Presence of Corporate Reputation Cues in Company Vacancy Texts Boosts Vacancy Attractiveness as Perceived by Employees

R. E. Loke [a] and F. Betten
Centre for Market Insights, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Abstract: Attracting the best candidates online for job vacancies has become a challenging task for companies. One thing that could influence the attractiveness of organisations for employees is their reputation that is an essential component of marketing research and plays a crucial role in customer and employee acquisition and retention. Prior research has shown the importance for companies to improve their corporate reputation (CR) for its effect on attracting the best candidates for job vacancies. Company ratings and vacancy advertisements are nowadays a massive, rich valued, online data source for forming opinions regarding corporations. This study focuses on the effect of CR cues that are present in the description of online vacancies on vacancy attractiveness. Our findings show that departments that are responsible for writing vacancy descriptions are recommended to include the CR themes citizenship, leadership, innovation, and governance and to exclude performance. This will increase vacancies' attractiveness which helps prevent labour shortage.

1 INTRODUCTION

Organisations face hiring challenges resulting from intensive knowledge work, labour shortages, competition for candidates, and workforce variety (Lee, 2011). For example, the labour shortage indicated by business owners in January 2020 was 22% in The Netherlands and constituted an obstacle to many business activities (Arbeidsmarkt in Cijfers - Intelligence Group, n.d.). Understanding how to attract the best candidates in recruiting as well as retaining employees helps companies to prevent hiring challenges and has become increasingly more important to companies over the years (Chapman et al., 2005; Küpper et al., 2021; Lievens et al., 2001). Hence, research and development (R&D) in the Human Resources (HR) divisions of companies has been quite intensive during the last decades (Bejtkovský, 2018; Chapman et al., 2005; Lee, 2011). For instance, there is a vast amount of literature now readily available for adequate matching tools between supply and demand that has become generally applicable after the transition to online availability of HR data on the public web (Bondarouk et al., 2013; Kureková et al., 2011; Yi et al., 2007). R&D in the

field of CR of companies has been equally active in this period and although traditionally being a survey-oriented field, a similar shift towards digitally based public data sources on the web has been made (Boselli et al., 2018; Lo et al., 2020; Spangler et al., 2007). Noteworthily, the CR construct has been occasionally addressed in HR research (Cable & Turban, 2003; Lemmink et al., 2003; Xie et al., 2015). Nevertheless, although it is vital to fully understand whether and how CR influences job seekers' attraction to potential employers, it never has been exploited with the explicit goal to boost the attractiveness of online job vacancies. It is known that an organisation's reputation is an essential factor that will receive attention when job seekers have no or limited information about the working conditions at a company (Ruth & York, 2004; Stockman et al., 2020). The analysis of online job vacancies represents a competitive advantage to labour market stakeholders compared to traditional survey-based analysis as it aims to diminish the time-to-market of analysis by moving towards a fact-based decision-making model (Boselli et al., 2018). However, according to Lo et al. (2020), the extraction of online semi-structured data with web scraper and crawler

[a] https://orcid.org/0000-0002-7168-090X

techniques that are consequently processed with big data solutions that enable the measurement of vacancies' text contents to better understand vacancies' attractiveness levels has mainly been overlooked until recently. We add to these findings by observing that no specific attention has been paid yet in literature to attempting to classify relevant CR concepts within vacancy texts for investigating whether the consciousness measurement of the presence of such triggered concepts ensures vacancy attractiveness. The current study aims to fill this gap in the literature.

This paper consists of an extensive literature review combined with quantitative data analysis to answer the main research question: Does including relevant CR keywords in a vacancy description positively influence online vacancy attractiveness? The structure of the paper is as follows. Relevant conceptual models as well as the formulation of the involved hypotheses from the theoretical framework will be discussed in section 2. In section 3, the methodology including data collection will be discussed. In section 4, the results of our analysis will be given. Finally, section 5 will describe the implications, recommendations and limitations of our research.

2 LITERATURE REVIEW

In this section, relevant work concerning CR and HR is discussed. Four models will be introduced in the following part. The first one is the reasoned-action attitude model by Albassami et al. (2015) discussed in 2.1.1 that states that the experiences with a company's reputation will eventually impact the behaviours concerning that company. This model is essential to understand one's behaviour concerning an organisation. Secondly, the hypothetical model of CR by Fombrun et al. (2015) consists of seven concepts discussed in 2.1.2. This model is one of the leading models in CR and has been cited many times; therefore, it will be used as a base in this study for defining CR concepts. Thirdly, the conceptual model by Neville et al. (2005) discussed in 2.1.3 supports one of the concepts by Fombrun et al. (2015) by explaining that Corporate Social Performance (CSP) influences CR. Finally, the conceptual framework by Küpper et al. (2021) makes the main subjects of this paper come together in 2.2. Their model shows the relationship between reputation, employer attraction, job search and choice. Studying these four models provides essential knowledge on CR in HR. Understanding is achieved by looking at the effect on

several stakeholders, the concepts that define CR, and the influence CR can have on the new job search. The conceptual framework with these four models provides the rich theoretical background that is needed to formulate our hypotheses in the upcoming sections.

In our research, certain CR cues as manifested in a vacancy text are the independent variables. Vacancy attractiveness is the dependent variable, and, in section 3, an explanation will follow on how it can be measured. Since natural language processing (NLP) in HR plays a prominent role, it is being introduced to the reader in section 2.3.

