# An Ontology-based Possibilistic Framework for Extracting Relevant Terms from Job Advertisements

Albeiro Espinal[1,2][a], Yannnis Haralambous[1][b], Dominique Bedart[2] and John Puentes[1]

[1]IMT Atlantique, Lab-STICC, CNRS UMR 6285, Brest, France

[2]DSI Global Services, Le Plessis Robinson, France

Abstract: In a traditional recruitment process, large amounts of resumes and job postings are often handled manually, which is very time-consuming. Existing machine learning techniques for automatic resume ranking lack accuracy in accessing relevant information in job offers, which is crucially needed in order to ensure the pertinence of resumes. We present a context-driven possibilistic framework for extracting such information from job postings, in the form of *relevant terms*. In our process, after considering the recruiters' specific organizational context, we analyze their term relevance evaluation strategies in job advertisements. By interviewing a group of recruiters and analyzing their behavior, we have derived a first set of *textual relevance markers*. Existing term-extraction methods from the literature were also applied to extract such textual relevance markers. We have evaluated all markers using cognitive uncertainty measures and we have integrated them into an ontology-based Belief-Desire-Intention architecture. Doing this, we have improved the F1 score and recall measures of existing state-of-the-art term extraction approaches by 20% and 29% respectively. Besides, our framework is open-ended: it is possible to add new textual markers at any time as nodes of a fuzzy decision tree, the calculation of which depends on the context and domain of job offers.

1 INTRODUCTION

In a recruitment process, recruiters interact with two essential documents: Curricula Vitae (CV) and Job Advertisements (JA). The former are documents sent by applicants to communicate their qualifications, while the latter are often written by recruiters to define the particular requirements of a new job opening (Cabrera-Diego et al., 2019). In general, recruiters are the main actors involved in managing these documents (Breaugh, 2013). They consider the pertinence of each applicant's CV and the requirements of the JA. Based on this comparison, they select the most relevant applicants to be contacted initially. This procedure is known as the "screening phase" (Cabrera-Diego et al., 2019).

A screening phase becomes significantly complex and time-consuming when recruiters need to evaluate a large number of CVs and JAs (Cabrera-Diego et al., 2019). This phenomenon is indeed globally experienced by multiple recruitment offices (Zhao et al., 2021). To reduce such workload, machine learning

methods have been proposed for automatically ranking applicants' CVs depending on the content of a JA. However, these methods do not achieve the expected identification of relevant information expressed in the JA. As a result, CVs are often ranked according to background information, unrelated to the given JA.

To understand and overcome this limitation, we studied the way recruiters evaluate information relevance in JAs. The study of recruiters' expertise is essential, since they develop strategies to identify the most relevant requirements of job positions.

As a first step in the modeling of the recruiter's cognitive process we chose to start with a very basic linguistic stratum, namely the one of *terms* (Frantzi et al., 2002; Cram and Daille, 2016). Therefore we identify "relevant terms" in the JA, which the recruiter expects to find in the CV. Of course, chances are the terms used in the CV may be different, but they interact with the relevant ones through a semantic network, therefore an ontology is necessary to access a network of similarities between terms.

To identify relevant terms we define the notion of *textual markers*, that is: text features on various linguistic strata that can be used for relevance evaluation. The main contribution of our study is to provide

[a] https://orcid.org/0000-0002-1907-3424

[b] https://orcid.org/0000-0003-1443-6115

163

an ontology-based possibilistic framework to define and evaluate textual markers associated to information relevance in JAs. This framework is sufficiently large to encompass the various recruiters' strategies and viewpoints. We also show how the organizational context, as represented by the recruiters, can be exploited for automating and improving the extraction of relevant information from JAs.

This article is organized as follows. Section 2 describes the state of the art. We summarize some key principles of the proposed framework in Section 3. The set of JA textual markers derived from the recruiters' perceptions and strategies is presented in Section 4. Experimental results and discussion, along with conclusions and perspectives are presented in Sections 5, 6 and 7, respectively.

2 STATE OF THE ART

In general, the automated ranking of resumes in relation to the content of a JA consists of three main steps, namely, pre-processing, representation, and ranking. Pre-processing includes text extraction from documents in digital format, like .pdf, .doc, or .txt files, followed by segmentation and semantic annotation of sections (Cabrera-Diego et al., 2019). Moreover, noise reduction is necessary to identify and remove stop words, as well as special symbols (Roy et al., 2020). Thereafter, by means of ontology-based (Çelik, 2016), document-embedding-based (Zhu et al., 2018), or word-count-based techniques (Cabrera-Diego et al., 2019), representations of CVs and JAs are defined, their contents are compared applying similarity metrics (Guo et al., 2016), and a ranking of resumes is obtained.

Various approaches have been proposed to determine ranking. For instance, recruiters' feedback has been combined with the evaluation of a relevance algorithm (Kessler et al., 2012) (Cabrera-Diego et al., 2019). Other techniques, centered on word embedding, used ranking methods based on neural networks such as CNNs, LSTMs, RNNs, and DNNs (Deng et al., 2018). Attention-based models were also defined to rank resumes with respect to JAs (Zhao et al., 2021) (Wang et al., 2021). Furthermore, joint embeddings were specified, in order to determine a common semantic space and compute cosine similarities of CVs and JAs (Zhu et al., 2018).

The process of ranking applicants' resumes encompasses a fundamental step, rarely studied in the literature, namely the task of explicitly identifying the most relevant information expressed in a JA, in order to rank the corresponding resumes. This question of

explicit information extraction from JA has been approached under the perspective of document indexing, at the document representation stage. Making use of an *n*-gram and a graph representation of each document, the RAKE (Rapid Automatic Keyword Extraction) (Rose et al., 2010) algorithm computes terms' frequency and degree, to identify the most relevant document terms. The YAKE! (Yet Another Keyword Extraction) algorithm (Campos et al., 2018) uses textual and statistical features to detect the most pertinent information expressed in documents. A parallel combination of textual and graph-based statistical features is used by the FRAKE (Fusional Real-Time Keyword Extraction) algorithm (Zehtab-Salmasi et al., 2021) for the same purpose. Finally, (Dagli et al., 2021) studied the viability of topical and BERT (Bidirectional Encoder Representations from Transformers) models for identifying the most relevant information in JAs.

Identifying the most relevant information in JAs is essential to ensure the pertinence of an automatic resumes ranking process. Yet, from the recruiter's point of view, relevant information in a JA is not equivalent to extracted information for document indexing. Additionally, existing studies in the field of document indexing do not consider the organizational context of JA processing. This point is critical since the under-representation of the organizational context limits significantly over time the application of machine learning methods (Martin Jr. et al., 2020).

