



**HAL**  
open science

## A Quality Assessment Framework for Information Extraction in Job Advertisements

Albeiro Espinal, Yannis Haralambous, Dominique Bedart, John Puentes

► **To cite this version:**

Albeiro Espinal, Yannis Haralambous, Dominique Bedart, John Puentes. A Quality Assessment Framework for Information Extraction in Job Advertisements. SN Computer Science, 2023, 4 (6), pp.787. 10.1007/s42979-023-02247-5 . hal-04599951

**HAL Id: hal-04599951**

**<https://hal.science/hal-04599951v1>**

Submitted on 4 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# A Quality Assessment Framework for Information Extraction in Job Advertisements

Albeiro Espinal<sup>1,2</sup> · Yannis Haralambous<sup>1</sup> · Dominique Bedart<sup>2</sup> · John Puentes<sup>1</sup>

Received: 3 April 2023 / Accepted: 12 August 2023  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2023

## Abstract

Efficient extraction of pertinent terms from job advertisements (JAs) relies on feature engineering, specifically the selection of textual markers. Yet, current methods lack comprehensive assessment of markers' quality in terms of their informative value and context-specific relevance. To address this issue, we introduce a fuzzy-inference-based framework, aimed at optimizing textual markers within JAs documents. This framework operates on three principles: ambiguity estimation of markers, assessment of the information quantity they convey, and their ongoing relevance evaluation within an organizational context. We demonstrated its effectiveness by applying it to 30 markers across 73 recruitment processes. The assessment resulted in the selection of five markers, improving the *F1*-score by 5% for relevant term extraction from JAs, via a possibilistic belief–desire–intention architecture. Six additional markers yielded a minor enhancement of results but significantly boosted model explainability. Thus, our framework significantly bolsters the extraction accuracy and clarity of the underlying model for JAs.

**Keywords** Job advertisement · Relevant terms extraction optimization · Mamdani inference · Fuzzy-inference-based textual markers assessment

## Introduction

Recruiters face a considerable challenge in manually evaluating a high volume of resumes and determining their relevance to job requirements during the recruitment process. In response, various automated methods have emerged to match job advertisements (JA) with resumes (CV) [1]. These

systems enable efficient candidate ranking and pre-selection by analyzing suitability to job specifics.

Despite the progress in CV ranking systems, aligning CVs with JAs' requirements automatically remains complex [2]. JAs' brevity, meant to concisely detail job specifics and the ideal candidate, presents two concerns. First, the size discrepancy between the JA and CVs can misdirect the system into assigning high relevance to candidates based on non-critical JA information. Second, since JAs often depict an ideal candidate, a complete match may be improbable, making it essential to accurately identify the most essential requirements [3].

To address these issues, improved information extraction from JAs is necessary, calling for the refinement of feature engineering processes for identifying pertinent textual markers [4]. This process should account for various information quality criteria such as relevance, ambiguity [5], the amount of extractable information, and alignment with recruiter expertise and organizational context [2].

Our study addresses a key research question: how can a quality assessment framework optimize textual markers to accurately extract the most relevant terms from JAs? Furthermore, how can this process effectively manage ambiguity, ensure alignment with recruiter expertise,

---

This article is part of the topical collection “Advances on Computational Intelligence 2022” guest edited by Joaquim Filipe, Kevin Warwick, Janusz Kacprzyk, Thomas Bäck, Bas van Stein, Christian Wagner, Jonathan Garibaldi, H. K. Lam, Marie Cottrell and Faiyaz Doctor.

---

✉ Albeiro Espinal  
albeiro.espinal@imt-atlantique.fr

Yannis Haralambous  
yannis.haralambous@imt-atlantique.fr

Dominique Bedart  
dominique.bedart@dsi-globalservices.fr

John Puentes  
john.puentes@imt-atlantique.fr

<sup>1</sup> IMT Atlantique, Lab-STICC, CNRS UMR 6285, Brest, France

<sup>2</sup> DSI Global Services, Le Plessis Robinson, France

maintain information pertinence, and uphold a context-driven, explainable approach?

Building upon prior work [6], we introduce a comprehensive, flexible, and context-sensitive methodology for selecting textual markers in JAs. Our methodology features four distinctive innovations aimed at improving automated CV ranking for job positions:

- *Comprehensive evaluation*: our approach explores relationships between textual markers and uses metrics such as Pearson correlation, precision, mutual information, ambiguity, and statistical significance for their holistic assessment.
- *Fuzzy-inference-based assessment*: We use a Mamdani-type inference engine to manage ambiguity and variability in JAs.
- *Holistic view of feature quality*: Our framework builds on three pillars: ambiguity estimation, information quantity assessment, and continuous evaluation of marker pertinence in the organizational context.
- *Adaptability to organizational context*: Our methodology ensures that the chosen markers stay relevant as organizations evolve by continually evaluating their pertinence.

This article is structured as follows. Section “[State of the Art](#)” describes the state of the art, and Sect. [JA Representation](#) outlines key principles for JA representation. We introduce a collection of textual relevance markers in Sect. [Textual Markers](#) and detail the proposed evaluation approach in Sect. [Description of the Framework for the Evaluation of Textual Markers](#). We then present our experimental findings and associated discussion in Sects. [Experimentation](#) and [Discussion](#), respectively. Finally, conclusions and perspectives are provided in Sect. [Conclusion](#).

## State of the Art

A “textual marker” identifies and represents a pertinent entity within a document [7], while also underscoring its significance within a social or organizational context [8]. Particularly, in the field of information extraction from JAs, the textual marker is vital to identify the most critical terms that ensure the quality of automated candidate CV ranking during recruitment [9]. Yet, some existing approaches bypass JA processing, heavily relying on direct feedback from recruiters [2].

Our literature review identified a gap: no methodologies have been explicitly designed for selecting textual markers based on their quality within JAs. However, information extraction methodologies have proposed various approaches to optimize identifying significant document features. These methodologies can be classified into four types: wrapper,

filter, embedded methods [10], and, spanning across these categories, fuzzy methods. Each of these methods will be discussed in the ensuing subsections.

## Wrapper Methods

Wrapper methods use a machine learning algorithm to evaluate the quality of feature subsets [10]. For instance, an improved particle swarm optimization method for selecting textual features was proposed by [11]. In addition, a combined Term Frequency-Inverse Document Frequency and Support Vector Machines Recursive Feature Elimination approach enhances feature selection for sentiment classification tasks [12]. Other alternatives include search optimization algorithms leveraging the Iterated Greedy metaheuristic [13] and genetic, sequential, and recursive elimination-based algorithms [14].

## Filter Methods

Filter methods evaluate feature quality based on their intrinsic characteristics, independent of machine learning algorithms. This generally results in less computational expense. Features are typically ranked based on statistical metrics, for example, Chi-squared, Gini-Index, Pearson correlation coefficient, precision, information gain, or mutual information entropy [15–21].

Novel filter-type approaches include the Maximal Information Coefficient [22], which detects linear and non-linear feature-target relationships. Other methods combine multiple metrics, for instance, the Fast Correlation Based Filter [23]. This technique estimates correlations among features by integrating various metrics and a backward selection procedure, paired with a sequential search strategy.

## Embedded Methods

Embedded methods integrate the feature-selection process within the learning algorithm, optimizing feature selection and model training simultaneously. Traditional methods of this type include the Least Absolute Shrinkage and Selection Operator which uses L1 regularization [24]. Decision trees are also part of this category, performing inherent feature selection [21]. Other methods combine principal component analysis (PCA), latent semantic analysis, and random projection [25].

## Fuzzy Methods

Fuzzy feature-selection techniques, adaptable and flexible alternatives to conventional selection methods, have emerged in intricate contexts characterized by uncertainty. For example, a fuzzy feature correlation approach

incorporates information gain to enhance feature identification and selection [26]. A fuzzy-rough feature-selection technique was used to reduce word dimensionality in texts, facilitating document processing [27]. In addition, a hybrid fuzzy marker selection method based on Mamdani-type fuzzy rule-based systems has demonstrated enhanced feature-selection performance compared to traditional filter approaches [28].

In summary, despite existing methodologies for identifying key document features, none specifically optimizes the selection of high-quality textual markers within JAs. Given their vital role in automating candidate CV ranking, our study seeks to address this gap by introducing a comprehensive methodology for marker selection in JAs. Our approach emphasizes quality assessment, the integration of recruiters' perspectives, explainability, and adaptability to organizational changes. More details follow in the subsequent sections.

## JA Representation

Automatic extraction of relevant terms from JAs is contingent upon creating an appropriate document representation [15]. This study aims to create such a representation by analyzing and modeling the organizational context where JAs are processed. This representation is likely to enhance the accuracy of artificial intelligence systems developed for JA processing [29].

### The Screening Process: Analysis of JAs and CVs

JAs and CVs serve as critical communication tools for recruiters and job applicants, conveying required and possessed professional skills, respectively [9]. Professional skills include knowledge, practical abilities, and interpersonal skills applicable to professional activities and tasks [30].

An important aspect to consider in this communication process is the economic activity sector in which the JA's job role is situated. Economic activity sectors, such as health-care, information and communications technology, manufacturing, management and so on, influence the specific requirements, skills, and terminologies used in JAs [30].

By integrating the understanding of professional skills and the corresponding economic activity sector, the screening phase of the recruitment process focuses on identifying suitable candidates by aligning applicant CVs with JA requirements [2]. This procedure heavily relies on recruiters' understanding of the organizational context [3, 31]. Post-screening, recruiters conduct further examinations and in-person interviews, if necessary.

The following section delves into the protocol for deriving textual markers in JAs, aligning with recruiters' strategies for effective information extraction.

## Organizational Context and Recruiters' Strategies

Building on our previous work [6], we utilized the UNC method [32] to capture key elements of the organizational context-including actors, artifacts, actions, and processes-relevant to JAs in hiring procedures [33]. Initial interviews with recruiters facilitated the creation of a mother-ontology, encapsulating JA's core concepts, and modeling of recruiter goals and enterprise processes. From studying recruiters' annotation strategies, we generated semantic rules, which served as textual markers of relevant information. Using a fuzzy decision tree, we identified the least ambiguous markers, integrated them into an ontology-based possibilistic agent architecture to simulate recruiters' behaviors, and provided insights into the context-sensitive relevance of information in JAs [29].

## Ontological Resources

Our approach relies on a mother-ontology that incorporates JA-related concepts and relationships. Specifically, we define a mother-ontology as an extensive ontology of module specifications. We enriched this ontology with organization-specific resources, such as the DSI Group's internal professional skills ontology and international ontologies like ESCO [30]. It also includes ontologies of professional skills frameworks, including O\*NET,<sup>1</sup> CIGREF,<sup>2</sup> and ROME,<sup>3</sup> reconstructed via transformations from text into Resource Description Framework triples. Furthermore, we incorporated an ontology derived from CVs and JAs, processed using ontology learning techniques to align professional skills with market knowledge. To integrate all ontologies into the mother-ontology, we implemented a hybrid approach which combines semantic similarity estimation using Bidirectional Encoder Representations from Transformers (BERT), terminological variation analysis, and ontology quality metrics [6]. In Fig. 1, we present an upper view of the resulting ontology.

## Beliefs, Desires and Intentions Architecture of the Agent

Considering the context-sensitive and uncertain nature of human semantic inferences [5], we incorporated this

<sup>1</sup> <https://www.onetonline.org/>.

<sup>2</sup> <https://www.cigref.fr/>.

<sup>3</sup> <https://www.pole-emploi.fr/employeur/vos-recrutements/le-rome-et-les-fiches-metiers.html>.