2.1 Corporate Reputation

An extensive literature review concluded with the following definition of CR: "Observers' collective judgments of a corporation based on assessments of the financial, social, and environmental impacts attributed to the corporation over time" (Barnett et al., 2006, p. 13). Ruth and York (2004) argue that companies with low reputation capital are often disadvantaged in recruiting and retaining top employees compared to companies with a good reputation. This partly explains the growing interest in CR among educators, administrators, and policymakers since they see CR as a precious asset that enables defending a competitive position (Tischer & Hildebrandt, 2014; Wang et al., 2016).

2.1.1 CR Among Stakeholders

Albassami et al. (2015) consider CR a significant corporate marketing component that differs among stakeholder groups. Their research focuses on CR as an independent variable that influences customer loyalty. Certain variables influence the beliefs about CR. First, the beliefs about CR control the attitude towards CR. Next, the attitude towards CR affects the intentions concerning CR and which variables finally influence the behaviours for CR. The experiences with a specific company and retrieved information from other sources about that company affect the beliefs about a company (Albassami et al., 2015).

Another topic that is highly related due to the high experience factor and that received attention in the last years is creating an image of a desirable and unique employer using online social media (Bondarouk et al., 2013; Fan et al., 2013). According to Bondarouk et al. (2013), focussing on this 'employer branding' attracts the best candidates. Abimbola and Vallaster (2007) agree with that by arguing that substantial firms are made by integrating

branding and reputation building in organisations' strategies. Furthermore, since employees are partly responsible for branding and online reputation building, researchers also focussed on their awareness and influence on their reputation (Helm, 2011; Schaarschmidt & Walsh, 2020; Xiong et al., 2013).

2.1.2 Corporate Reputation System

Fombrun et al. (2015) give seven underlying concepts that contribute to the reputation of a company and that form the basis of RepTrak; see Table 1. The concepts have been found by exhaustively testing across five different stakeholder groups in multiple industries and six countries.

Table 1: Concepts within the CR umbrella and their descriptive elements by Fombrun et al. (2015).

Concept	Description
Products	high quality, good value, stands behind, meets customer needs
Innovation	innovative, first to market, adapts to change
Workplace	rewards employees fairly, concern for employees' well-being, equal opportunities in workplace
Governance	open and transparent, behaves ethically, fair in doing business
Citizenship	protects environment, supports good causes, positive influence on society
Leadership	strong and appealing leader, clear vision of future, well organised, excellent managers
Performance	profitable, good financial results, strong growth prospects

2.1.3 CR in Financial Performance

Neville et al. (2005) investigated CR as a direct and significant factor influencing a firm's financial performance. In their conceptual model, corporate social performance (CSP) affects CR. This relationship is mediated by two factors: corporate strategy fit and reputation management capability. CSP refers "to the principles, practices, and outcomes of businesses' relationships with people, organisations, institutions, communities, societies, and the earth, in terms of the deliberate actions of businesses toward these stakeholders as well as the unintended externalities of business activity" (Corporate Social Performance - Management - Oxford Bibliographies, n.d.). Since CSP is about society, taking care of the environment and the earth, there are similarities to the Fombrun et al. (2015) concept of citizenship. Neville et al. (2005), Fombrun

et al. (2015), and the results by Peak Sales Recruiting (2018) all consequently consider citizenship affecting CR. Therefore, the first hypothesis will be formulated as follows:

H1: Including citizenship within a vacancy text positively influences a vacancy's attractiveness.

A positive relation is being expected; it is the first time in HR research that this hypothesis will be exhaustively tested with a massive online dataset at hand.

2.2 Corporate Reputation in HR

Multiple reasons underpin the importance for companies of leveraging their CR within vacancy texts. Three of them will be discussed. Firstly, companies' fundamental responsibility is to preserve and enhance their reputation, as reputation highly impacts stakeholders' beliefs and actions (Ruth & York, 2004). Therefore, reputation could also influence job seekers' interest to apply for vacancies. Secondly, research has been done on understanding vacancies' attractions and job choices. According to Chapman et al. (2005), six broad factors are found as predictors of applicant attraction: 1) job and organisational characteristics (evaluation of the job attributes or vacancy characteristics such as pay, benefits and company image and location), 2) recruiter characteristics (the perceptions of a recruiter's age, function and traits), 3) perceptions of the recruitment (the reception of personal attention and timely information), 4) perceived fit (applicants seek for a fit with the organisation or with the type of job), 5) perceived alternatives (viable employment opportunities) and 6) hiring expectancies (applicant's evaluations of the likelihood of being selected). The job and organisational characteristics and the perceived fit are based on a company's reputation. Therefore, vacancies' attractiveness could be based on a company's reputation. Thirdly, Küpper et al. (2021) created a conceptual model in which employer brand knowledge (which consists of familiarity, reputation, and image) influences employer attraction. That employer attraction, consequently, affects the new job search. Several studies concluded the positive impact that corporation reputation and employer attractiveness have on job seeker's job pursuit intentions (Lemmink et al., 2003; Sivertzen et al., 2013; Wayne & Casper, 2012; Xie et al., 2015).

In 2.2.1, factors important in the new job search will be discussed and matched with concepts of the conceptual model on CR by Fombrun et al. (2015). Consequently, in 2.2.2, two recruitment outcomes in the new job search will be elaborated: job pursuit

intentions and job-organisation attraction. In 2.2.3, the conceptual model of the antecedents of application intentions by Lemmink et al. (2003) will be discussed.

Just like in 2.1, the comparison of the similarities between the theories and the CR model by Fombrun et al. (2015) provides the theoretical background for the formulated hypotheses in the upcoming sections.