In this study, we aim to improve the task of identifying automatically the most relevant information in JAs, by giving a central role to the representation of the organizational context. This context is a strong component of the recruiters expertise, since their behavior is considerably influenced by the dynamics and needs of each enterprise (Breaugh, 2013). The ontology-based possibilistic framework that we propose intends to leverage organizational context integration, through the analysis of recruiters' strategies and viewpoints, for deriving context-driven textual markers of information relevance in JAs.

3 FRAMEWORK DESCRIPTION

Our framework is structured in three axes: The first axis analyses what is relevant for recruiters in JAs. We specify the organizational context of the documents and derive textual markers of information relevance. These markers are represented by an ontology. The second axis evaluates the pertinence of the markers from a recruiters' perspective. To this end, a fuzzy decision tree is constructed to identify unde-

tected relationships between markers. These relationships become new markers associated to information relevance in a JA. As for the third axis, it determines the best textual markers in a possibilistic agent architecture.

3.1 Definitions

Recruiters and job applicants write JAs and CVs respectively, with a single goal: to express the set of professional skills that they need or offer (Zhao et al., 2021). A professional skill can be defined as the mobilization of knowledge, know-how, and interpersonal abilities to achieve the goals of a professional activity (Le Vrang et al., 2014). A professional activity can be defined as a series of related tasks, requiring special knowledge and skills (Kiselyova et al., 2021).

A recruitment process is a sequence of steps undertaken to identify, engage, and retain the best applicant for a specific job position. With the help of field experts, recruiters identify the set of skills for a new job opening. The main needs are then expressed in a JA. Applicants respond to the JA by sending their resumes, which are manually evaluated by recruiters in order to determine whether or not the content of the resumes meets the JA requirements (Cabrera-Diego et al., 2019).

The screening phase demands the recruiter to gain knowledge of the organizational context in order to achieve successful results (Roy et al., 2020). Following this phase, the most suitable profiles are contacted on a personal basis in order to gain insight into their experience and skills. Finally, recruiters conduct face-to-face interviews with the top-ranked candidates.

3.2 Organizational Context and Recruiters' Strategies

In this section, we provide further details about the protocol that we propose for extracting and constructing textual features from recruiters' strategies and viewpoints. In general, a recruitment process takes place in the organizational context of an enterprise. We represent the organizational context of our domain, adapting an existing definition of societal context (Martin Jr. et al., 2020). Accordingly, the organizational context surrounding JAs is represented by identifying the main social actors, artifacts, actions on artifacts, and enterprise processes, associated with this type of document. Those aspects of the organizational context can significantly influence the lifecycle of CVs and JAs in the hiring processes. As a consequence, the representation of the organizational context is a fundamental preliminary step. Our ap-

proach makes use of the UNC-method, which was defined for representing organizational contexts (Zapata Jaramillo and Arango Isaza, 2009), as:

1. The organizational context representation is constructed together with recruiters from preliminary interviews (Zapata Jaramillo and Arango Isaza, 2009). Pre-conceptual schema and domain models are applied to identify and model the main contextual concepts and their relationships. From these representations, a mother-ontology is directly derived. This ontology covers the main concepts associated with JAs and the particular organizational context where these documents are processed.

2. Based on these diagrams, recruiters goals related to the life-cycle of JAs in the recruitment processes, are identified and represented in a hierarchical structure.

3. Process diagrams (Zapata Jaramillo and Arango Isaza, 2009) represent models of enterprise processes related to JAs.

4. Using a fishbone chart, the relationships between organizational problems related to JAs and their causes are also represented.

5. As suggested by (Zapata Jaramillo and Arango Isaza, 2009), all the previous diagrams are combined in a Process Explanatory Table in order to unify the representation of the organizational context.

6. From this representation and the derived ontology, the analysis of recruiters perceptions and strategies about information relevance in JAs is carried out.

7. Recruiters proceed to annotate relevant information in JA documents, handled in different recruitment processes. As recruiters read and annotate the most relevant information in JAs, each annotation is observed and described in detail, applying the controlled language proposed by (Zapata Jaramillo and Arango Isaza, 2009).

8. Our approach identifies and classifies common and transverse recruiters' behaviors when annotating documents into two categories: explicit ones (such as selecting a term) and implicit ones (such as refraining from selecting of a term). Transverse annotation behaviors are described by semantic rules. The mother-ontology conceived in step (1) is used to model the concepts and relations required by each rule. These semantic rules are textual markers of information relevance in JAs.

9. Cognitive uncertainty associated to derived textual markers is estimated by computing an ambiguity measure, as defined in (Yuan and Shaw, 1995). In Section 4, we formalize the process of evaluating textual markers.

10. Relevance markers with an ambiguity level less than or equal to an ambiguity threshold are selected and implemented in an ontology-based possibilistic agent architecture to simulate recruiters' behaviors. Relations between markers are analysed by calculating a fuzzy decision tree (Yuan and Shaw, 1995). In this tree, each marker is considered as a decision node. The least ambiguous relations derived from this tree are identified as new relevance markers in the agent's architecture, in order to improve its deliberation process of information relevance in JAs.

3.3 Ontological Resources

As a fundamental part of our framework, a mother-ontology contains the main concepts and relations inherent to the specific context of JAs. This ontology is complemented by integrating existing ontologies related to the specific organizational context. In our case, we integrated the following sources of knowledge. The internal ontology of DSI Group, a French consulting firm. It contains over 36,000 professional skills, lexically represented in French language. By complementing it with international ontologies like ESCO (Le Vrang et al., 2014), we also reconstructed ontologies that underpin professional skills frameworks such as O*NET[1], CIGREF[2], and ROME[3], based on text-to-RDF-triple transformations (Somodevilla García et al., 2018).

Additionally, we built an ontology based on 14,000 anonymous CVs and 2,000s JAs, in a semi-automatic manner, by applying ontology learning techniques (Alfonso-Hermelo et al., 2019). Specifically, we extracted and processed the professional skills' sections of these documents. Then, we integrated all ontologies to the mother-ontology by using a hybrid approach based on transformer models (BERT), a terminological variation analysis (Cram and Daille, 2016), and ontology quality metrics (Mc Gurk et al., 2017). In Figure. 1, we present an upper-view of the resulting ontology.

---

[1]https://www.onetonline.org/

[2]https://www.cigref.fr/

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

3.4 Beliefs, Desires and Intentions Architecture of the Agent

According to multiple empirical studies, "...humans' semantic inferences are uncertain and context-sensitive..." (Pavlick and Kwiatkowski, 2019). The same principle can be applied to the process of extracting relevant terms from JAs. In fact, recruiters *semantically infer* relevant terms by reading the JA's text, "...[assuming] common human understanding of language [and] common background knowledge..." (Pavlick and Kwiatkowski, 2019). Consequently, we propose a possibilistic architecture to incorporate human semantic reasoning and its associated uncertainty for enhancing the automatic extraction of relevant terms.