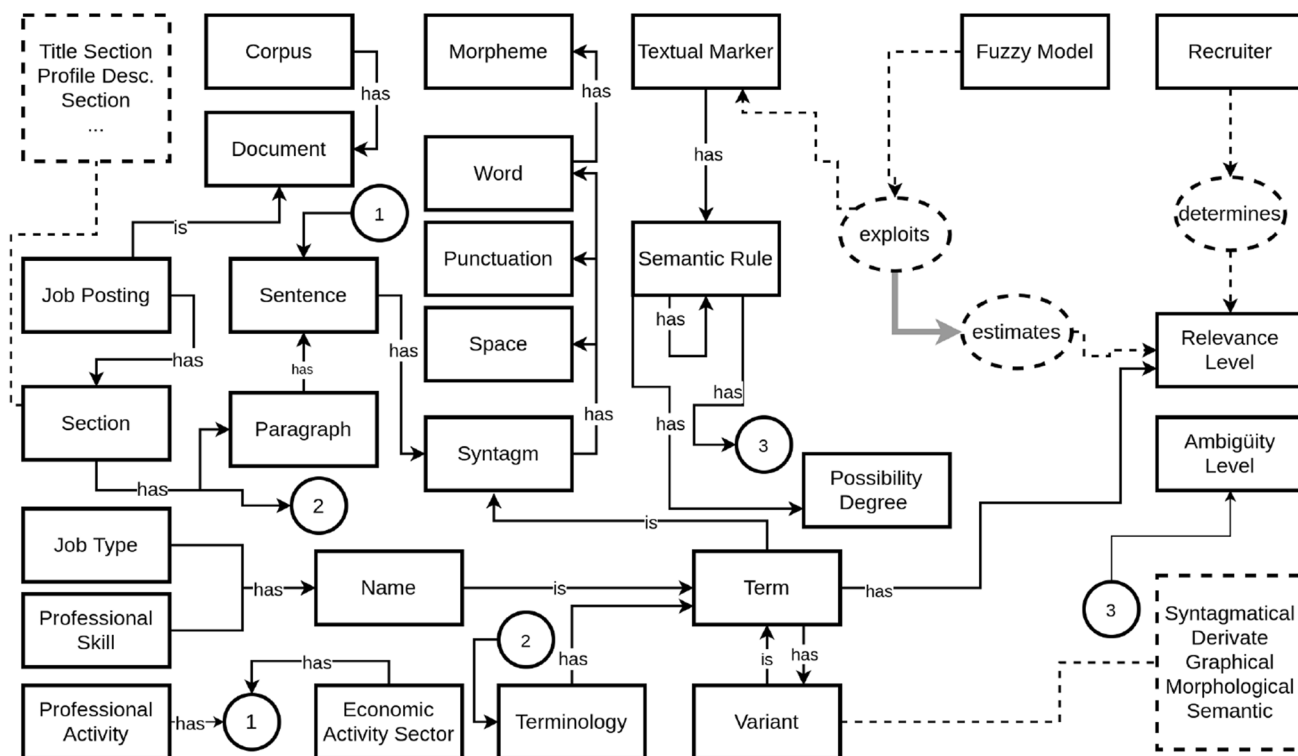


Fig. 1 Upper view of the mother-ontology, derived from the context representation of the previous research [6]

understanding into the extraction of pertinent terms from JAs. A possibilistic architecture was proposed [6], embodying human semantic reasoning and the associated uncertainty, thus enhancing the automatic extraction of relevant terms. This architecture is rooted in the agent’s beliefs, desires, and intentions (BDI) paradigm, which is connected to the modeling of variables using possibility theory [34] (Fig. 2). It consists of three primary modules.

The Beliefs Module is composed of two sub-parts, involving trust allocation ( $\tau$ ) to information sources like JAs, ontological resources, and textual markers, and a possibility distribution ( $\pi$ ) defining the agent’s beliefs ( $\mathcal{B}$ ). The agent can adjust beliefs to mitigate conflicts between contradictory or related information sources.

The Desires Module is represented by an utility distribution (formally a possibility distribution). It indicates whether or not the use of a textual relevance marker is convenient according to the current agent’s beliefs, desires, and desire-generation rules. Desire-generation rules, denoted as  $R_j$ , are logical conditions expressing dependence relationships between beliefs and desires [34]. The evaluation of desire-generation rules enables the agent to estimate to what extent, denoted as  $\Delta$ , it is justified to use each marker. Textual markers associated with positive degrees become justified desires, denoted as  $\Upsilon$ .

The Intentions Module outlines the actions that the agent decides to execute based on its beliefs, justified desires, and goals. These actions pertain to natural language processing methods essential for extracting relevant terms from JAs.

For the agent to make pertinent and explainable actions guided by its beliefs, desires, and intentions, a method to select the most relevant textual markers is crucial. These markers steer the agent in specific scenarios. An effective method for textual marker selection assists the agent’s decision-making regarding term relevance. In the following

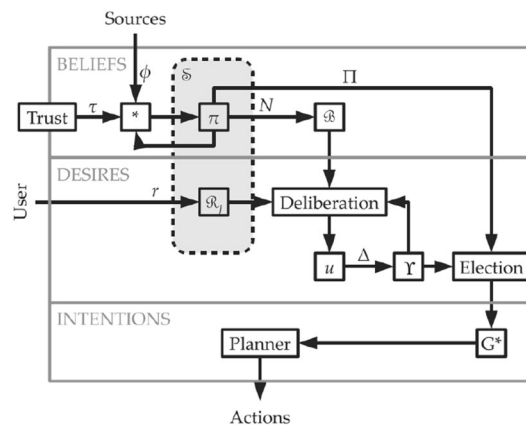


Fig. 2 Fundamental structure of the dynamic BDI agent [6] [34]

section, we introduce a set of textual markers to illustrate the proposed approach’s capabilities.

### Textual Markers

In this section, we start by referring back to a comprehensive formalization [6] for the concept of textual marker, ensuring compatibility with both the markers utilized in the YAKE! algorithm and those employed by recruiters in practice.

### Preliminary Definitions

A term can be defined as “a functional class of lexical units utilized in discourse” [35]. In our specific context, terms are identified based on their *termhood*, which is quantified using the *weirdness ratio*.

Term extraction is performed by applying specific morpho-syntactic patterns to multiple specialized corpora [35]. The majority of these extracted terms are noun phrases.

Let  $d_i$  be a JA belonging to a corpus  $C$  and  $T_{d_i} = \{t_1, t_2, \dots, t_n\}$  the set of terms of  $d_i$ .

Let  $R_{d_i} \subseteq T_{d_i}$  be the *set of most relevant terms* of  $d_i$ . Each term  $t_i \in R_{d_i}$  is considered as relevant under a possibility degree  $\alpha_{t_i}$ .

Let  $A_{d_i} = \{a_1, a_2, \dots, a_m\}$  be the set of sections of  $d_i$  (job description, profile details, etc). Each section  $a_i$  can be represented by a subset of terms from  $T_{d_i}$ . A term can belong to multiple sections.

Let  $E_{d_i} = \{e_1, e_2, \dots, e_p\}$  be a set of qualifying adjectives and nouns that are linked to a subset of terms in  $T_{d_i}$  by syntax dependencies.

Let  $O = \{o_1, o_2, \dots, o_s\}$  be a set of ontologies (as the one presented in Sect. **JA Representation**). Let  $c_{o_s} = \{c_{s,1}, c_{s,2}, \dots, c_{s,k}\}$  be the set of concepts of ontology  $o_s$  and  $T_{c_j} = \{t_{j,1}, t_{j,2}, \dots, t_{j,l}\}$  the set of terms lexically representing concept  $c_j$  in a given language.

The mother-ontology, denoted as  $O$ , encompasses a variety of concepts, including the hierarchical structure of textual documents, which comprises sections, paragraphs, sentences, syntagms, terms, words, morphemes, and so forth.

### Marker Definitions

Our proposal is an enhancement of the 16 textual markers explored in a previous study [6], to design the quality assessment framework. We will revisit the formalization of the first marker (Sect. **Textual Marker M1: Presence of Professional Skills or Job Types in Titles**) to illustrate the approach and, subsequently, we will provide a synopsis of the following

**Table 1** Brief description of textual markers proposed in a previous research [6], derived from recruiters’ behavior (R), and from a domain-independent approach like YAKE! (E)

Textual Marker #	Description	Type
$M_2$	Terms representing professional skills in a job description section or profile description section	R
$M_3$	Relevance of job posting sections	R
$M_4$	Terms dependent on pertinence expressions	R
$M_5$	Terms used in traces of professional activities descriptions	R
$M_6$	Terms representing high risk professional skills/ activities	R
$M_7$	Actions expressed in management JOs	R
$M_8$	BERT semantic similarity between professional skills	R
$M_9$	Relevance of the economic activity sector	R
$M_{10}$	Professional skill prerequisites	R
$M_{11}$	YAKE! Casing [36]	E
$M_{12}$	YAKE! Term position [36]	E
$M_{13}$	YAKE! Term frequency normalization [36]	E
$M_{14}$	YAKE! Term relatedness to context [36]	E
$M_{15}$	YAKE! Different sentences [36]	E
$M_{16}$	YAKE! Overall score [36]	E

15 markers (Sect. **Summary of Remaining Textual Markers from Prior Work [6] and Description of 14 New Textual Markers**). Originating from both expert recruiters’ practices (#1–#10) and the YAKE! term extraction technique (#11–#16), these markers are utilized in the evaluation experiments of our methodology.

### Textual Marker $M_1$ : Presence of Professional Skills or Job Types in Titles

*If a term within the title corresponds to one of the terms employed to represent professional skills or job types, it may be considered relevant.*

Let  $a_1 \in A_{d_i}$  be the title section of  $d_i$ . Let  $t_{a_1} = \{t_1, t_2, \dots, t_u\}$  be the set of terms contained in  $a_1$ .  $T_{c_j}$  is the set of terms lexically representing a professional skill or job type concept  $c_j$  in the ontology  $o_s$ . We request that:

$$\forall t_k \exists c_j [c_j \in o_s \wedge t_k \in T_{c_j} \wedge t_k \in t_{a_1}] \rightarrow t_k \in R_{d_i} \tag{1}$$

with a possibility degree  $\alpha_{t_{k,1}} \in [0, 1]$ .

### Summary of Remaining Textual Markers from Prior Work [6] and Description of 14 New Textual Markers

A concise description of the remaining 15 textual markers defined in our previous study is provided in Table 1.

**Table 2** Description of textual markers #17–#21 derived from contextual markers (#1–#10)

Marker #	Description	Formula	Possibility degree
$M_{17}$	Professional skills in the job description or profile description section of the JA, which are linked to pertinence expressions	This marker is defined as the conjunction between markers $M_2$ and $M_4$ : $\forall t_k (F_{M_2}(t_k) \wedge F_{M_4}(t_k)) \rightarrow t_k \in R_{d_i}$ where $F_{M_i}$ represents the set of logical conditions associated to markers $M_2$ and $M_4$	$\alpha_{t_{k,17}}(t_k) = \min(\alpha_{t_{k,2}}(t_k), \alpha_{t_{k,4}}(t_k))$
$M_{18}$	Upper-cased or capitalized terms in relevant JA sections	$\forall t_k (F_{M_3}(t_k) \wedge F_{M_{11}}(t_k)) \rightarrow t_k \in R_{d_i}$ where $F_{M_i}$ represents the set of logical conditions associated to markers $M_3$ and $M_{11}$	$\alpha_{t_{k,18}}(t_k) = \min(\alpha_{t_{k,3}}(t_k), \alpha_{t_{k,11}}(t_k))$
$M_{19}$	Specific professional skills in capital or initial capital letters	$\forall t_k (F_{M_2}(t_k) \wedge F_{M_8}(t_k) \wedge F_{M_{11}}(t_k)) \rightarrow t_k \in R_{d_i}$ where $F_{M_i}$ represents the set of logical conditions associated to markers $M_2, M_8$ and $M_{11}$	$\alpha_{t_{k,19}}(t_k) = \min(\alpha_{t_{k,2}}(t_k), \alpha_{t_{k,8}}(t_k), \alpha_{t_{k,11}}(t_k))$
$M_{20}$	Professional skills associated with various textual patterns related to recruiters' strategies about information relevance	$\forall t_k (F_{M_1}(t_k) \vee F_{M_2}(t_k) \vee F_{M_4}(t_k) \vee F_{M_9}(t_k)) \rightarrow t_k \in R_{d_i}$ where $F_{M_i}$ represents the set of logical conditions associated to markers $M_1, M_2, M_4$ and $M_9$	$\alpha_{t_{k,20}}(t_k) = \max(\alpha_{t_{k,1}}(t_k), \alpha_{t_{k,2}}(t_k), \alpha_{t_{k,4}}(t_k), \alpha_{t_{k,9}}(t_k))$
$M_{21}$	PCA method on contextual markers	we implemented the principal component analysis (PCA) method on the context-derived markers ( $M_1 - M_{10}$ ) to reduce them to a single dimension. This marker is defined as: $\forall t_k (t_k \in T_{d_i} \wedge is\_pca\_activated(t_k)) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,21}}(t_k) =   \rho_{t_k}  $ , where $\rho_{t_k}$ is the normalized PCA value of the term $t_k$

