which in our opinion, is an effective tool able to turn the wealth of data, otherwise unused, into useful, accessible and targeted information. Actually, the main objective was to provide a simple and accessible tool that can empower non-technical

16 https://www.bain.com/insights/introducing-the-net-promoter-system-loyalty-insights/

Table 6
Questions for the Perceived Value (PVL) and Switching benefit (SWB) constructs.

ID	Question	Construct
PVL 1	Considering the time and effort I have to spend, integrating the tool into my way of working is worthwhile.	Perceived value
PVL 2	Considering the loss that I incur, integrating the tool into my way of working is of good value.	
PVL 3	Considering the hassle that I have to experience, integrating the tool into my way of working is beneficial to me.	
SWB 1	Integrating the tool into my current way of working would enhance my effectiveness on the job.	Switching benefit
SWB 2	Integrating the tool into my current way of working would enable me to accomplish relevant tasks more quickly.	
SWB 3	Integrating the tool into my current way of working would increase my productivity.	
SWB 4	Integrating the tool into my current way of working would improve the quality of the work I do.	

Table 7
Questions for the Switching cost (SWC) and Self-efficacy for change (SFC) constructs.

ID	Question	Construct
SWC 1	I have already put a lot of time and effort into mastering the current way of working.	Switchin cost
SWC 2	It would take a lot of time and effort to integrate the tool into my current way of working.	
SWC 3	Integrating the tool into my current way of working could result in unexpected hassles.	
SWC 4	I would lose a lot in my work if I were to integrate the tool into my current way of working would.	
SFC 1	Based on my own knowledge, skills and abilities, using the tool in my everyday work activities would be easy for me.	Self-efficacy for change
SFC 2	I am able to integrate the tool in my current way of working without the help of others.	
SFC 3	I am able to integrate the tool into my current way of working reasonably well on my own.	

Table 8
Statistics related to the different constructs.

	Mean	Std Dev	Variance
PVL	5.52	1.17	1.36
SWB	5.5	1.15	1.31
SWC	3.3	.91	.82
SFC	5.39	1.11	1.24

Table 9
Correlation matrix among the four constructs.

	PVL	SWB	SWC	SFC
PVL	1	.984*	−.961*	.913*
SWB		1	−.938*	.967*
SWC			1	−.844*
SFC				1

* Significant at .05 level

users to access information from data without all the technical struggles involved. Moreover, to the best of our knowledge, our decision system is the first example of its kind.

To validate our concept, we distributed an anonymous survey to numerous professional figures involved in procurement, asking them to evaluate it. The findings back with our assumption, indicating overall satisfaction across all 33 participants.

The framework that we implemented is a proof-of-concept built on the requirement to first validate the design architecture and functionalities constrained by a limited amount of hardware resources. Therefore, some of our technical decisions are focused on these boundaries.

In future work, indeed, we consider the possibility of accessing even more extended data assets, mainly while the digitisation process of public administration advances and thereby, a larger number of APIs will be available. To improve our solution, we also plan to employ the latest results in the field of AI (Artificial Intelligence), e.g. *Large Language Models* [52] (LLMs). Such models can be used, e.g. in the form of chat for question answering over documents and data, by giving them access to such resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by *InnovaPuglia S.p.A.* and *Regione Puglia*, with the projects:

- "PAI - Sistemi prototipali per le applicazioni di e-procurement a supporto del RUP basati su tecnologie di Natural Language Processing, Machine Learning e Intelligenza Artificiale"
- "SIAP – Sistemi di supporto alle decisioni basati su tecnologie di intelligenza artificiale per il governo dei cicli di investimenti e appalti"

We acknowledge the support of the PNRR project FAIR - Future AI Research (PE00000013), Spoke 6 - Symbiotic AI (CUP H97G22000210007) under the NRRP MUR program funded by the NextGenerationEU.