Specifically, this architecture (Da Costa Pereira and Tettamanzi, 2010) is based on what has been defined as the agent's beliefs, desires, and intentions paradigm, linked to modeling of variables applying possibility (Figure. 2).

The module of beliefs is composed by two sub-modules. The first one consists of a distribution of trust degrees τ, which represents a trust level that the agent assigns to its sources ϕ. In our context, there are three categories of sources: JAs, ontological resources, and textual markers of information relevance. We define trust degrees through quality metrics, in order to detect possible inconsistencies, as malformed JA texts, ontology conformity issues (Mc Gurk et al., 2017), or ambiguous textual relevance markers. Each trust degree is associated to a tolerance threshold β_j. A source with a trust degree τ_j below a given tolerance threshold β_j will be rejected by the agent.

The second sub-module makes use of a possibility distribution π, which represents the agent's beliefs $\mathcal{B}$. Each term is characterized by a possibility distribution formed by the levels of relevance (possibility degrees) provided by each marker. This possibility distribution induces a measure of possibility Π (Da Costa Pereira and Tettamanzi, 2010). It shows to what extent a term is pertinent. Then, the possibility measure is associated to its dual necessity measure N (Da Costa Pereira and Tettamanzi, 2010). The latter can be interpreted as to what extent it is impossible not to conclude that a term is relevant. Note that possibility degrees of beliefs are limited to the trust degree of their respective information about knowledge sources. A belief modification operator $*$ is defined for this purpose (Da Costa Pereira and Tettamanzi, 2010).

The second agent's component is the module of desires, represented by an utility distribution (for-
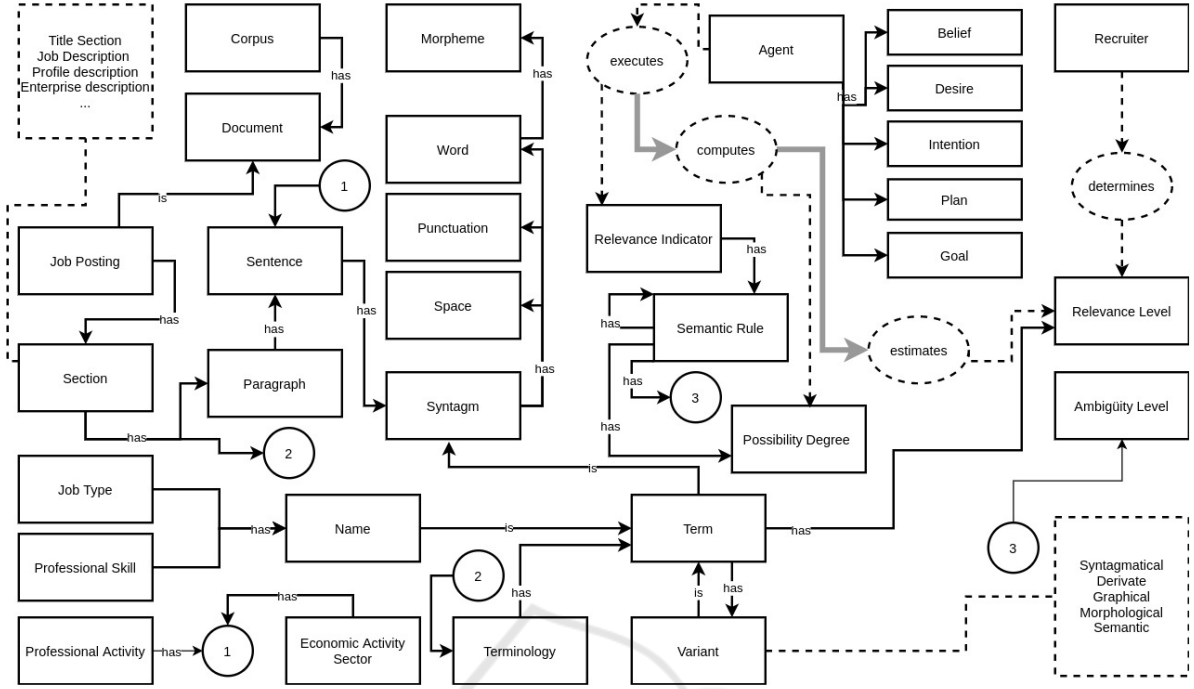
Figure 1: Upper-view of the mother-ontology created from the context representation.
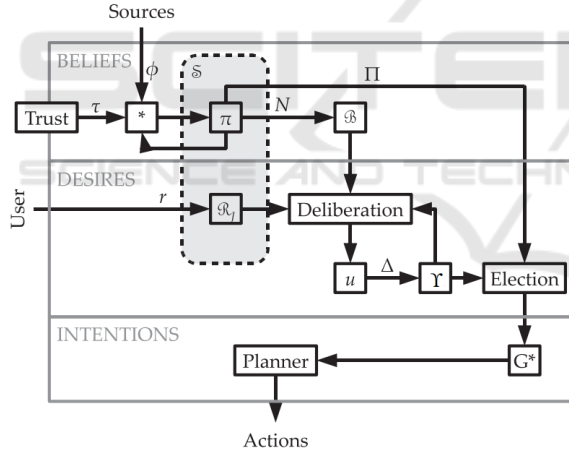

Figure 2: Basic architecture of the dynamic BDI agent (Da Costa Pereira and Tettamanzi, 2010).

mally a possibility distribution), which 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 R_J are logical conditions expressing dependence relationships between beliefs and desires (Da Costa Pereira and Tettamanzi, 2010). The evaluation of desire-generation rules allows the agent to estimate to what extent Δ it is justified to make use of each marker. Textual markers associated to positive degrees become justified desires Υ.

The third agent's component, namely intentions, consists of actions that the agent decides to execute based on its current beliefs, justified desires, and goals. A plan outlines the sequence of actions to be taken. In our context, these actions refer to natural language processing methods that are necessary to automatically extract the relevant terms from JAs.

4 TEXTUAL MARKERS

The algorithm YAKE! (Campos et al., 2018) already uses textual markers, but defined in an informal way. In this section we will formally define the notion of textual marker in a way that it is compatible with YAKE! textual markers, as well as with textual markers actually used by recruiters.

4.1 Preliminary Definitions

According to (Cram and Daille, 2016), a term is "a functional class of lexical units used in discourse". In our context, terms are identified based on their *termhood* (Frantzi et al., 2002), measured by using the *weirdness ratio* (Cram and Daille, 2016).