**Table 3** Description of textual markers #22–#30 corresponding to “fake” or false markers

Marker #	Description	Formula	Possibility degree
$M_{22}$	The relevance level of a term follows a random uniform distribution	This marker is defined as: $\forall t_k (t_k \in T_{d_i} \wedge is\_random\_marker\_activated(t_k)) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,22}}(t_k) = x_{t_k}$ , where $x_{t_k} \sim U(0, 1)$
$M_{23}$	Absolute value of the cosine function evaluated at the last term position in the JA	We execute the following query: $\forall t_k (t_k \in T_{d_i} \wedge is\_cosin\_marker\_activated(t_k)) \rightarrow t_k \in R_{d_i}$	Let $l_{t_k}$ be the last position of term $t_k$ in the JA's text. Then, $\alpha_{t_{k,23}}(t_k) =  cos(l_{t_k}) $
$M_{24}$	Does the term have an even number of nouns?	Let $n_{t_k}$ be a function that returns the number of nouns in a JA's term. This marker is defined as follows: $\forall t_k (t_k \in T_{d_i} \wedge is\_even(n(t_k))) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,24}} \in [0, 1]$
$M_{25}$	At least one of the term's positions in the text corresponds to a prime number	Let $P_{t_k} = p_{1,t_k}, p_{2,t_k}, \dots, p_{i,t_k}$ be the set of positions of term $t_k$ in the JA's text. We execute the following query: $\forall t_k \exists p_{i,t_k} (t_k \in T_{d_i} \wedge p_{i,t_k} \in P_{t_k} \wedge is\_prime(p_{i,t_k})) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,25}} \in [0, 1]$
$M_{26}$	In the terminology of the JA sorted alphabetically, is the term's position a perfect number?	Let $B_{d_i} = (t_{1,*}, t_{2,*}, \dots, t_{k,*})$ be a tuple containing the terms of the JA sorted alphabetically. The tuple $P_{d_i} = (p_{1,*}, p_{2,*}, \dots, p_{k,*})$ represents their respective positions numbered from 1 to $k$ . We query the following: $\forall t_k \exists t_{k,*} \exists p_{t_{k,*}} (t_k \in T_{d_i} \wedge refers\_to\_same\_term(t_{k,*}, t_k) \wedge is\_a\_position\_of(p_{t_{k,*}}, t_{k,*}) \wedge is\_perfect\_number(p_{t_{k,*}})) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,26}} \in [0, 1]$
$M_{27}$	The Term contains the Letter “x”	$\forall t_k (t_k \in T_{d_i} \wedge contains(t_k, \{x\})) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,27}} \in [0, 1]$
$M_{28}$	The number of times the term is used to create new Term Variations	Let $g(t_k)$ be a function that returns the number of term variants [35] associated with the term $t_k$ . This marker is defined as follows: $\forall t_k (t_k \in T_{d_i} \wedge is\_terms\_variants\_marker\_activated(t_k)) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,28}}(t_k) =   g(t_k)  $
$M_{29}$	The term is the Root of its Associated Syntactic Tree	Let $S_{d_i} = \{s_1, s_2, \dots, s_n\}$ be the set of phrases in the JA. For each phrase $s_i$ , at least one syntactic tree $a_{s_i,j}$ is associated. This marker is defined as follows: $\forall t_k \exists s_i \exists a_{s_i,j} (t_k \in T_{d_i} \wedge t_k \in s_i \wedge is\_root(t_k, s_i)) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,29}} \in [0, 1]$
$M_{30}$	The term contains at least one Upper-cased Letter (Anywhere in the Term)	$\forall t_k (t_k \in T_{d_i} \wedge contains\_uppercase\_letter(t_k)) \rightarrow t_k \in R_{d_i}$	$\alpha_{t_{k,30}} \in [0, 1]$

In addition to these markers, we introduce 14 new markers to the current study to further probe the efficiency and discernment of the marker assessment methodology. Specifically, we have derived four new markers (#17–#20) from the training process of a fuzzy decision tree, which are detailed in Table 2. Marker #21 corresponds to a PCA (Principal Component Analysis) marker derived from markers #1–#10. We have also introduced nine false markers (#22–#30). These are designed to test the method's capacity to distinguish irrelevant markers in the context of JAs. The false markers are presented in Table 3.

In the evaluation of these markers, it is essential to differentiate between 'truth value' and 'possibility degree' [37]. Each marker may have certain requirements or logical conditions determining its truth value for a given term, signifying whether a term is relevant based on the marker's conditions. Conversely, the possibility degree, exclusive to the context of application, signifies the level of confidence the agent places on the truth value. While truth value is binary, expressing relevance or non-relevance of a term, possibility degree communicates the agent's certainty level regarding this truth value, dependent on the quality metrics in the agent's trust module.

## Description of the Framework for the Evaluation of Textual Markers

This section details our holistic framework for evaluating textual markers in JAs, using a set of assumed markers and a JA corpus. Our approach involves:

- Estimating marker correlations to improve explainability.
- Analyzing marker activation frequency within the JA corpus to identify overused and underused markers.
- Measuring each marker's precision in a conventional machine learning model, evaluating their classification accuracy.
- Assessing marker ambiguity and mutual information entropy to measure associated uncertainty.
- Conducting a prospective-statistical study using a simple and highly explainable model like logistic regression to verify markers' behavior on a large scale.
- Identify markers with a minimum level of explainability by correlating possibility levels provided by the markers with recruiters' annotation strategies.

The results are fed into a Mamdani-type inference engine that uses logical rules to identify relevant markers for efficient term extraction in JAs. These refined markers are then

integrated into the BDI agent architecture to enhance its performance. Figure 3 illustrates our methodology.

## Correlation Analysis Between Markers

The evaluation begins with a correlation analysis of the textual markers using the Pearson correlation coefficient to explore linear relationships between markers and recruiters' annotations. This provides a preliminary indicator of explainability.

Let  $M_x$  and  $M_y$  represent two textual markers. The Pearson correlation coefficient,  $r_{xy}$ , is computed as

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} \quad (2)$$

where  $n$  is the number of terms in the JA,  $x_i$  and  $y_i$  denote the possibility degree provided by  $M_x$  and  $M_y$  for the term  $t_i$ ;  $\bar{x}$  and  $\bar{y}$  are the corresponding mean of the possibility degrees, and  $s_x$  and  $s_y$  are the respective standard deviations.

This analysis results in marker pairs  $(m_i, m_j)$  with correlations surpassing a set tolerance level  $\omega$ , allowing for a more thorough marker analysis.

## Preliminary Analysis of the Cumulative Possibility Degrees or Frequency of Markers in the JAs corpus

We estimate the total cumulative possibility degrees, or frequency, for each marker to identify those that may be overactivated or underactivated in the JA corpus. Let  $M_x$  be a textual marker providing a possibility degree  $M_x(t_i) = x_i$  for term  $t_i$  in the JA. We define the accumulated possibility degree, or frequency of a marker as follows:

$$l_{M_x} = \log \left( 1 + \sum_{i=1}^n x_i \right) \quad (3)$$

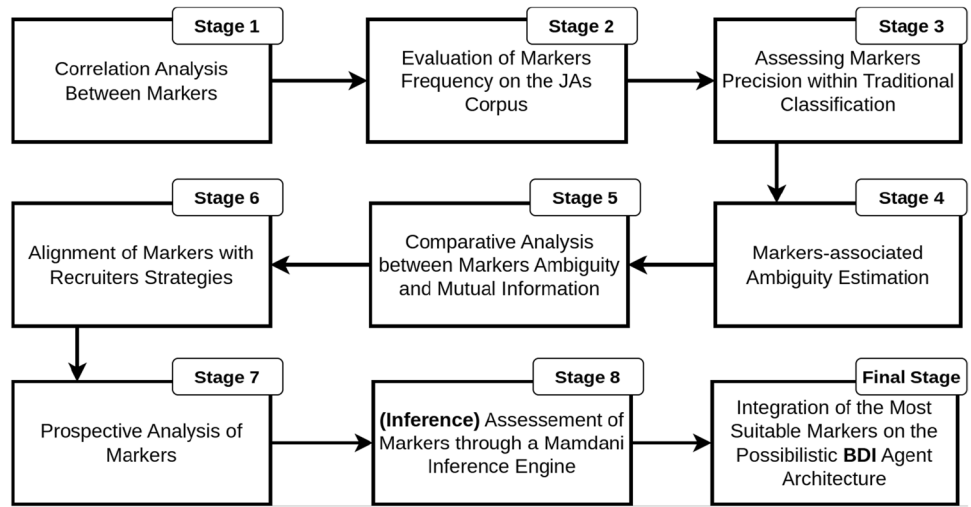
Markers with low activation compared to recruiter annotations may be either highly specific or unsuitable for extracting relevant terms. Conversely, markers with high activation may be able to identify broader sets of relevant terms, but could also be unsuitable due to excessive false positives. Interpretations of these outcomes guide the next evaluation stage.

## Assessing Marker Precision within Traditional Classification

In the third phase, we evaluate each marker's performance as a standalone model within a traditional classification



**Fig. 3** General description of the proposed approach



process. For a process consisting of classes  $C_1, C_2, \dots, C_m$  (in our case, relevant and non-relevant terms), a marker's precision on a specific term  $t_i$  is:

$$p(M_x, t_i) = \begin{cases} 1 & \text{if } M_x(t_i) > 0.5 \\ 0 & \text{if } M_x(t_i) \leq 0.5. \end{cases} \quad (4)$$

The average precision of the marker on the term set, for class  $C_m$ , is:

$$P(C_m | M_x) = \frac{\sum_{i=1}^n p(M_x, t_i)}{n} \quad (5)$$

where  $n$  is the number of JA terms linked to class  $C_m$  in the study corpus. Precision on relevant terms is 'positive precision', and on non-relevant terms, 'negative precision'. Markers fall into four categories based on these precision levels:

- *Type 1—High positive and negative precision:* accurately describe relevant and non-relevant terms.
- *Type 2—High positive, low negative precision:* act as an initial filter for relevant terms, lowering irrelevant ones but retaining most relevant ones.
- *Type 3—Low positive, high negative precision:* identify non-relevant terms (counter-relevance markers) and hint at aspects of relevant ones. High negative precision may be due to class imbalance.
- *Type 4—Low positive and negative precision:* do not contribute to recruiters' annotations and may be irrelevant.

Next, we measure ambiguity to evaluate how well the markers' relevance distribution matches the recruiters' term relevance distribution.

### Marker-Associated Ambiguity Estimation

This stage estimates each marker's analytical ambiguity related to the recruiters' term selection process. A classification task, like the one performed by recruiters, is a rational action with inherent uncertainties [5, 21]. Hence, we introduce a fuzzy-based ambiguity estimation stage, building on a method proposed in a previous study [6].

Defining  $U$  as the set of JA terms and  $C$  as a fuzzy set representing recruiters' perceived term relevance levels, we use the membership function  $\mu_C$  to map recruiters' annotations into  $C$ . The fuzzified recruiters' annotations on the JA are  $Z$ , while  $C_1$  and  $C_2$  denote subsets for relevant and non-relevant term levels, respectively.

Each term  $t$  can be described by a set of relevance markers  $M_k$ , obtained from recruiters' strategies and viewpoints. This is denoted as  $A(t) = M_1, M_2, \dots, M_k$ .