2.2.1 New Job Search

Peak Sales Recruiting surveyed in 2018 with recent college graduates who were asked to rank 21 items from most important to least important when considering a job opportunity. The top 5 items they found were: Work/Life Balance, Promotion Opportunities, Meaningful Work, Job Stability and Fun Work Culture (Millennials and the B2B Sales Industry [Study], n.d.). Both the first two as the last two items show similarities with the workplace concept of CR (Fombrun et al., (2015). Additionally, Meaningful Work could be matched with the citizenship concept by supporting good causes and protecting the environment.

2.2.2 Recruiting Outcomes

Section 2.1. shows the urgency of understanding how to leverage CR. Chapman et al. (2005) did a meta-analysis of 667 coefficients investigating relationships between recruiting outcomes. They used the following definition of recruiting by Breaugh (1992): "Employee recruitment involves those organisational activities that (1) influence the number and/or types of applicants who apply for a position and/or (2) affect whether a job offer is accepted" (p. 4). Two recruitment outcomes they found in their study are critical in the new job search. Firstly, job pursuit intentions comprise all variables that imply a person's willingness to apply for a job and visit a website. Secondly, job-organisation attraction has to do with the job seeker's overall evaluation of the job and the company (Chapman et al., 2005). These two factors are essential to determine a vacancy's attractiveness and whether a candidate would consider a position. Furthermore, Chapman et al. (2005) noted that job seekers mostly base their job choices on job attributes and vacancy characteristics.

2.2.3 Application Intentions

Lemmink et al. (2003) created a conceptual model containing the antecedents of application intentions. Their model starts with (A) information source use and (B) familiarity with a company supporting the

theories of Albassami et al. (2015) and Prabhu & Stewart (2001) by referring to essential sources that give information about a company that influence CR. The familiarity affects two factors within their model: (C) corporate image and (D) company employment image. These two influence the application intentions. The authors Lemmink et al. took a closer look at the variables that influence corporate image, namely, organisational management (E) and short-term experience (F). E is about the higher management committed to the organisation and about introducing innovations regularly. This description matches with the leadership and innovation concepts of Fombrun et al. (2015). Therefore, the second and third hypothesis are formulated as:

H2: Including leadership within a vacancy text positively influences a vacancy's attractiveness.

H3: Including innovation within a vacancy text positively influences a vacancy's attractiveness.

F is about appealing advertising of products and services, the applicant hearing positive things and offering the first choice for qualitative outcomes. Research by Stockman et al. (2020) confirms that a higher attraction for a vacancy is reached by having no negative worth of mouth (WOM). This description matches the products dimension of the Fombrun et al. (2015) model. Therefore, the fourth hypothesis is formulated as:

H4: Including products within a vacancy text positively influences a vacancy's attractiveness.

D is about job characteristics (G) and extra organisational benefits (H). G is about having a challenging job, many responsibilities, and various tasks. H is about having guidance through a career and attending exciting training. Both descriptions match the workplace concept of Fombrun et al. (2015). Moreover, as has been mentioned in section 2.2.1, four out of the five top 5 items that make an applicant to decide to apply for a company show similarities with this workplace concept (Millennials and the B2B Sales Industry [Study], n.d.). Therefore, the fifth hypothesis is formulated as:

H5: Including workplace within a vacancy text positively influences a vacancy's attractiveness.

Two concepts of the Fombrun (2015) conceptual model will be excluded from this research. There is insufficient overlap with the other conceptual models used in HR. Therefore, the sixth and seventh hypotheses are formulated as:

*H6: Including performance within a vacancy text does **not** affect a vacancy's attractiveness.*

*H7: Including governance within a vacancy text does **not** affect a vacancy's attractiveness.*

2.3 NLP in HR

NLP is a computerised approach to analyse texts (Kupiyalova et al., n.d.; Liddy, n.d.) and was already successfully deployed in the recruiting process. Boselli et al. (2018) have built an ML model to classify web job vacancies' titles and texts. This model enables evidence-based decision-making, which in turn supports policy designing and the evaluation of activities. According to Boselli et al. (2018), both the public and private markets are interested in analysing web job vacancies by text classification. Their machine learning model has implications for the e-recruitment process, i.e., matching candidates' profiles with the available job vacancies and descriptions. Another NLP application in HR is screening resumes by focussing on specific keywords. This could be an enormous task while looking for suitable candidates when many applicants apply (Cabrera-Diego et al., 2019; Roy et al., 2020). This is called a resume recommendation system. On the contrary, NLP applications are being used to help job seekers find vacancies that match their personnel interests. This is called a job recommendation system (T. Al-Otaibi, 2012).

The NLP opportunities to process vacancy description data are endless. Semantic search including keyword-based search enriched with synonyms that we do in this paper enables searching a text for meaning. To do this, the algorithm takes precise meaning or vocabulary from the text (Bast et al., 2016; Kupiyalova et al., n.d.).

The Natural Language Toolkit (NLTK) is an open-source Python library that includes extensive software, data, and documentation (Bird et al., 2009) that seems suitable for this study. Therefore, the analysis of this study will be based on the NLTK toolkit.

3 METHODOLOGY

3.1 Data Streams

Our work will focus on vacancies offered within the Netherlands. As many websites offer vacancies, it is essential to choose the best option. According to a blog by Business Insider (2020), www.indeed.com (hereafter referred to as Indeed) is the vacancy website most often used by active jobseekers but is not the favourite of employers and intermediaries. As our work focuses on preventing labour shortage, the focus needs to be on the jobseekers, making Indeed a suitable platform. The site has built-in filters to quickly select certain functions, keywords, companies, and regions that make Indeed easy to use. Sainju (2020) agrees with this as Indeed possesses an easy-to-use user interface and extensive features. Feng (2020) collected over 1,2 million job reviews from Indeed. Feng (202) used machine learning and NLP techniques to gather qualitative data about firms' dimensions of job satisfaction and management sentiment. Sainju (2020) also investigated employee reviews from Indeed to identify prominent aspects in employee reviews and automatically gather underlying topics that are notable factors related to employee satisfaction.