Terms are extracted by applying the most frequent morpho-syntactic patterns to multiple specialized corpora (Cram and Daille, 2016). Nearly all of them correspond to nominal phrases.

Let d_i be a JA belonging to a corpus C and $T_{d_i} =$

$\{t_1, t_2, \ldots, t_n\}$ the set of terms of d_i.

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

Let $A_{d_i} = \{a_1, a_2, \ldots, a_m\}$ 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 $E_{d_i} = \{e_1, e_2, \ldots, e_p\}$ be a set of qualifying adjectives and nouns that are linked to a subset of terms in T_{d_i} by syntax dependencies.

Let $O = \{o_1, o_2, \ldots, o_s\}$ be a set of ontologies (as the one presented in Section 3). Let $c_{o_s} = \{c_{s,1}, c_{s,2}, \ldots, c_{s,k}\}$ be the set of concepts of ontology o_s and $T_{c_j} = \{t_{j,1}, t_{j,2}, \ldots, t_{j,l}\}$ the set of terms lexically representing concept c_j in a given language.

The mother-ontology O contains, among other concepts, the structure of a textual document: sections, paragraphs, sentences, syntagms, terms, words, morphemes, etc.

4.2 Definitions of Markers

In this section, textual markers #1 to #10 have been obtained by monitoring the behavior of expert recruiters and by interacting with them. Textual markers #11 to #16 correspond to those of the YAKE! term extraction approach.

4.2.1 Textual Marker #1: Presence of Professional Skills or Job Types in Title Sections

"In the event that a term in the title matches one of the terms used to represent professional skills or job types, then it can potentially be considered as relevant."

Let $a_1 \in A_{d_i}$ be the title section of d_i. Let $t_{a1} = \{t_1, t_2, \ldots, t_u\}$ 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:

$$\forall t_k \exists c_j [c_j \in o_s \wedge t_k \in T_{c_j} \wedge t_k \in t_{a1}] \rightarrow t_k \in R_{d_i}, \quad (1)$$

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

4.2.2 Textual Marker #2: Terms Representing Professional Skills in a Job Description Section or Profile Description Section

In general, a term used to represent a professional skill in a job description section or in a profile description section is more likely to be chosen as a relevant term. Let s_2 and s_3 be the sets of terms in the job description section and the profile description section, respectively. Set $t_k \in T_{d_i}$. Let T_{c_j} be the set of terms

representing a professional skill concept c_j in the ontology o_s. We request that:

$$\forall t_k \exists c_j ((t_k \in s_2 \vee t_k \in s_3) \wedge t_k \in T_{c_j}) \rightarrow t_k \in R_{d_i}, \quad (2)$$

with a possibility degree $\alpha_{t_{k,2}} \in [0,1]$.

4.2.3 Textual Marker #3: Relevance of Job Posting Sections

"Recruiters will choose with a higher possibility degree terms used in the title, job description, and profile description sections, rather than terms belonging to other sections (company description, contract details, etc)."

This marker is not redundant with markers #1 and #2, since we do not request terms to be professional skills. Let $S = s_1 \cup s_2 \cup s_3 \subseteq T_{d_i}$, where: s_1 is the set of terms of the title section; s_2 is the set of terms of the job description section; and s_3 is the set of terms of the profile description section.

Let $t_m \in T_{d_i} \cap S$. Then, we request that:

$$\forall t_m \forall t_n (t_m \in T_{d_i} \wedge t_n \notin S) \rightarrow (P(t_m \in R_{d_i}) > P(t_n \in R_{d_i})), \quad (3)$$

with a possibility degree $\alpha_{t_{k,3}} \in [0,1]$. $P(t_* \in R_{d_i})$ denotes the possibility of t_* being chosen as a relevant term.

4.2.4 Textual Marker #4: Terms Dependent on Pertinence Expressions

"A term that bears a syntax dependency with a syntagm of the JA is more likely to be chosen as a relevant term."

- Let $t_k \in T_{d_i} \cap T_{c_j}$ for some c_j.
- We define a "pertinent expression" e_m as a syntagm used by the recruiter who wrote the JA (i.e., you *master C#*, *good knowledge* of cloud computing). Let us suppose that e_m is syntactically dependent with t_i. More specifically, let t_k be a qualifying adjective or a noun modifier directly dependent with e_m. Then, we request that:

$$\forall t_k \exists e_m (t_k \in T_{d_i} \wedge e_m \in E_{d_i} \wedge \text{is\_dependent}(t_k, e_m)) \rightarrow t_k \in R_{d_i}, \quad (4)$$

with a possibility degree $\alpha_{t_{k,4}} \in [0,1]$.

4.2.5 Textual Marker #5: Terms Used in Traces of Professional Activities Descriptions

"It will be more likely for a term representing a professional concept to be considered as relevant when a job description explicitly describes an interaction with this concept."

We define a trace of a professional activity description as a sentence in a job posting, describing a worker's action on an object. Be $b_j \in d_i$ a trace of a professional activity description represented by the

set of terms T_{b_j}. We request that b_j contains at least one verb and one dependent object. We assert that the terms t_k representing these objects will have a greater possibility degree of being chosen as relevant. Therefore:

$$\forall t_k(t_k \in T_{b_j} \land \text{is\_object}(t_k, b_j)) \rightarrow t_k \in R_{d_i}, \quad (5)$$

with a possibility degree $\alpha_{t_{k,5}} \in [0,1]$.

4.2.6 Textual Marker #6: Terms Representing High Risk Professional Skills/Activities

This marker intends to provide more relevance to terms representing professional skills or activities for which an employee's mistake *can significantly degrade the company's economic activity*. Value 0 indicates that a potential mistake will not have a substantial impact on the economic activity, while value 1 indicates that an error on this skill or activity will have a strong impact.

Depending upon the unique context of each enterprise, this textual marker allows the agent to adapt its own behavior to meet the specific needs of each job opening.

Let M be an ontology that contains the set of professional skills and activities of a given company. M contains a set of concepts $c_M = \{c_{M,1}, c_{M,2}, \ldots, c_{M,k}\}$. The recruiter manually assigns a risk level $\varepsilon_{c_{M,k}} \in [0,1]$ to each professional skill or activity.

Let s_j be a term in a job posting d_i representing a professional skill or activity in M. Among the concepts associated to s_j, let $c_{M,l}$ be the concept with the maximal risk level. If this risk level is greater than a threshold $\beta_{c_{M,l}}$, then s_j is chosen as a relevant term and:

$$\forall s_j \exists c_{M,l}(s_j \in T_{d_i} \land c_{M,l} \in M \land s_j \in T_{c_{M,l}}$$
$$\land \text{is\_greater\_than}(\varepsilon_{c_{M,l}}, \beta_{c_{M,l}}) \rightarrow s_j \in R_{d_i}, \quad (6)$$

with possibility degree $\alpha_{s_{j,6}} \in [0,1]$

4.2.7 Textual Marker #7: Actions Expressed in Management JAs

It is also important for recruiters to identify the type of actions required by management JAs. For instance, some jobs focus on team management, while others include accountability activities or even development tasks. This is because management jobs can be quite diverse.