Each relevance marker provides a degree of possibility for a term to be selected as relevant. We fuzzify each marker's possibility degree for term relevance by applying a membership function  $\mu_{M_k}$ , resulting in an evidence  $E_k$ . Although  $\mu_{M_k}$  was built in an equivalent manner as  $\mu_C$ , the specific codomain of each marker  $M_k$  was taken into account. Comparing this with the fuzzified recruiters' annotations  $Z$ , we assess the possibility of describing  $Z$  based on  $E_k$ . To do this, we use the Hamming distance between fuzzy sets to measure how closely  $E_k$  approximates  $Z$ :

$$S(E_k, Z) = \sum_{t \in U} \left( 1 - \left| \mu_{M_k}(t) - \mu_Z(t) \right| \right) \quad (6)$$

To allow for a direct interpretation of scores, we add the 1 constant to the expression, adapting an equation from a previous study [6]. High scores reflect markers closely aligned

with recruiters’ viewpoints, while low scores suggest a divergence. In this way, the Hamming distance measures the degree to which a term is relevant or not, based on the available evidence  $E_k$ .

We normalize this distance to account for potential class imbalance, represented as  $\|S(E_k, Z)\| \in [0, 1]$ . We then determine the possibility  $\pi$  of classifying a term as relevant ( $Z_1$ ) or non-relevant ( $Z_2$ ) based on recruiters’ strategies and viewpoints:

$$\pi(Z_i | E_k) = \frac{S(E_k, Z_i)}{\max(S(E_k, Z_1), S(E_k, Z_2))} \tag{7}$$

Ambiguity ties in with possibility [21]. Given  $E_k$ , a term’s ambiguity decreases as its relevance certainty increases. From  $\pi(Z_i | E_k)$ , we can compute the ambiguity level for marker  $M_k$  or evidence  $E_k$ :

$$G(E_k) = g(\pi(Z|E_k)) = \sum_{i=1}^n (\pi_i^* - \pi_{i+1}^*) \ln(i) \tag{8}$$

where  $\pi^* = \{\pi_1^*, \pi_2^*, \dots, \pi_n^*\}$  is the possibility distribution  $\pi(Z|E_k)$  permuted and sorted so that  $\pi_i^* \geq \pi_{i+1}^*$  for  $i \in \{1, \dots, n\}$  and  $\pi_{n+1}^* = 0$ . The ambiguity function  $G$  reflects the degree to which it can be inferred whether a term is relevant or not, based on the relevance marker  $M_k$ . A value of 0 denotes no ambiguity, and a value of  $\ln(n)$  denotes the maximum ambiguity level [21]. Subsequently, we normalize the ambiguity of the markers, denoted as  $N(E_k) \in [0, 1]$ , and introduce a penalty factor as follows:

$$N(E_k) = \min(\|G(E_k)\| + \gamma, 1) \tag{9}$$

We apply a penalty factor,  $\gamma$ , to markers failing to provide explicit class membership information for JA terms. The definition of  $\gamma$  is

$$\gamma = \begin{cases} 0 & \text{if } \exists Z_i (\|S(E_k|Z_i)\| > \beta) \\ 1 - \max(\|S(E_k|Z_i)\|) & \text{otherwise} \end{cases} \tag{10}$$

Here,  $\beta$  represents the tolerance factor determining the marker’s minimum clear information level about at least one class within the class set.

The penalty factor accounts for the potential impact on decision-making when a marker fails to deliver clear information despite demonstrating satisfactory performance in conventional classification. This situation can potentially lead to decisions based on weak or insufficient evidence.

The ambiguity metric enables estimating the separation accuracy between relevant and non-relevant terms from the recruiters’ viewpoints. Therefore, markers can be categorized based on ambiguity:

- *Low ambiguity markers*: These markers accurately separate relevant and non-relevant terms.
- *High ambiguity markers*: These markers struggle to differentiate between relevant and non-relevant terms. They may still perform adequately in traditional term classification processes, but could be negatively impacted when the variables they rely on are slightly modified within a different JA corpus.

### Parallel Analysis of Ambiguity and Amount of Information

To understand each marker’s potential in reducing uncertainty during decision-making, we analyze its ambiguity alongside its mutual information entropy. Let us consider a marker  $M_x$  and function  $Y$  (recruiters’ annotations on the JA terms). The mutual information entropy is as follows:

$$I(Y;M_x) = H(Y) - H(Y|M_x) \tag{11}$$

where  $H(Y)$  is the entropy of recruiters’ annotations, and  $H(Y|M_x)$  is the conditional entropy of  $M_x$  against recruiters’ annotations. The comparative analysis identifies four types of markers:

- *Type 1. Low entropy and High ambiguity*: The marker shows weak influence and provides a fuzzy distinction between relevant and non-relevant terms.
- *Type 2. Low entropy and Low ambiguity*: Despite the marker’s insignificant relationship with recruiters’ views, it may still help distinguish relevant from non-relevant terms.
- *Type 3. High entropy and High ambiguity*: The marker greatly influences the decision but fails to distinctly separate relevant from non-relevant terms.
- *Type 4. High entropy and Low ambiguity*: Ideal markers that show a significant correlation with recruiters’ views and effectively separate relevant from non-relevant terms.

After characterizing the markers, we move to analyze their explainability within the organizational context.

### Alignment of Markers with Recruiters Strategies

We ensure the explainability of selected markers by aligning them with recruiters’ strategies regarding JA term relevance. This alignment results in tuples  $(m_n, s_m, t_j)$  representing the marker, recruiter strategy, and strategy type (implicit or explicit), respectively.

To this end, the Apriori algorithm is selected for its efficiency, ease of interpretation, robustness, and ability to handle large datasets [38]. It generates a list of frequent tuples that are easily interpretable, a significant advantage as explainability is crucial in our application case. Furthermore, its adjustable sensitivity to noise and outliers enhances its adaptability.

Thus, a marker's relevance is determined by its association with a minimum level of explainability as per the recruiters' information selection strategies. This level is defined by a support level  $\alpha$  based on the Apriori algorithm's output. Notably, in this phase, examining the confidence level is not vital, as our focus lies in revealing recurring marker and recruiters' strategy associations. Detailed exploration of their predictive relationship is reserved for the next phase.

### Prospective Analysis of Markers

Moving beyond the focus on explainability, we also seek to determine the statistical significance of each marker in larger datasets. We utilize logistic regression (LR) to examine if a statistically significant relationship exists between markers and recruiters' viewpoints. Correlations identified in the first stage of the methodology (Sect. [Correlation Analysis Between Markers](#)) are a critical input, ensuring stability during regression training [39].

The choice of LR is guided by several notable advantages [40]. First, LR offers exceptional explainability, yielding clear and interpretable results. Second, its efficiency in identifying pertinent features in relation to a dependent variable aligns with our goal of information extraction. Lastly, LR is particularly resistant to overfitting when handling smaller datasets, which bolsters model stability and generalizability.

### Mamdani Fuzzy Inference Engine for Marker Selection and Evaluation

Since the selection process of relevant and non-relevant markers can be highly relative and variable according to each organizational context in which JAs are processed, fuzzy logic is incorporated to manage ambiguity and uncertainty during the evaluation process [21]. This approach is especially beneficial given the subjectivity and variability of human and organizational factors in JAs. Consequently, we have chosen a standard Mamdani fuzzy-inference engine for marker selection [41]. This method has proven efficacy in feature-selection tasks [28], demonstrates comparative

performance despite computational complexity [42], and provides interpretability [43].

In the context of our proposed assessment methodology, the Mamdani engine constructs a relevance scale for the set of markers, incorporating recruiters' perspectives. Results from each stage of our methodology are processed as follows.

#### Definition of the Engine's Input and Output Fuzzy Sets

For each marker  $M_x$  under evaluation, the inference system takes  $m$  input variables. These variables represent quality metrics (e.g., ambiguity, entropy, statistical significance) from  $m$  evaluation stages in our methodology. Through a fuzzy rule-based inference process, these inputs transform into output variables, which indicate specific quality criteria tailored to the organizational context. While our study focuses on a single output variable, representing the overall relevance level of each marker, the formalization below allows for multiple quality assessment criteria using multiple output variables.

Formally, the inference system consists of  $m$  input variables  $X_i (i = 1, \dots, m)$ , representing different marker quality metrics, and  $p$  output variables  $Y_k (k = 1, \dots, p)$ , denoting marker quality criteria. We define  $n_i$  input fuzzy sets  $A_{i,j} (j = 1, \dots, n_i)$  for each quality metric  $X_i$  and  $q_k$  output fuzzy sets  $B_{k,l} (l = 1, \dots, q_k)$  for each evaluation criterion  $Y_k$ .

By using a variable number of fuzzy sets for each quality metric, we can adjust specificity and rigor levels, enabling the development of robust fuzzy-quality rules for markers tailored to the unique needs of the organizational context dealing with JAs.

#### Creation of the Fuzzy-quality-rules Base

We define fuzzy-quality rules based on the quality metrics of the markers, allowing for an association between different metric values and each output variable.

Thus, we generate  $R$  if-then rules linking the engine input and output variables. Each rule  $r (r = 1, \dots, R)$  takes the form:

$$\begin{aligned} \text{IF } X_1 \text{ is } A_{1,j_1} \text{ AND } \dots \text{ AND } X_m \text{ is } A_{m,j_m} \text{ THEN} \\ Y_1 \text{ is } B_{1,l_1} \text{ AND } \dots \text{ AND } Y_p \text{ is } B_{p,l_p} \end{aligned} \quad (12)$$

where  $X_i$  is the quality metric  $i$ ,  $A_{i,j_i}$  is the fuzzy category  $j_i$  of metric  $i$ ,  $Y_p$  is the quality assessment criterion (output variable)  $p$ , and  $B_{p,l_p}$  is the fuzzy category  $l_p$  of the assessed criterion  $p$ .

These rules create the fuzzy-quality-rule base of the inference system, which the engine uses to estimate the evaluation criterion fulfillment based on the marker quality scores. The number of rules and the system’s sensitivity to them depend directly on the unique quality requirements of each organizational context processing JAs. These factors are foundational to fuzzy logic, and their flexibility suits each varying context [44].

**Marker Quality Scores Fuzzification**

Given a crisp input vector  $x(x_1, \dots, x_m)$ , which represents the values  $x_i$  of the quality metrics for the textual marker, we calculate the membership degree  $\mu_{A_{ij}}(x_i)$  for each quality category  $A_{ij}$  using its corresponding membership function  $f_{ij}(x_i)$ :

$$\mu_{A_{ij}}(x_i) = f_{ij}(x_i) \tag{13}$$

Here,  $f_{ij}(x_i)$  is the membership function for the fuzzy category  $A_{ij}$ , and  $x_i$  is the crisp input value for the quality metric (or input variable)  $X_i$ .

The result of fuzzification is a set of membership degrees, which represent how much the input values belong to each fuzzy category. These membership degrees are used in the subsequent steps of the fuzzy-inference process.

**Rules Evaluation**

The firing strength  $w_r$  of each quality rule, representing the degree to which the rule is activated, is calculated as follows:

$$w_r = T(\mu_{A_{1j_1}}(x_1), \dots, \mu_{A_{mj_m}}(x_m)) \tag{14}$$

Here, T is a T-norm aggregation operator, the minimum function, which combines the membership degrees of the antecedent fuzzy categories (input fuzzy sets) for a given rule. The firing strength  $w_r$  thus reflects the applicability of a rule in the given input values context.

**Consequences Evaluation**

We compute the clipped output fuzzy set  $B_{k,l_r}^r$  for each rule  $r$ , based on the firing strength obtained in the previous step:

$$\mu_{B_{k,l_r}^r}(y_k) = T(w_r, \mu_{B_{k,l_r}}(y_k)) \tag{15}$$

In this equation,  $\mu_{B_{k,l_r}}(y_k)$  is the membership degree of the clipped output fuzzy set  $B_{k,l_r}^r$  for rule  $r$ .  $w_r$  is the firing

strength of rule  $r$ , and  $\mu_{B_{k,l_r}}(y_k)$  is the membership degree of the original output fuzzy category (or fuzzy set)  $B_{k,l_r}$ . In addition, the term  $y_k$  represents any value within the domain of the output variable  $Y_k$ .