Indeed has compared to other websites the highest number of monthly visitors who are most likely active job seekers since there is no focus on specific educational levels. Moreover, there are many features available such as location, tenure, and salary.
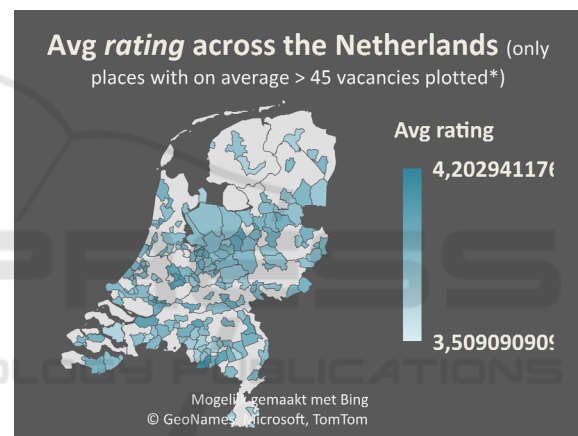

Figure 1: The average rating across the Netherlands.

3.2 Data Collection

The web crawler was developed in-house with Scrapy 2.4.1 (web spider for crawling websites). The scraper's first and most important task is to analyse the target website https://nl.indeed.com/vacatures-in-Nederland. After producing the proper scraping script, 255.104 vacancies from Indeed were scraped. Nine different languages were detected in these vacancies. Only the English and Dutch vacancies were used for the analysing part. For each vacancy, the following elements were scraped: 1) title, 2) company, 3) location, 4) description, 5) day published, 6) company rating from 1 to 5 and 7) URL. Because we will focus on CR cues being present in the description to find the influence on vacancy attractiveness, the description and the rating are the most important variables in this study. In Figure 1, the average rating across the Netherlands is plotted. The

highest average rating was 4,2 in Alblasserdam in the province of Zuid-Holland. The lowest average rating in this plot is 3.5. The figure shows that the average rating is not evenly distributed across municipalities in the country.

3.3 Vacancy Text Pre-Processing

The data collected from Indeed were robustly stored and pre-processed. For each of the vacancies, the following steps were taken: (1) Data cleaning was done to remove HTML tags and unnecessary signs (i.e. \n or <200). In addition, cleaning was done to create additional columns like the number of employees that rated a company and to merge columns with company and location values. Also, the column having the number of days that a vacancy was online needed to be cleaned to just contain the digit showing the age of the vacancy. After these steps, 150.316 vacancies were kept and used in the following steps. (2) Word tokenisation: each of the words within the vacancy description column was tokenised. (3) Stemming and upper to lowercase: each word has been suffix-stripped to its word stem, and each of the terms was lowercased. We note that tokenisation helps interpret the meaning of a text by analysing the sequence of the words while stemming is a method for reducing derived words to their stem, root, or origin form (Haji Ali, 2015). Changing the upper cases to lower cases is helpful in the process of comparing the composed lists of words to the vacancy descriptions, as it standardizes the form of writing.

3.4 Vacancy Text Processing

After the vacancies were pre-processed, the following steps were taken in NLTK.

As specified in the book on NLP by Bird et al. (2009), you need to define a collection of text files that is called a corpus. As there are no existing corpora that precisely describe the seven concepts of CR, a Taylor-made corpus has been created. A separate text file has been composed and computed for each of the seven concepts, including keywords and synonyms. Computing synonyms can be done via several Python packages such as Wordnet, Textblob and PyDictionary. Wordnet is an NLTK corpus reader which gives access to the Open Multilingual WordNet (WordNet Interface, n.d.). Textblob provides an API for processing NLP tasks like sentiment analysis, classification and translation using Wordnet's library (TextBlob: Simplified Text Processing — TextBlob 0.16.0 Documentation, n.d.). PyDictionary uses Wordnet for understanding

meanings, Google for translations, and synonym.com for receiving synonyms (Pandas Documentation — Pandas 1.2.4 Documentation, n.d.). As these options all use Wordnet as input, and the tokenising and stemming has been systematically done using NLTK as well, the retrieving of the synonyms will be done using NLTK's package Wordnet. As this package is written in English, the synonyms received from Wordnet were consequently translated to Dutch as the data frame consisted of both English and Dutch vacancies. Consequently, to generate the Dutch corpus, the English keywords were translated to Dutch using https://www.deepl.com/translator.