Be d_i a management job posting. The agent detects management JAs based on a Latent Dirichlet Allocation model, trained on 14,000 curricula vitae. Let t_k be a verbal term of d_i. If t_k belongs to the trace of a professional activity f_j and corresponds to the head of its syntactic tree, then it potentially is a relevant term.

We define this marker as:

$$\forall t_k \exists f_j(f_j \in d_i \land t_k \in f_j \land \text{is\_management}(d_i) \land$$
$$\text{is\_verb}(t_k) \land \text{is\_head\_of}(t_k, f_j)) \rightarrow t_k \in R_{d_i}, \quad (7)$$

with a possibility degree $\alpha_{t_{k,7}} \in [0,1]$.

4.2.8 Textual Marker #8: BERT Semantic Similarity of Professional Skills

"Specific terms used to represent professional skills that are semantically close (in the sense of BERT) to already discovered relevant terms, will be considered relevant."

Let $t_1 \in R_{d_i}$ and $t_2 \in T_{d_i}$. We define the specificity function (Cram and Daille, 2016) of a term $f(t)$ as its relative frequency in a specific corpus C_s, divided by its relative frequency in a multi-language corpus C_L.

Furthermore, we define $g(t_1, t_2)$ as the BERT semantic similarity between two terms. We used a SBERT model pre-trained on the Wikipedia corpus (Reimers and Gurevych, 2019) to derive semantically meaning from complex terms. This model was fine-tuned on the professional skill standards CIGREF, e-CF, c2i, and ROME, being formalized as follows:

$$\forall t_1 \forall t_2(t_1 \in R_{d_i} \land g(t_1, t_2) > 0) \rightarrow t_2 \in R_{d_i}, \quad (8)$$

with a possibility degree defined by the normalized equation :

$$\alpha_{t_{2,8}} = \|(1 - \alpha_{t_1}) * g(t_1, t_2) * f(t_2))\|. \quad (9)$$

4.2.9 Textual Marker #9: Relevance of the Economic Activity Sector

"Terms referring to an economic activity sector required by a job posting (e.g., finance, banks, aeronautics, etc.) will be chosen as potentially relevant."

This implies that:

$$\forall t_k(t_k \in T_{d_i} \land \text{is\_sector\_requirement}(t_k)) \rightarrow t_k \in R_{d_i}, \quad (10)$$

with a possibility degree $\alpha_{t_{k,9}} \in [0,1]$.

Economic activity sectors are identified by aligning job posting terms and economic activity concept labels, provided by ESCO, O*NET, ROME, and CIGREF standards.

4.2.10 Textual Marker #10: Professional Skill Prerequisites

Let there be a *prerequisite relation* between two professional skills c_1 and c_2 as given in an ontology o_i. Relations of this type can be obtained from ontologies such as ESCO. If c_2 is a prerequisite of c_1 and c_1 is considered as relevant (under a certain possibility degree) then c_2 inherits the possibility degree of c_1.

$$\forall t_1 \forall t_2 \exists c_1 \exists c_2(c_1 \in o_i \land c_2 \in o_i \land t_1 \in T_{c_1} \land t_2 \in T_{c_2} \land$$
$$\text{is\_prerequisite}(c_1, c_2) \land t_1 \in R_{d_i}) \rightarrow t_2 \in R_{d_i}, \quad (11)$$

with a possibility degree $\alpha_{t_{k,10}} \in [0,1]$. This possibility degree is equal to the possibility degree of $t_1 \in R_{d_i}$.

4.2.11 Textual Marker #11: YAKE! Casing

Upper-cased terms tend to be more relevant. In our context, this YAKE! marker is related to the recruiters' behavior, as they tend to capitalize terms pertaining to professional skills:

$$\forall t_k(t_k \in T_{d_i} \wedge \text{is\_upper\_cased}(t_k)) \to t_k \in R_{d_i} \quad (12)$$

We define the possibility degree of this rule based on the normalized YAKE! equation:

$$\alpha_{t_{k,11}}(t_k) = \left\| \frac{\max(\text{TF}(U(t_k)), \text{TF}(A(t_k)))}{\ln(\text{TF}(t_k))} \right\|, \quad (13)$$

where $\text{TF}(U(t_k))$ is the number of times that t_k appears uppercased, $\text{TF}(A(t_k))$ is the number of occurrences of t_k as an acronym (for details see (Campos et al., 2018)) and $\text{TF}(t_k)$ is the term frequency.

4.2.12 Textual Marker #12: YAKE! Term Position

The hypothesis of this marker is that terms appearing at the beginning of the document tend to be more relevant.

$$\forall t_k(t_k \in T_{d_i} \wedge \text{is\_position\_marker\_activated}(t_k)$$
$$\to t_k \in R_{d_i}, \quad (14)$$

with a possibility degree given by the following normalized YAKE! equation:

$$\alpha_{t_{12}}(t_k) = \| \ln(\ln(3 + \text{Median}(\text{Sent}(t_k)))) \|, \quad (15)$$

where $\text{Sent}(t_k)$ is the set of positions of the sentences containing t_k.

4.2.13 Textual Marker #13: YAKE! Term Frequency Normalization

The most commonly used terms are more relevant. Hence:

$$\forall t_k(t_k \in T_{d_i} \wedge \text{is\_frequency\_marker\_activated}(t_k)$$
$$\to t_k \in R_{d_i}, \quad (16)$$

with a possibility degree given by the following normalized equation proposed by YAKE!:

$$\alpha_{t_{k,13}}(t_k) = \left\| \frac{\text{TF}(t_k)}{\text{MeanTF} + \sigma} \right\|, \quad (17)$$

where $\text{TF}(t_k)$ is the number of occurrences of t_k, which is balanced by the mean and standard deviation of frequency.

4.2.14 Textual Marker #14: YAKE! Term Relatedness to Context

This YAKE! marker is based on the following hypothesis: "The higher the number of different terms that

co-occur with a candidate term t on both sides, the least significant term t will be":

$$\forall t_k(t_k \in T_{d_i} \wedge \text{is\_relatednes\_activated}(t_k)) \to t_k \in R_{d_i}, \quad (18)$$

with a possibility degree obtained from the normalized YAKE! equation:

$$\alpha_{t_{k,14}} = \left\| 1 + (DL + DR\cdots) * \frac{\text{TF}(t_k)}{\max \text{TF}} \right\|, \quad (19)$$

where

$$DL[DR] = \frac{|A_{t,w}|}{\sum\limits_{k \in A_{t,w}} CoOccur_{t,k}}. \quad (20)$$

$|A_{t,w}|$ corresponds to the number of different terms in a window of size w and TF is the term frequency.