In this way, we modify the original output fuzzy sets (consequent part of the rule) by clipping them based on the firing strength  $w_r$ . This step ensures that the quality rule’s influence on the final output is proportional to the rule’s firing strength.

**Output Aggregation**

We combine the clipped output fuzzy sets into a single aggregated fuzzy set  $B_k$  for each quality assessment criterion  $Y_k$ :

$$\mu_{B_k}(y_k) = S(\mu_{B_{k,l_1}}^1(y_k), \dots, \mu_{B_{k,l_r}}^r(y_k), \dots, \mu_{B_{k,l_q}}^q(y_k)) \tag{16}$$

In this equation,  $\mu_{B_k}(y_k)$  is the membership degree of the aggregated output fuzzy set  $B_k$  for the evaluation criterion  $Y_k$ . S is a S-norm aggregation operator, the maximum function, which combines the membership degrees of the clipped output fuzzy from each rule.

The result is a single fuzzy set that embodies the collective impact of all quality rules on the output variable(s).

**Output Defuzzification**

The crisp output value for each assessment criterion  $Y_k$  is computed using the centroid method, chosen for its balance of accuracy, response time, and resource efficiency:

$$y_k^* = \frac{\int y_k \mu_{B_k}(y_k) dy_k}{\int \mu_{B_k}(y_k) dy_k} \tag{17}$$

In this equation,  $y_k^*$  is the crisp output value for the quality assessment criterion (output variable)  $Y_k$ . The integral in the numerator calculates the weighted sum of the output variable values ( $y_k$ ) multiplied by their corresponding membership degrees in the aggregated output fuzzy set ( $\mu_{B_k}(y_k)$ ). The integral in the denominator calculates the sum of the membership degrees in the aggregated output fuzzy set ( $\mu_{B_k}(y_k)$ ). The crisp output value  $y_k^*$  is obtained by dividing the weighted sum (numerator) by the sum of membership degrees (denominator).

The proposed Mamdani-type inference engine simplifies the decision-making process, providing a crisp value for each criterion for every marker. This enables a robust understanding of the textual markers’ quality within the

specific JA's application context and the possibilistic BDI agent architecture, enhancing system performance and optimizing outcomes in the JA information retrieval organizational environment.

## Experimentation

The evaluation of the proposed methodology was conducted within the human resources department of DSI Group,<sup>4</sup> a French consulting firm operating in the information and communications technology sector. We evaluated the relevance of 30 defined textual markers across 73 recruitment processes, which included 73 JAs and 361 corresponding CVs. The goal was to enhance the automatic extraction of key terms from JAs for improved matching with CVs. The recruitment processes spanned several days or weeks, allowing recruiters to gain a deep understanding of each JA's key requirements.

The evaluation process comprised four stages:

1. *Term annotation in JAs*: Ten recruiters annotated critical requirements from a corpus of 73 JAs, identifying 580 relevant terms and 15,487 non-relevant terms.
2. *Division of the collected dataset into training and test partitions or folds*: The 10-fold cross-validation technique was employed, with the metrics for marker relevance evaluation (ambiguity, entropy, etc.) being assessed for each fold, and their average reported. Furthermore, since the ambiguity metric can be highly sensitive to class imbalance, a class balancing process is carried out within each fold to more rigorously estimate it, ensuring that 35% of the terms are relevant and 65% are not relevant [45].
3. *Textual marker selection*: We examined the performance of 30 textual markers to model the recruiters' annotations.
4. *BDI Architecture implementation*: The markers identified as the most relevant after the evaluation process were introduced into the architecture of a Possibilistic BDI agent. Their relevance was further confirmed through a two-step automatic extraction process of relevant terms. Initially, this process was applied to the test dataset folds (14 JAs per fold) of the current study's corpus. Subsequently, it was applied to a previously analyzed corpus, consisting of 20 JAs [6].

## Evaluated Textual Markers

As highlighted in this study, we evaluated 30 textual markers, which included markers from a previous study [6] ( $M_1-M_{16}$ ), additional ones derived from a fuzzy decision tree ( $M_{17}-M_{20}$ ), and one marker obtained via the PCA method for dimension reduction ( $M_{21}$ ). In addition, we tested nine false markers for comparison ( $M_{22}-M_{30}$ ), consisting of logical rules that, in principle, have no meaningful relevance.

## Mamdani Inference Engine Settings

Details of the Mamdani-type inference engine's implementation include:

- *Markers' possibility degrees calculation*: Considering that the objective of this investigation is not focused on the estimation of the marker's possibility degrees, a simplified approach was used for assessing quality metrics. For markers without a specific possibility degree formula  $\alpha_{i,k}$ , their possibility degrees were assumed equal to their truth values.
- *Fuzzy categories*: A minimum of 2 quality categories were selected for the less relevant quality metrics in the organizational context, and a maximum of 7 quality categories for the more critical metrics (e.g., ambiguity and negative precision) to model them with more detailed and rigorous rules.
- *Membership functions*: We used standard triangular membership functions for all quality categories.
- *Evaluation criterion*: We assessed the overall quality (QA) of each textual marker based on their metrics.

## Experimental Results

This section delineates the evaluation outcomes at each stage of our quality assessment framework. The figures presented show the applicability and usefulness of quality metrics in the specific organizational context. They unveil the markers' relevance, visually indicating how the modeling rigor for each metric and the associated fuzzy rules can be adjusted when introduced into the inference system:

1. *Identification of correlations*: Fig. 4 illustrates the Pearson correlation levels between each marker and the variable R, which denotes the recruiters' annotations or viewpoints.
2. *Evaluation of markers frequency*: Fig. 5 presents the logarithm of the accumulated possibility levels or frequency of each marker.
3. *Assessment of markers precision on both relevant and non-relevant terms*: Fig. 6 displays the precision evalu-

<sup>4</sup> <https://www.group-dsi.com/>

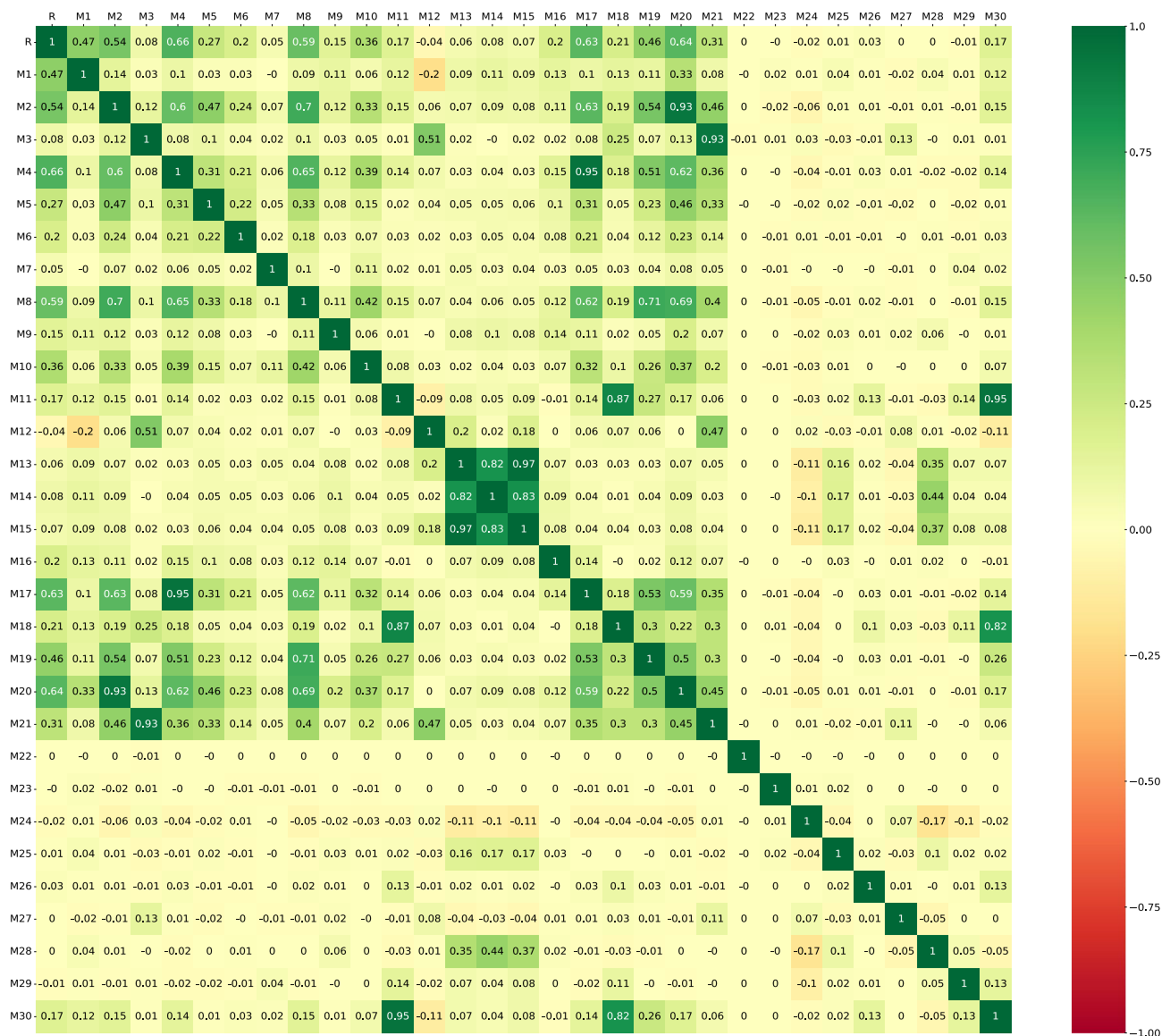


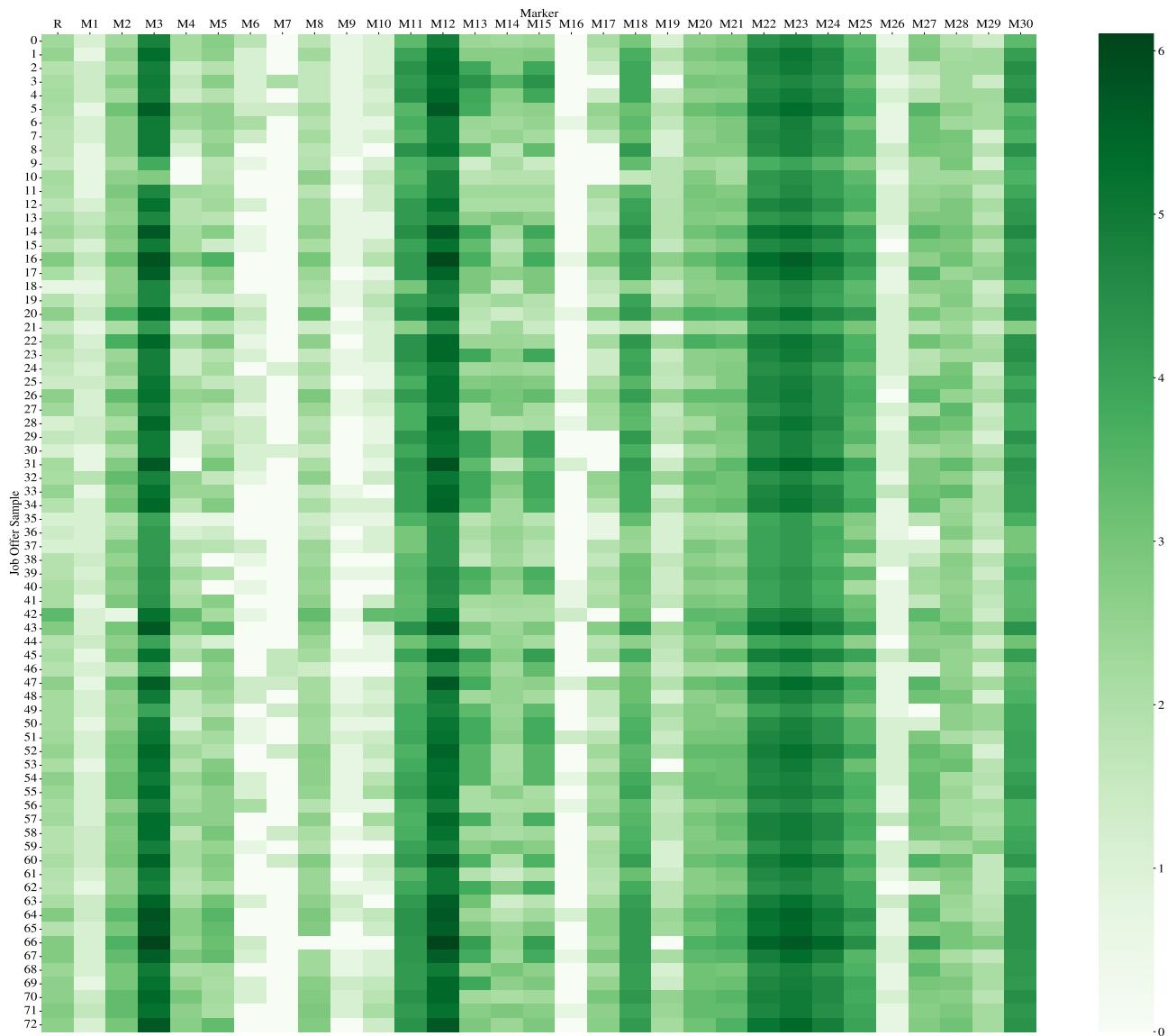
Fig. 4 Pearson correlation coefficients between the markers (M<sub>x</sub>) and recruiters annotations (R). Darker colors indicate higher levels of correlation

ation results of each marker on pertinent (x-axis) and non-pertinent terms (y-axis).