The input for the keyword text files is a combination of different keywords from relevant theory in the field and the synonym output created by Wordnet. In all cases, the descriptions of the seven concepts by Fombrun et al. (2015) were used as input: (1) Product is defined as high quality, good value, stands behind and meets customer needs. The synonyms of a product are merchandise and ware, so these words and product are added to the list of keywords. In addition, quality and prime refer to quality and good value. As the keyword search only allows single words, "stands behind" and "meets customer needs" were omitted. (2) Innovation is defined as innovative, first to market and adapts to change. The synonyms of innovation are invention, excogitation, conception, design, initiation, founding, foundation, institution, origination, creation, introduction, instauration. Creativity has been added to this list as this word is also essential in innovating. (3) Workplace is defined as rewards employees somewhat, concern for employees' well-being and equal opportunities in the workplace. There are no synonyms of the workplace. Based on Fombrun et al. (2015), the following keywords are added to the keywords: workplace, rewarding, wages, well-being and equal. Peak Sales Recruiting argue that work/life balance, promotion opportunities, job stability and fun work culture are vital when considering job opportunities. Therefore, career, promotion, stability, culture, fun and balance were added. Lemmink et al. (2003) consider having guidance through a career and attending exciting training as contributing to the company employment image, which is also part of the workplace. Therefore, guidance, training, education and developing were added. Also, the company employment image exists of having a challenging job, many responsibilities, and various tasks. Consequently, variety, diversity and challenging were also added. (4) Governance is defined as open and transparent, behaves ethically and fair in doing business. Therefore, governance, open, transparent,

ethical, fair, and honest were added as keywords. (5) Citizenship is defined as protects the environment, supports good causes and positive influence on society). Neville et al. (2015) connect corporate social performance to reputation management by focussing on society, taking care of the environment and the earth and Peak Sales Recruiting found that doing meaningful work is one essential factor in considering job opportunities. Combining these sources, the following keywords were formulated: citizenship, environment, society, sustainability, earth, world, globe, community, meaning and significance. (6) Leadership is defined as strong and appealing leadership, clear vision of future, well organised and excellent managers. Lemmink et al. (2003) discuss the effect of the higher management committed to the organisation to an organisation's image. Therefore, the keywords leadership, lead, leader, manager, and vision were added to the keywords. (7) Performance is defined as profitable, good financial results and strong growth prospects. Therefore, the words performance, profit, grow, prospect, and results were defined as keywords.

As stated before, the keywords all needed to be stemmed as the description texts were also stemmed. There is a stemmer for English words and another one for Dutch words in NLTK that we used for this task. In Tables 2 and 3 an overview of the stemmed text files containing all keywords is given.

The next step was to match the text files containing all CR keywords to the descriptions of the scraped vacancies. After this matching, seven new columns were created called 'products', 'innovation', 'workplace', 'governance', 'citizenship', 'leadership', and 'performance'. These columns were filled with a 0 in case that no relevant keyword was present in the vacancy description text and with a 1 when a relevant keyword was present. After creating the two different data frames in the Pandas Python package (e.g., one where workplace is present (1) and one where workplace is not present (0), the average rating from the rating column for both data frames was calculated and could be compared. The results of these comparisons will be discussed in section 4.

Table 2: English keywords per CR concept.

Concept	Keywords (English)
Products	product + merchandis + ware + quality + prime
Innovation	innov + excogit + concept + design + initi + found + institut + origin + creation + introduct + instaurat + creativ
Workplace	workpl + career + varieti + divers + guidanc + train + develop + educ + reward

	+ wage + wellb + equal + balanc + promot + fun + cultur + stabil + challeng
Governance	govern + open + transpar + ethic + fair + honest
Citizenship	citizenship + environ + societi + sustain + earth + world + globe + communiti + meaning + signific
Leadership	leadership + lead + leader + manag + vision
Performance	perform + profit + grow + prospect + result

Table 3: Dutch keywords per CR concept.

Concept	Keywords (Dutch)
Products	koopwar + war + productie + kwaliteit
Innovation	uitvind + innovatie + innover + innovatief + concept + ontwerp + initiatie + opricht + introductie + introducer + instauratie + creatief + creativiteit
Workplace	werklek + carrièr + verschil + verscheid + ontwikkel + begeleid + belon + opleid + gelijk + evenwicht + plezier + welzijn + gezell + stabiel + cultur + uitdag
Governance	bestur + transparant + ethisch + eerlijk
Citizenship	milieu + samenlev + duurzam + aard + wereld + gemeenschap + maatschappij + maatschappelijk + betekenis
Leadership	leiderschap + leid + directeur + begeleid + visie + werkgev
Performance	uitvoer + winst + winstgev + groeiend + vooruitzicht + resultat

3.5 Measuring a Vacancy's Attractiveness

A way must be found to measure a vacancy's attractiveness, the dependent variable in this work. Several studies show the influence of CR in the recruitment process. Cable & Turban (2003) acknowledge a vital role for CR in recruitment: "results from 339 individuals suggested that job seekers' reputation perceptions affected job pursuit because (a) individuals use reputation as a signal about job attributes, and (b) reputation affects the pride that individuals expect from organisational membership" (p. 2244). The attractiveness of a vacancy is therefore dependent on the job seekers' reputation perceptions. Furthermore, Xie et al. (2015) found that a company's reputation influences employer brand attractiveness and Boselli et al. (2018) figured that reputation affects application intentions. According to Ahamad (2020), there is an impact of reviews and ratings on job seekers decision. This has to do with the portrayal of the advantages and disadvantages of working at a company. Therefore, one of the possible measurements to

measure a vacancy's attractiveness is evaluating the rating given by (former) employees. Such rating variable and employee reviews will contribute to the job seekers' reputation perceptions of companies. At Indeed, the rating of a company varies from 0 to 5 and is given by (former) employees. For completeness sake, next to the rating, also the number of raters has been scraped per company.

To test the seven hypotheses, separate data frames will be created. For example, one data frame containing vacancies in which leadership is addressed and one data frame where no keyword referring to leadership is mentioned. From both data frames, the average rating will be calculated. The results of this will be shown in section 4. Should the data frame containing leadership keywords have a higher average rating (i.e., 3.8) than the data frame containing no keywords of leadership (i.e., 3,5), then the hypothesis will be assumed to be true, provided appropriate statistical tests apply (i.e., when the standard deviation is about 0.1).