4.2.15 Textual Marker #15: YAKE! Different Sentences

"A term will be more relevant depending on how often it is used within different sentences," represented as:

$$\forall t_k(t_k \in T_{d_i} \wedge \text{is\_sentences\_marker\_activated}(t_k))$$
$$\to t_k \in R_{d_i}, \quad (21)$$

with a possibility degree obtained from the normalized equation:

$$\alpha_{t_{k,15}} = \left\| \frac{SF(t_k)}{\#\text{Sentences}} \right\|, \quad (22)$$

where $SF(t_k)$ is the number of sentences containing t_k and #Sentences is the total number of sentences of d_i.

4.2.16 Textual Marker #16: YAKE! Overall Score

We include the global relevance score proposed by YAKE! which is based on markers #11, #12, #13, #14, and #15 (Campos et al., 2018). Let $t_k \in d_i$. A term is considered as "partially relevant" if it's predicted as such by the overall score:

$$\forall t_k(t_k \in T_{d_i} \wedge \text{is\_predicted\_by\_yake}(t_k)) \to t_k \in R_{d_i}, \quad (23)$$

with a possibility degree $\alpha_{t_{k,16}} \in [0,1]$.

4.3 Evaluation of Relevance Markers

According to (Yuan and Shaw, 1995), "a classification task can be viewed as a rational action that a decision maker will take in consistency with his or her knowledge". In our case, the recruiter's task of annotating relevant terms is associated with uncertainties, which have been measured in the literature through ambiguity measures (Yuan and Shaw, 1995). Based on ambiguity estimation, we propose a protocol for evaluating the pertinence of relevance markers.

4.3.1 Protocol Formalization

Let $U = \{t\}$ be the set of terms of a given JA. Let C be a fuzzy set representing the recruiter perceived levels of term relevance in a JA. C is represented by the membership function μ_C, which maps recruiters' annotations into C. C is constituted by two fuzzy subsets: the subset C_1 of relevant term levels and the subset C_2 of non relevant term levels. Both fuzzy categories are represented by triangular functions. We will call R (resp. R_1, R_2) the subset of C (resp. C_1, C_2) obtained from the fuzzification of recruiters annotations on the JA.

Each JA term t can be described by a set of relevance markers I_k, obtained from recruiters strategies and viewpoints. We denote $A(t) = \{I_1, I_2, \ldots, I_k\}$.

Each relevance marker provides a degree of possibility for a term to be selected as relevant. The next step involves fuzzifying these possibility degrees by applying a membership function μ_{I_k}. Although μ_{I_k} was built in an equivalent manner as μ_C, the specific codomain of each marker I_k was taken into account. We interpret the result of this fuzzification process as an evidence E_k.

After fuzzifying I_k and the recruiters' annotations, we evaluate the possibility of describing the fuzzified recruiters' annotations R based on the evidence E_k obtained from I_k. More specifically, we evaluate the ambiguity level of the following rule: If E_k Then R. To do it, a subsethood measure between the evidence E_k and the experts' classification R has been proposed (Yuan and Shaw, 1995). In our context, as a substitute for subsethood, we use the Hamming distance between fuzzy sets to reflect more explicitly to what extent recruiter's knowledge R approximates E_k:

$$S(E_k, R) = \sum_{t \in U} \left| \mu_{I_k}(t) - \mu_R(t) \right|. \quad (24)$$

The Hamming distance measures the degree to which a term is relevant or not, based on the available evidence E_k. Applying this distance, we define the possibility π of classifying a term as relevant (R_1) or non relevant (R_2), in relation to recruiter's strategies and viewpoints as:

$$\pi(R_i \mid E_k) = \frac{S(E_k, R_i)}{\max(S(E_k, R_1), S(E_k, R_2))}. \quad (25)$$

Possibility is intrinsically related to the concept of ambiguity (Yuan and Shaw, 1995). Based on E_k, the more likely we are to deduce that a term is relevant or not, the less ambiguity there is. From $\pi(R \mid E_k)$, we estimate the ambiguity level associated to marker I_k, or equivalently, to the evidence E_k as:

$$G(E_k) = g(\pi(R|E_k)) = \sum_{i=1}^{n} (\pi_i^* - \pi_{i+1}^*) \ln(i), \quad (26)$$

where $\pi^* = \{\pi_1^*, \pi_2^*, \ldots, \pi_n^*\}$ is the possibility distribution $\pi(R \mid E_k)$ permuted and sorted so that $\pi_i^* \geq \pi_{i+1}^*$ for $i \in \{1, \ldots, n\}$ and $\pi_{n+1}^* = 0$. In our case, $n = 2$ as we evaluate ambiguity when deciding whether a term is relevant (R_1) or not (R_2) based on I_k.

This ambiguity function G indicates to what degree it can be inferred that a term is pertinent or not, depending on the relevance marker I_k. A value of 0 means that there is no ambiguity and a value of $\ln(n)$ represents the maximal ambiguity level (Yuan and Shaw, 1995).

Since $A = \{I_1, I_2, \ldots, I_k\}$ is the set of relevance markers, based on the least ambiguous markers with an ambiguity level below or equal to a threshold σ, we train a fuzzy decision tree (Yuan and Shaw, 1995). We replace the classical information entropy measure with the previously presented ambiguity metric. As proposed by (Yuan and Shaw, 1995), the path of branches from the roots to the leaves is converted to simplified logical rules. These new rules represent logical relations between markers associated to information relevance. The least ambiguous ones are added as new sources of information in the agent architecture. This allows us to enrich the deliberation of the agent. The process of building the fuzzy decision tree is an important step in evaluating the pertinence of markers. A fuzzy decision tree can be used for discovering unseen relationships between them, and also for automating the classification task of relevant terms' prediction.

5 EXPERIMENTAL RESULTS

We evaluated the proposed framework in the setting of the recruitment office of DSI Group. We interviewed four recruiters, which we refer to in this study as A, B, C, and D. These recruiters spent multiple days or even weeks leading recruitment processes and managing their associated JAs, exchanging on a daily basis with management and technical experts. In other words, they gained an in-depth understanding of the most fundamental needs of their JA by acquiring high levels of contextual knowledge.

Based on preliminary interviews with these recruiters, the organizational context surrounding the JAs was represented. Then, we interviewed an expert recruiter (recruiter A) in order to identify which job requirements were most relevant to the managed JAs. During this interview, we asked the recruiter the following question about five JAs: what are the requirements expressed in this JA that you would not accept a candidate without?