- 4. *Ambiguity estimation*: The estimated ambiguity levels of each of the evaluated markers are illustrated in Fig. 7.
- 5. *Comparison of markers ambiguity and mutual information entropy*: Fig. 8 displays a comparative chart consisting of the normalized ambiguity dimension (y-axis) and the normalized mutual information entropy dimension (x-axis) of each marker.
- 6. *Prospective analysis*: The prospective analysis, conducted by training a LR model, is presented in Table 4 (in the column P<sup>?</sup>). It displays whether the textual markers were detected as highly significant across the folds of

the 10-fold cross-validation process, with a confidence level of 95% (Wald test).

- 7. *Alignment of textual markers with respect to recruiters strategies*: Table 4 also presents (in the column labeled “Rec?”), both context-dependent and context-independent markers that showed minimal association with the recruiters’ strategies.
- 8. *Comprehensive quality results*: Table 4 also delivers the comprehensive quality outcomes at each evaluation stage. The QA column represents the defuzzification result from the inference engine, while the Pertinence column signifies the fuzzy category corresponding to each marker’s relevance. Complementarily, Table 5 pre-



**Fig. 5** Log of accumulated possibility degrees of textual markers ( $M_x$ ) and recruiters annotations (R). Over-activated markers are depicted in a darker color and underactivated markers in a lighter color

sents an example of a complex fuzzy quality rule derived from the application case.

9. *BDI Agent performance*: Finally, Table 6 presents the performance results of the BDI agent on the test dataset folds in the task of relevant term extraction from JAs, using the most optimally identified textual markers. It's important to highlight that the performance of the evaluated state-of-the-art techniques on this dataset was not superior to the results we had already reported in our previous research [6].

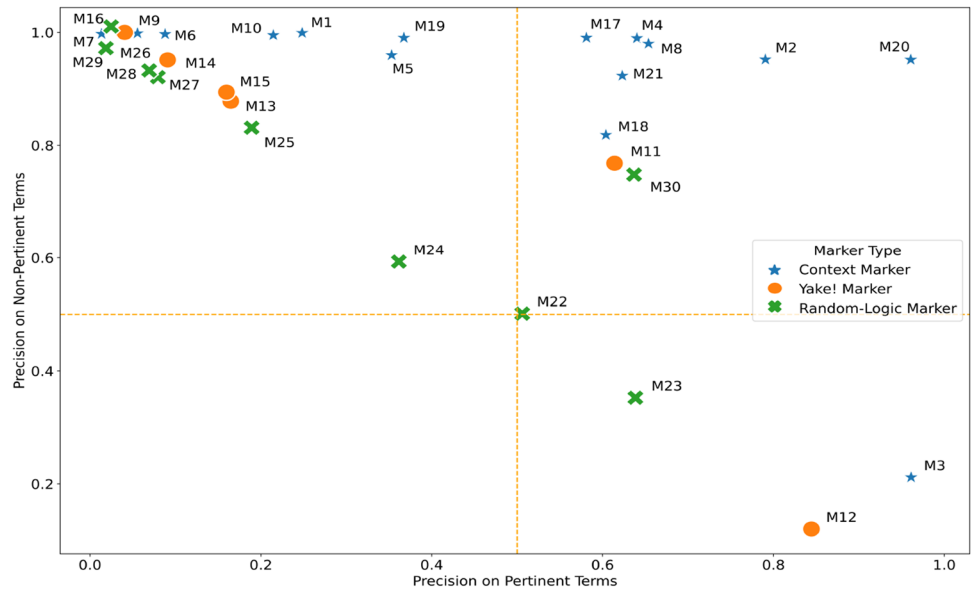
Next, Table 7 provides the performance results on a JAs corpus from the earlier study [6], illustrating the performance outcomes prior to and following the imple-

mentation of our proposed textual marker assessment methodology.

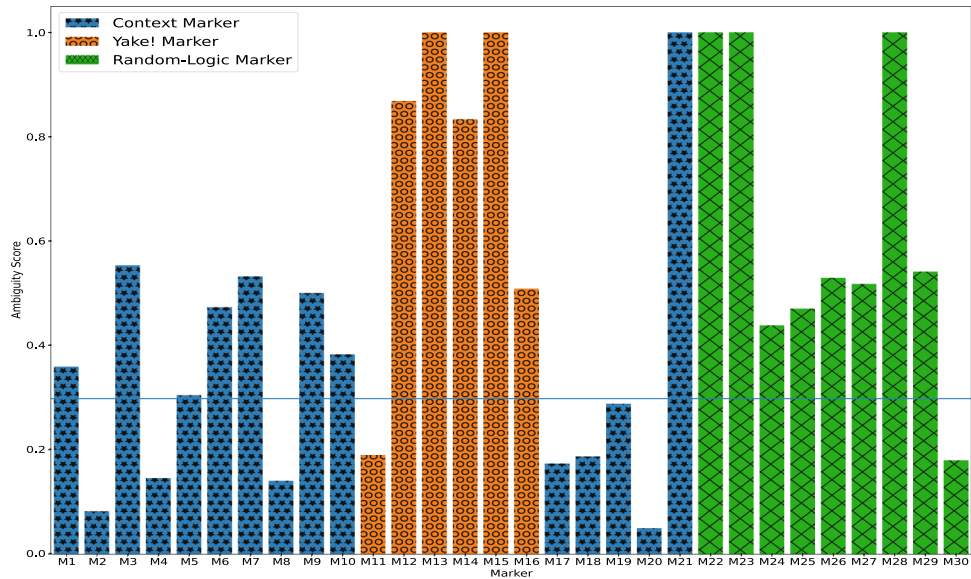
To obtain these results, we employed classical metrics in Information Retrieval: Precision@N, Recall@N, and  $F1$ -score@N. Here, @N refers to the number of terms annotated by a recruiter as relevant in the JA. Therefore, these metrics are defined as follows:

- **Precision@N**: Precision at N assesses the proportion of terms correctly predicted by the model among the top N terms identified as relevant. Precision@N is normalized by the total number of the top N terms identified by the model. It is formally defined as

**Fig. 6** Markers precision on pertinent and non-pertinent terms. Blue stars for context-derived markers, orange circles for YAKE! markers, and green crosses for false markers



**Fig. 7** Textual markers estimated ambiguity degrees. Context-derived markers are indicated in blue with stars, YAKE! markers in orange with circles, and false markers in green with crosses



$$Precision@N = \frac{|{\{T\} \cap \{R\}}|}{N} \tag{18}$$

where  $\{T\}$  is the set of terms identified by the model among top  $N$  and  $\{R\}$  is the set of terms annotated by recruiters as relevant.

- **Recall@N:** Recall at rank  $N$  measures the proportion of truly relevant terms that the model manages to identify among the top  $N$ . Recall@ $N$  is normalized by the total number of relevant terms in the

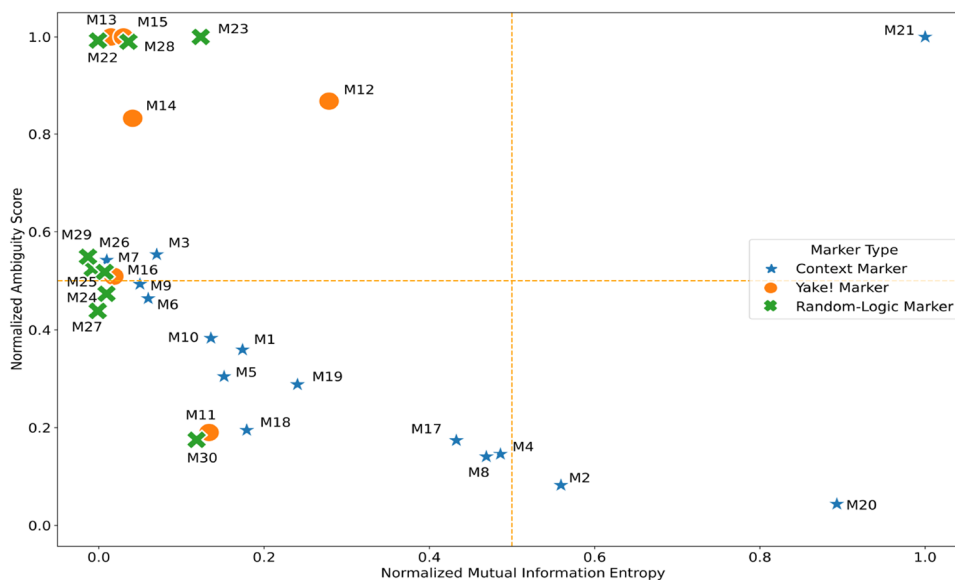
JA, denoted as  $|\{R\}|$  and annotated by recruiters. It is formally defined as

$$Recall@N = \frac{|{\{T\} \cap \{R\}}|}{|\{R\}|} \tag{19}$$

- **F1-score@N:** F1-score at rank  $N$  provides a harmonic mean of precision and recall at rank  $N$ . It is formally defined as



**Fig. 8** Textual markers ambiguity and mutual information entropy. Blue stars for context-derived markers, orange circles for YAKE! markers, and green crosses for false markers



**Table 4** Results of the textual markers evaluation on the studied corpus. The applied evaluation metrics are Pearson (Pearson correlation coefficient), Prec+ (precision on pertinent terms), Prec- (precision on non-pertinent terms), Amb. (normalized textual marker ambiguity), MI (normalized marker mutual information), P? (whether the marker was identified as statistically significant across the dataset folds), and Rec? (whether the marker is minimally associated with recruiter strategies). The QA (Quality Assessment) column corresponds to the Mamdani inference engine score obtained using the defuzzification centroid method. The Pertinence column indicates the corresponding fuzzy category of the QA score. The standard deviation of each evaluated marker metric ranged between approximately 0 and 0.03, except for marker 24, which exhibited a standard deviation of 0.21 for the ambiguity metric