References

[1] A. Davila, M. Gupta, R. Palmer, Moving procurement systems to the internet:: The adoption and use of e-procurement technology models, Eur. Manag. J. 21 (1) (2003) 11–23.
[2] P. Phillips, W. Piotrowicz, E-procurement: How does it enhance strategic performance? 2006, Retrieved May 18 (2006).
[3] H.Z. Henriksen, V. Mahnke, J.M. Hansen, Public e-procurement adoption: economic and political rationality, in: 37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the, IEEE, 2004, pp. 9–pp.
[4] S. Kajewski, A. Weippert, E-tendering: Benefits, challenges and recommendations for practice, in: Clients Driving Innovation CRC Construction Innovation International Conference Proceedings, Australian Cooperative Research Centre for Construction Innovation, 2004, pp. 1–11.
[5] E. Dawson, S. Christensen, W. Duncan, P.J. Black, E. Ernest Foo, R. Du, Security and Legal Issues in E-Tendering, CRC for Construction Innovation, 2005.
[6] E. Seah, S. Profile, Dos and don'ts for e-tendering–a quantity surveying perspective, Project Control Professional 46 (5) (2008) 18.
[7] S. Liu, A.H. Duffy, R.I. Whitfield, I.M. Boyle, Integration of decision support systems to improve decision support performance, Knowl. Inf. Syst. 22 (2010) 261–286.
[8] D.J. Power, Decision Support Systems: Concepts and Resources for Managers, Greenwood Publishing Group, 2002.
[9] W.H. Inmon, Building the Data Warehouse, John wiley & sons, 2005.
[10] E.F. Codd, Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate, 1993, http://www.arborsoft.com/papers/coddTOC.html.
[11] H.J. Watson, Tutorial: business intelligence–past, present, and future, Commun. Assoc. Inf. Syst. 25 (1) (2009) 39.
[12] R. Mohemad, A.R. Hamdan, Z.A. Othman, N.M.M. Noor, Decision support systems (DSS) in construction tendering processes, 2010, arXiv preprint arXiv:1004.3260.
[13] L. Doulos, I. Sioutis, P. Kontaxis, G. Zissis, K. Faidas, A decision support system for assessment of street lighting tenders based on energy performance indicators and environmental criteria: Overview, methodology and case study, Sustainable Cities Soc. 51 (2019) 101759.
[14] M.M. Kumaraswamy, S.M. Dissanayaka, Developing a decision support system for building project procurement, Build. Environ. 36 (3) (2001) 337–349.
[15] V. Bobar, K. Mandic, M. Suknovic, Bidder selection in public procurement using a fuzzy decision support system, Int. J. Decis. Support Syst. Technol. (IJDSST) 7 (1) (2015) 31–49.
[16] K. Dobi, J. Gugić, D. Kancijan, AHP as a decision support tool in the multicriteria evaluation of bids in public procurement, in: Proceedings of the ITI 2010, 32nd International Conference on Information Technology Interfaces, IEEE, 2010, pp. 447–452.
[17] M. Hasnain, M.J. Thaheem, F. Ullah, Best value contractor selection in road construction projects: ANP-based decision support system, Int. J. Civ. Eng. 16 (2018) 695–714.
[18] K. Lam, S.T. Ng, H. Tiesong, M. Skitmore, S. Cheung, Decision support system for contractor pre-qualification—artificial neural network model, Eng. Constr. Archit. Manag. 7 (3) (2000) 251–266.
[19] S.T. Ng, R.M. Skitmore, CP-DSS: decision support system for contractor prequalification, Civ. Eng. Syst. 12 (2) (1995) 133–159.
[20] R.B. Velasco, I. Carpanese, R. Interian, O.C.G. Paulo Neto, C.C. Ribeiro, A decision support system for fraud detection in public procurement, Int. Trans. Oper. Res. 28 (1) (2021) 27–47, http://dx.doi.org/10.1111/itor.12811, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/itor.12811, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/itor.12811.
[21] M. Fazekas, I.J. Tóth, L.P. King, An objective corruption risk index using public procurement data, Eur. J. Crim. Policy Res. 22 (2016) 369–397.
[22] African Union, Corruption perceptions index 2018, 2019.
[23] D. Kaufmann, A. Kraay, M. Mastruzzi, The worldwide governance indicators: Methodology and analytical issues1, Hague J. Rule Law 3 (2) (2011) 220–246.
[24] M.A. Golden, L. Picci, Proposal for a new measure of corruption, illustrated with Italian data, Econ. Polit. 17 (1) (2005) 37–75.
[25] B. Tóth, M. Fazekas, Á. Czibik, I.J. Tóth, Toolkit for detecting collusive bidding in public procurement, 2014, With examples from Hungary.
[26] R. Nai, E. Sulis, R. Meo, Public procurement fraud detection and artificial intelligence techniques: a literature review, 2022.
[27] J. Cutroni, Google Analytics, Publisher O'Reilly Media, 2007.
[28] M. Card, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann, 1999.
[29] S. Few, Show me the numbers, Analytics Pres (2004).
[30] A. Eberhardt, M.S. Silveira, Show me the data! a systematic mapping on open government data visualization, in: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, in: dg.o '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 1–10, http://dx.doi.org/10.1145/3209281.3209337.
[31] B. Ansari, M. Barati, E.G. Martin, Enhancing the usability and usefulness of open government data: A comprehensive review of the state of open government data visualization research, Gov. Inf. Q. 39 (1) (2022) 101657, http://dx.doi.org/10.1016/j.giq.2021.101657, URL https://www.sciencedirect.com/science/article/pii/S0740624X21000939.
[32] A. Graves, J. Hendler, Visualization tools for open government data, in: Proceedings of the 14th Annual International Conference on Digital Government Research, in: dg.o '13, Association for Computing Machinery, New York, NY, USA, 2013, pp. 136–145, http://dx.doi.org/10.1145/2479724.2479746.
[33] A. Guarino, N. Lettieri, D. Malandrino, P. Russo, R. Zaccagnino, Visual analytics to make sense of large-scale administrative and normative data, in: 2019 23rd International Conference Information Visualisation (IV), IEEE, 2019, pp. 133–138.
[34] R. Madan, M. Ashok, AI adoption and diffusion in public administration: A systematic literature review and future research agenda, Gov. Inf. Q. (2022) 101774.
[35] A. Mavidis, D. Folinas, From public E-procurement 3.0 to E-procurement 4.0; a critical literature review, Sustainability 14 (18) (2022) 11252.
[36] P. Basile, A. Caputo, M.D. Ciano, G. Grasso, G. Rossiello, G. Semeraro, SEPIR: a semantic and personalised information retrieval tool for the public administration based on distributional semantics, Int. J. Electron. Gov. 9 (1–2) (2017) 132–155.
[37] P. Kanerva, J. Kristoferson, A. Holst, Random indexing of text samples for latent semantic analysis, in: Proceedings of the Annual Meeting of the Cognitive Science Society, Vol. 22, (22) 2000.
[38] 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, in: CEUR Workshop Proceedings, 2947, CEUR-WS.org, 2021, URL https://ceur-ws.org/Vol-2947/paper2.pdf.
[39] 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 Linguistics, CLiC-It 2021, Milan, Italy, January 26-28, 2022, in: CEUR Workshop Proceedings, 3033, CEUR-WS.org, 2021, URL https://ceur-ws.org/Vol-3033/paper28.pdf.
[40] G. Salton, A. Wong, C.-S. Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1975) 613–620.
[41] S. Robertson, H. Zaragoza, et al., The probabilistic relevance framework: BM25 and beyond, Found. Trends Inf. Retr. 3 (4) (2009) 333–389.
[42] P. Basile, A. Caputo, G. Semeraro, Integrating sense discrimination in a semantic information retrieval system, in: Information Retrieval and Mining in Distributed Environments, Springer, 2011, pp. 249–265.
[43] P. Basile, A. Caputo, A.L. Gentile, M. Degemmis, P. Lops, G. Semeraro, et al., Enhancing semantic search using N-levels document representation, SemSearch 334 (2008) 29–43.