3.6 Statistics

To determine whether the seven concepts of CR significantly influence the rating, a multiple linear regression (MLR) model was made using R. The output is shown in Figure 2. The vacancy's rating is the dependent variable and stands before the tilde sign. The respective CR columns stand behind the tilde sign being the independent variables – the predictors – that could influence the rating. As shown in the output, six out of the seven variables show highly significant p-values. The variable for performance however is an insignificant predictor (p-value of 0,95). Looking from the 2nd row to the 8th row, the respective coefficients indicate the impact of a variable on the rating. To illustrate: for every time workplace is called in a vacancy description, the rating lowers by 0,03 points, and for every time citizenship is named in the description, the rating goes up by 0,09. The rejection or acceptance of the formulated hypotheses in section 4 will partly be based on Figure 2.

The day that the vacancy has been published, the number of ratings and the language (English or Dutch) also impact the rating, as mediator, mediator, and moderator, respectively. Note that the Dutch language is not defined because of singularities with the English language; these two variables are perfectly correlated.

Having these 10 predictors in the model to predict the rating, the adjusted R-squared is 0,1261. This means that the predicting variables can explain 12,61% of the variation in the rating. Given the

complexity of the topic and of the involved analysis such weak influence could be expected and is commonly acceptable, especially since most of the explanatory variables are statistically significant.

Figure 2: Output multiple linear regression model where the rating is the dependent variable and workplace, innovation, leadership, product, governance, performance and citizenship are the independent variables and published, rating count and English and Dutch language are mediators and moderators.

Of course, we evaluated carefully whether the assumptions for applying MLR were not validated. The VIF (Variance Indicator Factor) indicates whether a predictor has a strong linear relationship with other predictors. As a rule of thumb, it is said that when the largest VIF is greater than 10, there is a reason for concerns. For our model, all VIFs were around 1, so there were no reasons for concerns. We also checked the assumptions of independent errors using the Durbin-Watson test. The conservative rule is that values less than one or greater than three should raise alarm bells. The closer to 2, the better. In this study, the Durbin-Watson statistic was 1,63.

When considering the outcomes from Figure 2, the conclusion is that there is a significant relationship between the predictors workplace, innovation, leadership, products, governance, citizenship, and the outcome variable (rating). However, not between performance and rating. The two mediators and the moderator significantly attribute to the model.

4 RESULTS

The results are based on all 158.351 vacancies, containing 3.775 companies and 3.849 different locations. Most vacancies are coming from Teleperformance Benelux, a call centre that takes over the customer support of other companies. The keywords that belong to workplace are the most

common within vacancy texts. Innovation is the least frequent category within the vacancy descriptions.

To get to the results in Table 4, the average ratings of two data frames have been compared. In the first data frame mentioned, the cues of a CR concept were present. In the second data frame, those cues were absent. As mentioned before, the CR concepts play the role of independent variables, and the average rating plays the role of dependent variable that reflects vacancy attractiveness.

Table 4: Number of vacancies and average rating per CR element.

Concept in vacancy description	Number of vacancies	Average rating
Citizenship keywords present	110.757	3.875
Citizenship keywords not present	39.559	3.763
Innovation keywords present	39.467	3.890
Innovation keywords not present	110.849	3.830
Leadership keywords present	105.249	3.876
Leadership keywords not present	45.067	3.774
Workplace keywords present	147.663	3.846
Workplace keywords not present	2.653	3.836
Product keywords present	95.356	3.843
Product keywords not present	54.960	3.850
Governance keywords present	68.524	3.873
Governance keywords not present	81.792	3.822
Performance keywords present	95.514	3.850
Performance keywords not present	54.802	3.837

The average ratings in the last column of Table 4 will contribute to accepting or rejecting the formulated hypotheses. Concerning H1, the average rating is higher when citizenship keywords are present in the description text (3.875 vs 3.763). Moreover, as can be seen in Figure 2, there is a significant relationship between citizenship and rating, with a high significance level. For example, for every vacancy text where citizenship is present, the attractiveness of the vacancy will increase by 0,09. Therefore, H1 is accepted. Similar reasoning can be applied for the hypotheses on leadership and innovation (H2 and H3 both accepted).

Although similar reasoning can also be applied for the hypothesis on products, the average rating is higher when product keywords are not present in the description text (3.843 vs 3.850) (H4 rejected).

Also, although similar reasoning can also be applied for the hypothesis on workplace, the

multivariate regression in Figure 2 shows that for every vacancy where workplace is present, the attractiveness of the vacancy will decrease by 0,032 (H5 rejected).

Regarding performance, there is indeed as has been expected no significant contribution of this variable in the MLR (H6 accepted).

Regarding governance, there is contrary to what has been expected a significant positive influence of this variable in the MLR (H7 rejected).

To summarize, in our research it has been found on one side of the spectrum that including citizenship, leadership and innovation cues in vacancy texts increases a vacancies' attractiveness both on theoretical grounds as well as empirical grounds. The same holds for governance but it is only on empirical grounds. On the other side of the spectrum, it has been found that including performance cues makes no sense both theoretically and empirically. Including products and workplace cues makes no sense empirically.

5 CONCLUSION

5.1 Research Implications

In this research, we make at least two significant contributions. Firstly, little attention has been paid to detecting cues of CR within vacancies and investigating whether such cues impact vacancy attractiveness. This study was the first time that hypotheses have been tested on systematically detecting CR concepts within vacancies' descriptions. Secondly, the developed NLP approach has been successfully applied to semantic search regarding CR. In NLTK we designed our own custom corpora for Dutch and English language by creating Taylor-made corpus per language for CR in HR relevant for this study. Synonyms were generated with Wordnet for the chosen keywords. For the keyword search, the words within the corpora have been stemmed, as all vacancy descriptions were also stemmed. This approach ensured that, for example, "innovative" (ENG), "innovation" (ENG), "innovatief" (NL) and "innoverend" (NL) were all matched with innovation. These four words all became innov after the stemming, so "innov" was one of the keywords in the corpus of the category innovation.