Marker	Pearson	Prec+	Prec-	Amb	MI	P?	Rec?	QA	Pertinence
M <sub>1</sub>	0.47	0.25	0.99	0.36	0.17	Yes	Yes	0.50	Medium
M <sub>2</sub>	0.54	0.79	0.95	0.08	0.56	Yes	Yes	0.82	High
M <sub>3</sub>	0.08	0.96	0.21	0.55	0.07	No	Yes	0.50	Medium
M <sub>4</sub>	0.66	0.64	0.99	0.15	0.49	Yes	Yes	0.84	High
M <sub>5</sub>	0.27	0.35	0.96	0.3	0.15	Yes	Yes	0.50	Medium
M <sub>6</sub>	0.2	0.09	0.99	0.47	0.05	Yes	Yes	0.50	Medium
M <sub>7</sub>	0.05	0.01	0.99	0.53	0.01	No	Yes	0.50	Medium
M <sub>8</sub>	0.59	0.65	0.98	0.14	0.47	Yes	Yes	0.84	High
M <sub>9</sub>	0.15	0.05	0.99	0.5	0.03	Yes	Yes	0.50	Medium
M <sub>10</sub>	0.36	0.21	0.99	0.38	0.14	Yes	Yes	0.50	Medium
M <sub>11</sub>	0.17	0.61	0.77	0.19	0.13	Yes	Yes	0.68	Medium-high
M <sub>12</sub>	-0.04	0.84	0.12	0.87	0.28	No	Yes	0.25	Low-medium
M <sub>13</sub>	0.06	0.16	0.89	1.00	0.01	No	No	0.17	Low
M <sub>14</sub>	0.08	0.09	0.95	0.83	0.04	No	No	0.24	Low
M <sub>15</sub>	0.07	0.16	0.89	1.00	0.03	No	No	0.17	Low
M <sub>16</sub>	0.2	0.04	0.99	0.51	0.02	No	No	0.21	Low
M <sub>17</sub>	0.63	0.58	0.99	0.17	0.43	Yes	Yes	0.81	High
M <sub>18</sub>	0.21	0.6	0.82	0.19	0.18	Yes	Yes	0.74	Medium-high
M <sub>19</sub>	0.46	0.37	0.99	0.29	0.24	Yes	Yes	0.59	Medium-high
M <sub>20</sub>	0.64	0.96	0.95	0.05	0.88	Yes	Yes	0.83	High
M <sub>21</sub>	0.31	0.62	0.92	1.00	1.00	Yes	Yes	0.25	Low-medium
M <sub>22</sub>	0.00	0.51	0.5	1.00	0.01	No	No	0.16	Low
M <sub>23</sub>	0.00	0.64	0.35	1.00	0.12	No	No	0.18	Low
M <sub>24</sub>	-0.02	0.36	0.59	0.44	0.00	No	No	0.19	Low
M <sub>25</sub>	0.01	0.19	0.83	0.47	0.00	No	No	0.17	Low
M <sub>26</sub>	0.03	0.02	0.99	0.53	0.01	No	No	0.19	Low
M <sub>27</sub>	0.00	0.08	0.92	0.52	0.01	No	No	0.19	Low
M <sub>28</sub>	0.00	0.07	0.93	1.00	0.03	No	No	0.17	Low
M <sub>29</sub>	-0.01	0.02	0.97	0.54	0.00	No	No	0.17	Low
M <sub>30</sub>	0.17	0.64	0.75	0.18	0.13	Yes	Yes	0.70	Medium-high

**Table 5** Example of a fuzzy rule for marker quality assessment (QA) evaluation

<b>IF</b>	Recruiters strategies association	<b>IS NOT</b>	LOW
<b>AND</b>	Ambiguity	<b>IS</b>	(VERY LOW OR LOW OR LOW-MEDIUM)
<b>AND</b>	P-value recurrency	<b>IS NOT</b>	LOW
<b>AND</b>	Negative precision	<b>IS</b>	(HIGH OR VERY HIGH)
<b>AND</b>	Positive precision	<b>IS NOT</b>	LOW
<b>AND</b>	Mutual information	<b>IS NOT</b>	LOW
<b>AND</b>	Pearson correlation	<b>IS NOT</b>	LOW
<b>THEN</b>	Marker QA	<b>IS</b>	MEDIUM-HIGH

**Table 6** Experimental results on the JAs dataset test folds. Mean and standard deviation of precision@N, recall@N, and F1-score@N levels of the BDI Agent

Metric/method	Mean value	Standard deviation
Recall@N, Precision@N F1-score@N <sup>1</sup>	0.56	0.028
Recall@2N	0.67	0.038
Precision@2N	0.34	0.019
F1-score@2N	0.45	0.026

<sup>1</sup> Recall@N, Precision@N and F1-score@N are equivalent at @N

$$F1 - score@N = 2 \cdot \frac{Precision@N \cdot Recall@N}{Precision@N + Recall@N} \tag{20}$$

## Discussion

The analysis of accumulated possibility levels (Fig. 5) delineates two distinct groups of markers: the high activation frequency group ( $M_3, M_{12},$  and  $M_{23}$ ) and the low activation frequency group ( $M_6, M_7, M_9,$  and  $M_{16}$ ). High frequency

markers could be instrumental in pinpointing document sections rich in relevant terms, whereas low frequency markers might detect relevant terms in specific contexts. Marker  $M_7$ , for instance, proves particularly effective in examining JAs in the management sector.

Correlation evaluation (Fig. 4) discloses significant correlations between YAKE! markers  $M_{13}, M_{14},$  and  $M_{15}$ , due to their dependence on term frequency in the text. The mutual influence of the random logic marker  $M_{30}$  and the YAKE! marker  $M_{11}$  is notable, attributable to their focus on upper-case term detection. Similarly, context markers  $M_2, M_4,$  and  $M_8$  manifest positive correlations, attributed to their emphasis on professional skills. Understanding these correlations can offer insights into the interrelation of markers and their efficiency.

On the other hand, from the perspective of this same metric, composite markers  $M_{17}-M_{20}$ , derived from the simplification of a decision fuzzy tree, correspond strongly with the context markers that formed their foundation, underlining the recruiters’ focus on professional skills. Conversely, false markers  $M_{22}-M_{29}$  correlate negligibly with recruiters’ perspectives. This discrepancy underscores the value of correlation analysis in comprehending marker relevance.

In terms of precision (Fig. 6), contextual markers ( $M_1 - M_{10}$ ) and composite marker  $M_{20}$  exhibit high performance in identifying irrelevant terms in JAs—a crucial attribute given the generally limited set of highly relevant terms in these documents. Context-derived markers such as  $M_2, M_4, M_8,$  and composite markers  $M_{17}, M_{20},$  and  $M_{21}$  demonstrate a high precision for both relevant and irrelevant terms. However,  $M_3$  and  $M_{12}$ , while efficient in recognizing relevant terms, display lower precision for irrelevant terms. This phenomenon likely results from their broad classification of terms in specific document sections as potentially relevant.

The random marker  $M_{22}$ , which maintains a 50% precision for each term class, is remarkable. However, false markers, which are more restrictive, tend to show higher precision for irrelevant terms and lower for relevant ones,

**Table 7** Experimental results on a previous corpus of JAs [6]. Precision, recall, and F1-score levels of each method on a previous corpus of 20 JAs applying RAKE [46], FRAKE [47], BERT topics approach [48], YAKE! [36] and our agent Before [6] and After applying the textual markers assessment methodology

Metric/method	RAKE	FRAKE	BERT	YAKE!	Before	After
Recall@N, Precision@N F1-score@N <sup>1</sup>	0.02	0.09	0.17	0.10	0.38	0.55
Recall@2N	0.08	0.17	0.20	0.27	0.56	0.64
Precision@2N	0.04	0.08	0.10	0.13	0.28	0.32
F1-score@2N	0.05	0.11	0.14	0.18	0.38	0.43

<sup>1</sup> Recall@N, Precision@N and F1-score@N are equivalent at @N

influenced by the high prevalence of non-relevant terms in the JA corpus. This can introduce uncertainty in the evaluation parameter. Consequently, the accuracy of contextual markers, YAKE! marker  $M_{11}$ , and random logic marker  $M_{30}$  in representing relevant and irrelevant terms within JAs is significant.

Ambiguity analysis (Fig. 7) shows low ambiguity in markers  $M_2$ ,  $M_4$ ,  $M_5$ , and  $M_8$ . However, most context-independent and random logic markers demonstrate high ambiguity. Notably, marker  $M_{21}$  exhibits considerable ambiguity despite satisfactory precision in a traditional classification process, potentially complicating term classification and increasing ambiguity in approximate reasoning processes as the one performed by the BDI agent.

When evaluating mutual information entropy (Fig. 8), markers  $M_2$ ,  $M_4$ ,  $M_8$ ,  $M_{17}$ , and  $M_{20}$ , and to a lesser extent  $M_1$ ,  $M_{19}$ ,  $M_{12}$ ,  $M_{10}$ , and  $M_5$ , are marked by low ambiguity and relatively high mutual information entropy. Conversely, random marker  $M_{22}$  and context-independent markers such as  $M_{13}$  and  $M_{15}$ , present high ambiguity and low information entropy. Most false markers also show minimal mutual information entropy and high ambiguity levels, confirming their inappropriateness.

In the prospective analysis of markers (Table 4), context markers  $M_1$ ,  $M_2$ ,  $M_4$ ,  $M_5$ ,  $M_6$ ,  $M_8$ ,  $M_9$ ,  $M_{10}$ , and YAKE! markers  $M_{11}$  and  $M_{12}$  show significant systematic associations with recruiters' annotations on the evaluated dataset folds. Certain strategies associated with these markers include recruiters often incorporating a representative name of the profession in the job title along with pertinent skills ( $M_1$ ); selecting terms representing professional skills located in relevant document sections ( $M_2$ ); associating expressions of relevance with important skills ( $M_4$ ); providing details on the type of interaction required from the worker regarding relevant terms ( $M_5$ ); using highly relevant terms (e.g., within the document's title) to denote other terms that are semantically related and of high relevance to the job position ( $M_8$ ); and accentuating the significance of a term if it corresponds to a pertinent economic sector of the job opening or, more prominently, an essential professional skill ( $M_{20}$ ). This last marker indicates, to a further extent, that professional skills are typically emphasized through a variety of methods, including their incorporation into the job title, their placement within the most pertinent sections of the JA (like the job description), or their amplification through explicit or implicit relevance expressions. A term is likely deemed potentially irrelevant if it does not meet any of these criteria.

In evaluating markers' relationships with the organizational context, the analysis of markers such as YAKE!'s  $M_{11}$  and  $M_{30}$  (random logic) emphasizes their limitations, particularly in discerning relevant terms. Despite recruiters

often writing professional skill names in uppercase or capitalized letters, these context-independent markers still categorize wrongly irrelevant terms as relevant. This misclassification includes terms such as the first ones following a period, company names, geographic locations, and list titles. These errors indicate the need for refining these markers and present opportunities for improvement.

The Mamdani-type fuzzy-inference system emphasizes the utility of markers  $M_2$ ,  $M_4$ ,  $M_8$ ,  $M_{17}$ , and  $M_{20}$  in extracting information from JAs. These markers have resulted in a 5% increase in the  $F1$ -score compared to the previous study [6], as illustrated in Table 7. This high performance remains relatively consistent across the test dataset folds. Most context-derived markers, such as  $M_1$ ,  $M_5$ ,  $M_6$ ,  $M_7$ ,  $M_8$ ,  $M_9$ ,  $M_{10}$ ,  $M_{18}$  and  $M_{19}$  were classified as relatively relevant due to their statistical significance, explainability, high precision on non-relevant terms, and medium-low ambiguity levels. Even though  $M_3$  was considerably penalized for its diminished precision, it can serve an essential role in initial filtering. Conversely,  $M_{21}$  exhibits low performance, explained by its highly fuzzy separation between relevant and non-relevant terms. The remaining markers ( $M_{22}$ ,  $M_{13}$ ,  $M_{14}$ ,  $M_{15}$ ,  $M_{16}$ , and most false markers) were considered weakly relevant.

It should be noted that markers identified as highly and moderately relevant align with recruiters' selection strategies. Some of these markers, notably  $M_1$ ,  $M_6$ ,  $M_7$ ,  $M_9$ ,  $M_{10}$  and  $M_{19}$ , have high precision for non-relevant terms. Despite their low positive precision and relative ambiguity in the studied corpus, their statistical significance suggests potential higher relevance in other JA study corpora. Incorporating these markers into the agent's information extraction process improved its  $F1$ -score@2N by 0.7–1% in the test dataset folds and the previous study corpus [6]. Thus, these markers do not negatively impact the agent's performance but enhance its explainability, providing the agent with a more comprehensive understanding due to the diverse hypotheses these markers offer.