According to personal viewpoints and knowledge,

Table 1: Experimental results. Precision, recall, and F1-score levels of each method on 20 job advertisements applying RAKE (Rose et al., 2010), FRAKE (Zehtab-Salmasi et al., 2021), BERT topics approach (Dagli et al., 2021), YAKE! (Campos et al., 2018), and our Agent.

Metric/Method	RAKE	FRAKE	BERT	YAKE!	Our Agent
Recall@N, Precision@N F1-Score@N[4]	0.02	0.09	0.17	0.10	**0.38**
Recall@2N	0.08	0.17	0.20	0.27	**0.56**
Precision@2N	0.04	0.08	0.10	0.13	**0.28**
F1-Score@2N	0.05	0.11	0.14	0.18	**0.38**

this recruiter marked a set of terms that represented the most essential requirements of each JA. The majority of annotated terms corresponded to professional skills. A small number of terms related to locations or key aspects of previous responsibilities. The recruiter annotated a minimum of 4 terms and a maximum of 10 terms per JA. In this specific application case, JAs had, in average, 100 terms.

By applying our framework, we derived textual markers of information relevance associated to the recruiter strategies and viewpoints. These markers were formalized, implemented, and became the cornerstone of a BDI agent. We evaluated the validity and reproducibility of this agent on 20 recruitment processes managed by recruiters B, C and D. This process was executed as follows:

- We asked recruiters B, C and D to annotate the most relevant information expressed in the set of JAs associated to their recruitment processes.

- We compared their annotations in relation to the most relevant terms predicted by the agent derived from recruiter A's strategies and viewpoints.

- We compared the performance of the proposed agent to state of the art term extraction methods. We used precision@K, recall@K and F1-score@K metrics to asses the suitability of each model, in predicting the top N and top $2N$ most relevant terms of each annotated JA (where N represents the number of terms annotated by the recruiter on each document).

Even if theoretically our approach should be compared to other approaches specifically designed for JAs, we discarded that aim for two reasons. To the best of our knowledge, currently there are not automated and open source methods for extracting specifically relevant information from JA. Also, there are no public JA corpora annotated for the extraction of relevant information. For these reasons, we compared our approach to the performance of domain-independent state of the art term extraction methods. In their respective papers, these approaches were evaluated on annotated academic documents, news texts, and scientific articles (Campos et al., 2018). Therefore, this is the first time that they are evaluated on JAs.

Table 1 presents the results of our experiments. As indicated by the metrics, our algorithm's results are significantly better than the other four algorithms. Particularly, compared to the best results of those algorithms (most of them from YAKE!), our approach improvements vary from 15% to 29%, being 56% for Recall@2N the highest performance.

Regarding the performance of the BDI agent, three aspects of its behavior were essential for achieving and preserving a superior F1-Score at @N and @2N. Firstly, context-driven markers associated with lower levels of ambiguity, significantly influence the extraction process of the agent. For instance, markers #1, #4 and #5 have usually been sufficient per se for determining whether a term (under the conditions of each marker) is relevant or not. Secondly, context-driven markers with medium ambiguity levels tend to reduce their associated cognitive uncertainty, when used together with other markers. Thirdly, we observed that the inclusion of cognitive uncertainty measures to determine trust degrees, allowed the agent to control the relative ambiguity associated to some markers as #12.

Finally, our results reflect the complexity of extracting relevant terms from short documents such as JAs. Work on YAKE! had already evidenced this issue on corpora made up of short length documents. Various corpora were involved in their work, e.g., WWW, KDD and pak2018 (Campos et al., 2018). In our case, YAKE! has a similar behavior (18%) when compared to its reported maximal of 17.2% for these datasets of short documents. On the other hand, a maximal F1-Score@2N of 5% and 14% for RAKE (n-gram) and BERT (embeddings) respectively, could be attributed to the inadequacy of document representations to reproduce the concepts of an organizational context. In fact, the n-gram approach tends to underrepresent terms, while the embedding approach may not be aligned with the inherent knowledge of a given organizational context, even after a fine-tuning process. It also appears that the FRAKE algorithm

is limited by the short length of JAs, given that it is centered on statistical graph-based markers. According to these elements, with a maximal F1-Score of 38%, the performance of our approach illustrates that context-driven markers can complement relevance information extraction in JAs, by integrating knowledge of the specific organizational context, from a recruiter's perspective.

6 DISCUSSION

The automatic extraction of relevant terms from JAs is a complex task that has been rarely studied, despite its key role to determine the ranking of CVs during the screening phase of recruitment. We propose an approach that relies on a multisource ontology and a possibilistic framework, utilized by a dynamic BDI agent model, to analyze simple rules that define 16 textual markers. Given the current lack of specific approaches for this task, we compared our approach to the closest related techniques, i.e., automatic term extraction approaches for document indexing. Out of 16 textual markers, we defined 10 markers upon an expert recruiter's strategies and viewpoints, while 6 were integrated from the YAKE! algorithm. Results show that the defined BDI agent performs better than automatic term extraction algorithms, in extracting relevant information from JAs. Obtained performances suggest that representing the organizational context of JAs, in terms of recruiter's strategies and viewpoints, is likely to improve the identification of information relevance, beyond relevant terms.

Also, we observed that our implementation is likely to detect more coherent sets of relevant terms than other approaches. Given that the agent extracts terms following a terminological analysis (Frantzi et al., 2002; Cram and Daille, 2016), it detects complex terms often chosen as relevant by recruiters. Besides, we observed that some textual representations, such as the one proposed by RAKE (Rose et al., 2010), underrepresent complex terms, reducing predictability considerably.

An overview of results reveals that performance measures of all algorithms, except Recall@2N of our algorithm, are below the expected baseline performance. This can be explained by the fact that statistical markers on which are based the works of (Campos et al., 2018) (Rose et al., 2010) (Zehtab-Salmasi et al., 2021) are insufficient to determine appropriately term relevance in JAs. Our results show that context-based qualitative relevance markers are essential for achieving a better F1-score, given that they are independent of the JA's size. E.g., it is not unusual to find JAs

where the extent of the enterprise description is large compared to the job description. As a consequence, statistical markers tend to give a higher score to terms from the former section that aren't necessary to describe essential job requirements.

After performing combinatorial analysis of textual markers, we found that the agent gets a better F1-score by executing textual markers as follows. The first step is to generate a population of beliefs about terms relevance, using independent markers with lower ambiguity levels. For instance, marker #1 based on titles contents or marker #6 focused on financial impact of professional skills/activities. Then, dependent markers with low ambiguity levels (such as #10) can be applied in order to reinforce the current population's beliefs about terms relevance.