Nowadays, one data stream is specifically relevant in measuring CR in vacancies: online vacancies. Vacancies were already scraped and analysed several times before in literature, but the focus was primarily on extracting titles and skills.

This study provides a new application of web scraping by focusing on concepts of reputation within vacancy descriptions. The conceptual model by Fombrun et al. (2015) is the most leading conceptual model used in this study since it is leading in CR research. The seven concepts of CR were systematically compared to other conceptual models on CR by Albassami et al., Lemmink et al., Neville et al., and Küpper et al in this study. The goal of this comparison was to gather enough theoretical foundation to formulate relevant hypotheses. It is important to note that in the conceptual model by Küpper et al. employer branding is described and that these researchers state that familiarity, reputation, and image contribute to employer attraction as well as that employer attraction consequently affects job search and choice as well as motivation and retention. This model has been used as the bridge between CR and job vacancies in HR.

5.2 Practical Implications

Departments responsible for writing vacancy descriptions are recommended to smartly cover the themes citizenship, leadership, innovation, and governance as well as to strategically not waste too much energy on including company performance details. This will increase vacancy attractiveness, which helps prevent labour shortage.

All premises on citizenship, leadership and innovation are backed up in this study by relevant literature, confirmed on the massive Indeed dataset that we collected and reinforced with multivariate regression. For the governance theme, relevant literature will need to be examined and evaluated to theoretically ground the empirical findings that have been shown in this study.

5.3 Limitations and Future Research

In this research, synonyms of CR keywords were retrieved from NLTK's wordnet package. However, synonyms of all words in vacancy descriptions could also be extracted and focussed on to apply a two-way synonym mechanism instead of a one-way synonym mechanism only. On the one hand, one would expect better results, as more synonyms would match with the created CR corpora. On the other hand, it could cause some computational difficulties as data frame sizes would increase exponentially.

It will be a challenge to test and experiment with more or other NLP techniques which could cause more accurate results soon. The chosen NLTK way of stemmed keyword searching may not be the most feasible option to make recommendations on the stemmed keywords in description texts as context is not considered. Also, adding wordnet's similarity scores (followed with thresholds) could improve the analysis (Ezzikouri et al., 2019; WordNet Interface, n.d.). Future research should focus on applying advanced semantic search to description texts instead of looking for keywords to be able to incorporate multi-token keywords of Fombrun, such as "meets customer needs". With the context of words and multi-token keywords included, the results of NLP would even be more accurate.

ACKNOWLEDGEMENTS

REFERENCES

Abimbola, T., & Vallaster, C. (2007). Brand, organisational identity and reputation in SMEs: An overview. *Qualitative Market Research: An Int. J.*, *10*(4), 341–348. https://doi.org/10.1108/13522750710819685

Ahamad, F. (2020). Impact of online job search and job reviews on job decision. *WSDM 2020 – Proc. of the 13th International Conference on Web Search and Data Mining*, 909–910. https://doi.org/10.1145/3336191.3372184

Albassami, F. A., Alqahtani, S. S., & Saleh, M. A. (2015). Proposing a Conceptual Model of Corporate Reputation Associated with Relationship Marketing and Customer Loyalty. *Int. J. of Marketing Studies*, *7*(6), 102. https://doi.org/10.5539/ijms.v7n6p102

Arbeidsmarkt in cijfers - Intelligence Group. (n.d.). Retrieved May 8, 2021, from https://intelligence-group.nl/nl/nieuws/arbeidsmarkt-in-cijfers

Barnett, M. L., Jermier, J. M., & Lafferty, B. A. (2006). Corporate Reputation: The Definitional Landscape. *Corporate Reputation Review*, *9*(1), 26–38. https://doi.org/10.1057/palgrave.crr.1550012

Bast, H., Buchhold, B., & Haussmann, E. (2016). Semantic search on text and knowledge bases. *Foundations and Trends in Information Retrieval*, *10*(2–3), 119–271. https://doi.org/10.1561/1500000032

Bejtkovský, J. (2018). Factors influencing the job search and job selection in students of generation y in the Czech republic in the employer branding context.

Management and Marketing, *13*(3), 1133–1149. https://doi.org/10.2478/MMCKS-2018-0028

Bird, S., Klein, E., Loper, E., Bird, S., Klein, E., Loper, E., Bird, C. S., Klein, E., & Loper, E. (2009). *Natural Language Processing With Python*.

Bondarouk, T., Ruël, H., Axinia, E., & Arama, R. (2013). What is the future of employer branding through social media? Results of the delphi study into the perceptions of hr professionals and academics. In *Advanced Series in Management* (Vol. 12, Issue Aug. 2019). https://doi.org/10.1108/S1877-6361(2013)0000012006

Boselli, R., Cesarini, M., Mercorio, F., & Mezzanzanica, M. (2018). Classifying online Job Advertisements through Machine Learning. *Future Generation Computer Systems*, *86*, 319–328. https://doi.org/10.1016/j.future.2018.03.035

Breaugh, J. A. (1992). *Recruitment: Science and practice.* PWS- Kent.