Throughout the experimentation, the variability in each evaluated marker metric across the dataset folds, measured by the standard deviation, remained comparatively small. The lowest observed value was near zero, while the upper bound hovered around 0.03. This is primarily attributable to the markedly low ratio of pertinent terms in every JA. As a result, markers typically display consistent levels of performance—either good, medium, or low—across all assessed quality metrics for each fold. However, marker  $M_{24}$  deviated from this pattern with a standard deviation of 0.21 on the ambiguity metric, which can be ascribed to pronounced ambiguity peaks in JAs where recruiters favor a higher number of simple (single-word) terms as relevant.

## Conclusion

The primary aim of this study was to address the following research question: “How can a quality assessment framework optimize the use of textual markers to accurately extract the most relevant terms from JAs? Furthermore, how can this process effectively manage ambiguity, ensure alignment with recruiter expertise, maintain information quantity, and uphold a context-driven, explainable approach?”

We proposed a hybrid methodology to select and optimize textual markers for automatic term extraction in JAs. This approach assesses markers based on the information they provide and their relevance in the specific organizational context processing these documents. Our methodology incorporates a multidimensional quality assessment framework to characterize markers, addressing ambiguity, information quantity, and relative significance, while aligning with recruiters’ expertise.

One limitation was the high cost of creating a reliable corpus of annotated JAs. Consequently, we aim to gradually apply the methodology to more recruitment processes, recruiters, and organizational contexts, to enhance our findings’ generalizability across different application domains, like job search engines.

Future research will focus on characterizing the ambiguity function via a fuzzy approach tailored to each organization’s unique attributes. We will also explore the areas of rule number optimization and sensitivity analysis within different organizational contexts for JA information extraction. In addition, we aim to study the correlation between recruiters’ and non-expert annotators’ perspectives to discern whether the knowledge encapsulated by the most relevant textual markers is general or specialized.

**Acknowledgements** This study would not have been possible without the invaluable collaboration of Bassem Mahmoud, Director of the Human Resources Department at DSI Group, alongside the recruitment team, who provided us with valuable expert knowledge and experience.

**Funding** This study was funded by DSI Group, France.

**Data Availability** In our specific context of application, and given the nature of our field of research centered on automatic processing of CVs and JAs, both the data and the code encompass sensitive and confidential information, rendering them unsuitable for publication.

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Ethical Approval** All the procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed Consent** All the individual participants included in the study, or their legal guardians where applicable, provided informed consent.

## References

1. Wang X, Jiang Z, Peng L. A deep-learning-inspired person-job matching model based on sentence vectors and subject-term graphs. *Complexity*. 2021;2021:1–11. <https://doi.org/10.1155/2021/6206288>.
2. Cabrera-Diego LA, El-Béze M, Torres-Moreno JM, Durette B. Ranking résumés automatically using only résumés: a method free of job offers. *Expert Syst Appl*. 2019;123:91–107. <https://doi.org/10.1016/j.eswa.2018.12.054>.
3. Roy PK, Chowdhary SS, Bhatia R. A machine learning approach for automation of resume recommendation system. *Procedia Comput Sci*. 2020;167:2318–27.
4. Jović A, Brkić K, Bogunović N. A review of feature selection methods with applications. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE; 2015, pp. 1200–5.
5. Pavlick E, Kwiatkowski T. Inherent disagreements in human textual inferences. *Trans Assoc Comput Linguist*. 2019;7:677–94.
6. Espinal A, Haralambous Y, Bedart D, Puentes J. An ontology-based possibilistic framework for extracting relevant terms from job advertisements. *Proc Fourteenth Int Jt Conf Comput Intell*. 2022. <https://doi.org/10.5220/0011521700003332>.
7. Jurafsky D, Martin JH. *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River: Pearson Prentice Hall; 2009. p. 77–8.
8. Jensen MM. Linking place and mind: Localness as a factor in socio-cognitive salience. *Front Psychol*. 2016;7:74–74. <https://doi.org/10.3389/fpsyg.2016.01143>.
9. Zhao J, Wang J, Sigdel M, Zhang B, Hoang P, Liu M, Korayem M. Embedding-based recommender system for job to candidate matching on scale; 2021. [arXiv:2107.00221v1](https://arxiv.org/abs/2107.00221v1)
10. Pintas JT, Fernandes LAF, Garcia ACB. Feature selection methods for text classification: a systematic literature review. *Artif Intell Rev*. 2021;54(8):6149–200. <https://doi.org/10.1007/s10462-021-09970-6>.
11. Lu Y, Liang M, Ye Z, Cao L. Improved particle swarm optimization algorithm and its application in text feature selection. *Appl Soft Comput*. 2015;35:629–36.
12. Nafis NSM, Awang S. An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access*. 2021;9:52177–92.
13. Gokalp O, Tasci E, Ugur A. A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Syst Appl*. 2020;146:113176.
14. Feofanov V, Devijver E, Amini M-R. Wrapper feature selection with partially labeled data. *Appl Intell*. 2022;52(11):12316–29.
15. Manning CD. *An introduction to information retrieval*. Cambridge: Cambridge University Press; 2009.
16. Shang W, Huang H, Zhu H, Lin Y, Qu Y, Wang Z. A novel feature selection algorithm for text categorization. *Expert Syst Appl*. 2007;33(1):1–5. <https://doi.org/10.1016/j.eswa.2006.04.001>.
17. Nasir IM, Khan MA, Yasmin M, Shah JH, Gabryel M, Scherer R, Damaševičius R. Pearson correlation-based feature selection for document classification using balanced training. *Sensors*. 2020;20(23):6793.

18. Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. *Proc Fourteenth Int Conf Mach Learn.* 1997;97:412–20.
19. Lee C, Lee GG. Information gain and divergence-based feature selection for machine learning-based text categorization. *Inf Process Manag.* 2006;42(1):155–65. <https://doi.org/10.1016/j.ipm.2004.08.006>.
20. Paninski L. Estimation of entropy and mutual information. *Neural Comput.* 2003;15(6):1191–253.
21. Yuan Y, Shaw MJ. Induction of fuzzy decision trees. *Fuzzy Sets Syst.* 1995;69(2):125–39. [https://doi.org/10.1016/0165-0114\(94\)00229-Z](https://doi.org/10.1016/0165-0114(94)00229-Z).
22. Kinney JB, Atwal GS. Equitability, mutual information, and the maximal information coefficient. *Proc Natl Acad Sci.* 2014;111(9):3354–9.
23. Kumbhar P, Mali M. A survey on feature selection techniques and classification algorithms for efficient text classification. *Int J Sci Res.* 2016;5(5):1267–75.
24. Hara S, Maehara T. Enumerate lasso solutions for feature selection. *Proc AAAI Conf Artif Intell.* 2017. <https://doi.org/10.1609/aaai.v31i1.10793>.
25. Zainuddin N, Selamat A, Ibrahim R. Hybrid sentiment classification on twitter aspect-based sentiment analysis. *Appl Intell.* 2018;48:1218–32.
26. Li Q, He L, Lin X. Dimension reduction based on categorical fuzzy correlation degree for document categorization. In: 2013 IEEE International Conference on Granular Computing (GrC); 2013, p. 186–190.
27. Zuo Z, Li J, Anderson P, Yang L, Naik N. Grooming detection using fuzzy-rough feature selection and text classification. In: 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE); 2018, pp. 1–8.
28. Aghaeipoor F, Javid MM. A hybrid fuzzy feature selection algorithm for high-dimensional regression problems: an mrmr-based framework. *Expert Syst Appl.* 2020;162:113859. <https://doi.org/10.1016/j.eswa.2020.113859>.
29. Martin Jr. D, Prabhakaran V, Kuhlberg J, Smart A, Isaac WS. Extending the machine learning abstraction boundary: a complex systems approach to incorporate societal context; 2020. [arXiv:2006.09663](https://arxiv.org/abs/2006.09663)
30. Le Vrang M, Papantoniou A, Pauwels E, Fannes P, Vandenstein D, De Smedt J. ESCO: boosting job matching in Europe with semantic interoperability. *Computer.* 2014;47(10):57–64.
31. Breugh JA. Employee recruitment. *Annu Rev Psychol.* 2013;64:389–416.
32. Zapata Jaramillo CM, Arango Isaza F. The UNC-method: a problem-based software development method. *Ingeniería e Investigación.* 2009;29:69–75.
33. Hovy D, Yang D. The importance of modeling social factors of language: theory and practice. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*; 2021, pp. 588–602.
34. Da Costa Pereira C, Tettamanzi AGB. An integrated possibilistic framework for goal generation in cognitive agents. *Proc Ninth Int Conf Auton Agents Multiagent Syst.* 2010;1:1239–46.
35. Cram D, Daille B. Terminology extraction with term variant detection. In: *Proceedings of ACL-2016 System Demonstrations*; 2016, pp. 13–18.
36. Campos R, Mangaravite V, Pasquali A, Jorge AM, Nunes C, Jatowt A. Yake! collection-independent automatic keyword extractor. In: Pasi G, Piwowarski B, Azzopardi L, Hanbury A, editors. *Advances in information retrieval.* Cham: Springer; 2018. p. 806–10.
37. Dencoux T, Dubois D, Prade H. In: Marquis P, Papini O, Prade H, editors. *Representations of uncertainty in artificial intelligence: probability and possibility.* Cham: Springer; 2020. p. 69–117. [https://doi.org/10.1007/978-3-030-06164-7\\_3](https://doi.org/10.1007/978-3-030-06164-7_3).
38. Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases. In: *Proceedings of the 20th international conference on very large data bases.* VLDB 94, vol 1215. San Francisco: Morgan Kaufmann Publishers Inc; 1994. p. 487–99.
39. Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied logistic regression*, vol. 398. 3rd ed. New York: John Wiley and Sons; 2013. p. 313–76.
40. Rosenfeld A, Richardson A. Explainability in human-agent systems. *Auton Agent Multi-Agent Syst.* 2019;33(6):673–705. <https://doi.org/10.1007/s10458-019-09408-y>.
41. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Man Mach Stud.* 1975;7(1):1–13. [https://doi.org/10.1016/S0020-7373\(75\)80002-2](https://doi.org/10.1016/S0020-7373(75)80002-2).
42. Blej M, Azizi M. Comparison of mamdani-type and sugeno-type fuzzy inference systems for fuzzy real time scheduling. *Int J Appl Eng Res.* 2016;11(22):11071–5.
43. Precup R-E, Hellendoorn H. A survey on industrial applications of fuzzy control. *Comput Ind.* 2011;62(3):213–26. <https://doi.org/10.1016/j.compind.2010.10.001>.
44. Duțu L-C, Mauris G, Bolon P. A fast and accurate rule-based generation method for mamdani fuzzy systems. *IEEE Trans Fuzzy Syst.* 2018;26(2):715–33. <https://doi.org/10.1109/TFUZZ.2017.2688349>.
45. Lin W-C, Tsai C-F, Hu Y-H, Jhang J-S. Clustering-based under-sampling in class-imbalanced data. *Inf Sci.* 2017;409–410:17–26. <https://doi.org/10.1016/j.ins.2017.05.008>.
46. Rose S, Engel D, Cramer N, Cowley W. *Automatic keyword extraction from individual documents.* New York: John Wiley and Sons; 2010. p. 1–20. <https://doi.org/10.1002/9780470689646.CH1>.
47. Zehtab-Salmasi A, Feizi-Derakhshi M-R, Balafar M-A. FRAKE: fusional real-time automatic keyword extraction. 2021. [arXiv:2104.04830](https://arxiv.org/abs/2104.04830)
48. Dagli R, Shaikh AM, Mahdi H, Nanivadekar S. Job descriptions keyword extraction using attention based deep learning models with BERT. In: *3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*; 2021, pp. 1–6.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.