[44] 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, http://dx.doi.org/10.18653/v1/w17-1003.
[45] P.C. Mandal, Net promoter score: a conceptual analysis, Int. J. Manag. Concepts Philos. 8 (4) (2014) 209–219.
[46] J. Brooke, et al., SUS-a quick and dirty usability scale, Usability Eval. Ind. 189 (194) (1996) 4–7.
[47] J.R. Lewis, The system usability scale: past, present, and future, Int. J. Hum.-Comput. Interact. 34 (7) (2018) 577–590.
[48] R. Likert, A technique for the measurement of attitudes, Arch. Psychol. (1932).
[49] J. Sauro, J.R. Lewis, Quantifying the user experience: practical statistics for user research, Morgan Kaufmann, 2016, http://dx.doi.org/10.1016/C2010-0-65192-3.
[50] I. Mahmud, T. Ramayah, S. Kurnia, To use or not to use: Modelling end user grumbling as user resistance in pre-implementation stage of enterprise resource planning system, Inf. Syst. 69 (2017) 164–179, http://dx.doi.org/10.1016/j.is.2017.05.005.
[51] H.-W. Kim, A. Kankanhalli, Investigating user resistance to information systems implementation: A status quo bias perspective, MIS Q. (2009) 567–582.
[52] W.X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al., A survey of large language models, 2023, arXiv preprint arXiv:2303.18223.