7 CONCLUSION AND PERSPECTIVES

Multiple analyses and resources are necessary to automatize a recruitment process. While a rich ontology is essential to cover the wide spectrum of professional qualifications, skills, and experience, the study of JAs permits to identify relevant terms intended to facilitate CV ranking. Additionally, the ambiguity of employed language must also be taken into account. Since statistical parameters do not provide optimal representations of relevant document terms in a JA, we propose to correlate human decisions with identified textual markers. This approach relies strongly on terminological analysis and hypotheses formulated after observing how recruiters work. Our results indicate that the terminological analysis is a fundamental step for improving the linguistic quality and contextual relevance of information extraction from JAs. Our framework has therefore the potential to assess the relevance of text markers within specific contexts, and to be adapted to the evolving organizational context of each company.

Further work is still required to confirm the various insights that have been identified, in particular concerning recruiters' way of working when matching JAs and CVs, as well as the relevance of other models to represent the semantic space of relevant terms. A larger corpus and more recruiters will be necessary to evaluate whether the applicability of textual relevance markers can be affected by variations in the organizational context. Compatible automatic approaches to identify relevant terms in CVs to match JAs are foreseen, integrating the recruiter's expertise. Other machine learning models will be defined looking for a more precise representation of the intrin-

sic and extrinsic relationships between the concepts constituting a specific organizational context in which JAs are processed.

REFERENCES

Alfonso-Hermelo, D., Langlais, P., and Bourg, L. (2019). Automatically Learning a Human-Resource Ontology from Professional Social-Network Data. In *Canadian AI 2019: Advances in Artificial Intelligence*, volume 11489 of *LNAI*, pages 132–145. Springer.

Breaugh, J. A. (2013). Employee Recruitment. *Annual Review of Psychology*, 64:389–416.

Cabrera-Diego, L. A., El-Béze, M., Torres-Moreno, J. M., and Durette, B. (2019). Ranking résumés automatically using only résumés: A method free of job offers. *Expert Systems with Applications*, 123:91–107.

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A. M., Nunes, C., and Jatowt, A. (2018). YAKE! Collection-Independent Automatic Keyword Extractor. In *Advances in Information Retrieval*, pages 806–810. Springer.

Çelik, D. (2016). Towards a semantic-based information extraction system for matching résumés to job openings. *Turkish Journal of Electrical Engineering and Computer Sciences*, 24(1):141–159.

Cram, D. and Daille, B. (2016). Terminology extraction with term variant detection. In *Proceedings of ACL-2016 system demonstrations*, pages 13–18.

Da Costa Pereira, C. and Tettamanzi, A. G. B. (2010). An integrated possibilistic framework for goal generation in cognitive agents. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1*, pages 1239—1246.

Dagli, R., Shaikh, A. M., Mahdi, H., and Nanivadekar, S. (2021). 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)*, pages 1–6.

Deng, Y., Lei, H., Li, X., and Lin, Y. (2018). An improved deep neural network model for job matching. In *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 106–112.

Frantzi, K. T., Ananiadou, S., and Tsujii, J. (2002). The C-value/NC-value Method of Automatic Recognition for Multi-word Terms. *Research and Advanced Technology for Digital Libraries*, 1513:585 – 604.

Guo, S., Alamudun, F., and Hammond, T. (2016). RésuMatcher: A personalized résumé-job matching system. *Expert Systems with Applications*, 60:169–182.

Kessler, R., Béchet, N., Roche, M., Torres-Moreno, J. M., and El-Bèze, M. (2012). A hybrid approach to managing job offers and candidates. *Information Processing and Management*, 48(6):1124–1135.

Kiselyova, E. I., Koroshchenko, K. R., and Robson, G. (2021). Legal Horizons Content of the Job Description: Features and Areas of Concern. *Legal Horizons*, 14(2):63–69.

Le Vrang, M., Papantoniou, A., Pauwels, E., Fannes, P., Vandensteen, D., and De Smedt, J. (2014). ESCO: Boosting job matching in Europe with semantic interoperability. *Computer*, 47(10):57–64.

Martin Jr., D., Prabhakaran, V., Kuhlberg, J., Smart, A., and Isaac, W. S. (2020). Extending the Machine Learning Abstraction Boundary: A Complex Systems Approach to Incorporate Societal Context. arXiv 2006.09663.

Mc Gurk, S., Abela, C., and Debattista, J. (2017). Towards Ontology Quality Assessment. http://ceur-ws.org/Vol-1824/ldq_paper_2.pdf.

Pavlick, E. and Kwiatkowski, T. (2019). Inherent disagreements in human textual inferences. *Transactions of the Association for Computational Linguistics*, 7:677–694.

Reimers, N. and Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 3982–3992. Association for Computational Linguistics.

Rose, S., Engel, D., Cramer, N., and Cowley, W. (2010). *Automatic Keyword Extraction from Individual Documents*, chapter 1, pages 1–20. John Wiley and Sons.

Roy, P. K., Chowdhary, S. S., and Bhatia, R. (2020). A Machine Learning approach for automation of Resume Recommendation system. *Procedia Computer Science*, 167:2318–2327.

Somodevilla García, M., Vilariño Ayala, D., Pineda, I., Somodevilla García, M., Vilariño Ayala, D., and Pineda, I. (2018). An Overview of Ontology Learning Tasks. *Computación y Sistemas*, 22(1):137–146.

Wang, X., Jiang, Z., and Peng, L. (2021). A Deep-Learning-Inspired Person-Job Matching Model Based on Sentence Vectors and Subject-Term Graphs. *Complexity*, 2021:1–11.

Yuan, Y. and Shaw, M. J. (1995). Induction of fuzzy decision trees. *Fuzzy Sets and Systems*, 69(2):125–139.

Zapata Jaramillo, C. M. and Arango Isaza, F. (2009). The UNC-method: a problem-based software development method. *Ingeniería e Investigación*, 29:69–75.

Zehtab-Salmasi, A., Feizi-Derakhshi, M.-R., and Balafar, M.-A. (2021). FRAKE: Fusional Real-time Automatic Keyword Extraction. arXiv 2104.04830.

Zhao, J., Wang, J., Sigdel, M., Zhang, B., Hoang, P., Liu, M., and Korayem, M. (2021). Embedding-based Recommender System for Job to Candidate Matching on Scale. arXiv 2107.00221v1.

Zhu, C., Zhu, H., Xie, F., Ding, P., Xiong, H., Ma, C., and Li, P. (2018). Person-Job Fit: Adapting the Right Talent for the Right Job with Joint Representation Learning. *ACM Transactions on Management Information Systems*, 9:1–17.