Cable, D. M., & Turban, D. B. (2003). The Value of Organisational Reputation in the Recruitment Context: A Brand-Equity Perspective. *J. of Applied Social Psychology*, *33*(11), 2244–2266. https://doi.org/10.1111/j.1559-1816.2003.tb01883.x

Cabrera-Diego, L. A., El-Béze, M., Torres-Moreno, J. M., & Durette, B. (2019). Ranking résumés automatically using only résumés: A method free of job offers. *Expert Systems with Applications*, *123*, 91–107. https://doi.org/10.1016/j.eswa.2018.12.054

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organisations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *J. of Applied Psychology*, *90*(5), 928–944. https://doi.org/10.1037/0021-9010.90.5.928

Corporate Social Performance - Management - Oxford Bibliographies. (n.d.). Retrieved April 17, 2021, from https://www.oxfordbibliographies.com/view/document/obo-9780199846740/obo-9780199846740-0099.xml

Ezzikouri, H., Madani, Y., Erritali, M., & Oukessou, M. (2019). A new approach for calculating semantic similarity between words using WordNet and set theory. *Procedia Computer Science*, *151*(2018), 1261–1265. https://doi.org/10.1016/j.procs.2019.04.182

Fan, D., Geddes, D., & Flory, F. (2013). The toyota recall crisis: Media impact on toyota's corporate brand reputation. *Corporate Reputation Review*, *16*(2), 99–117. https://doi.org/10.1057/crr.2013.6

Farhadi, R., & Nanda, V. (2021). What do employees know? Quality perception and 'over-satisfaction' in firms going public. *J. of Corporate Finance*, *66* (Nov. 2020), 101779. https://doi.org/10.1016/j.jcorpfin.2020.101779

Feng, S. F. (2020). Job Satisfaction, Management Sentiment, and Financial Performance: Text Analysis with Job Reviews from Indeed. com. *Authorea Preprints*, 1–18. https://www.authorea.com/doi/pdf/10.22541/au.159596366.69672878

Fombrun, C. J., Ponzi, L. J., & Newburry, W. (2015). Stakeholder tracking and analysis: The RepTrak® System for measuring corporate reputation. *Corporate Reputation Review*, *18*(1), 3–24. https://doi.org/10.1057/crr.2014.21

Haji Ali, N. (2015). *Porter Stemming Algorithm for Semantic Checking Nakhoda Dashboard : The Use of Analytics in Analyzing Student Engagement in e-Learning View project*. https://www.researchgate.net/publication/260385215

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Helm, S. (2011). Employees' awareness of their impact on corporate reputation. *J. of Business Research*, *64*(7), 657–663. https://doi.org/10.1016/j.jbusres.2010.09.001

Kupiyalova, A., Satybaldiyeva, R., & Aiaskarov, S. (n.d.). *Semantic search using Natural Language Processing*. https://doi.org/10.1109/CBI49978.2020.10065

Küpper, D. M., Klein, K., & Völckner, F. (2021). Gamifying employer branding: An integrating framework and research propositions for a new HRM approach in the digitized economy. *Human Resource Management Review*, *31*(1), 100686. https://doi.org/10.1016/j.hrmr.2019.04.002

Kureková, L. M., Beblavý, M., & Thum-Thysen, A. (2011). *Using online vacancies and web surveys to analyse the labour market: a methodological inquiry*. https://doi.org/10.1186/s40172-015-0034-4

Lee, I. (2011). Modeling the benefit of e-recruiting process integration. *Decision Support Systems*, *51*(1), 230–239. https://doi.org/10.1016/j.dss.2010.12.011

Lemmink, J., Schuijf, A., & Streukens, S. (2003). The role of corporate image and company employment image in explaining application intentions. *J. of Economic Psychology*, *24*(1), 1–15. https://doi.org/10.1016/S0167-4870(02)00151-4

Liddy, E. D. (n.d.). *SURFACE SURFACE Center for Natural Language Processing School of Information Studies (iSchool) 2001 Natural Language Processing 1*. Retrieved April 18, 2021, from https://surface.syr.edu/cnlp

Lievens, F., Decaesteker, C., Coetsier, P., & Geirnaert, J. (2001). Organisational attractiveness for prospective applicants: A person-organisation fit perspective. *Applied Psychology*, *50*(1), 30–51. https://doi.org/10.1111/1464-0597.00047

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–184. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Lo, H. C., Koedijk, K. G., Gao, X., & Hsu, Y. T. (2020). How do job vacancy rates predict firm performance? A web crawling massive data perspective. *Pacific Basin Finance J.*, *62*. https://doi.org/10.1016/j.pacfin.2020.101371

Luo, N., Zhou, Y., & Shon, J. J. (2016). Employee satisfaction and corporate performance: Mining employee reviews on glassdoor.com. *2016 Int. Conf. on Information Systems, ICIS 2016*, *2015*, 1–16.

McNeil, T. (2021). *Stronger than the pandemic.* https://www.linkedin.com/feed/update/urn:li:activity:6780139838592155648/?updateEntityUrn=urn%3Ali%

3Afs_feedUpdate%3A%28V2%2Curn%3Ali%3Aacti
vity%3A6780139838592155648%29

Melville, P., Gryc, W., & Lawrence, R. D. (2009).
Sentiment analysis of blogs by combining lexical
knowledge with text classification. *Proc. of the ACM
SIGKDD Int. Conf. on Knowledge Discovery and Data
Mining*, 1275–1283. https://doi.org/10.1145/155701
9.1557156

Millennials and the B2B Sales Industry [Study]. (n.d.).
Retrieved May 12, 2021, from https://www.peak
salesrecruiting.com/blog/millennials-and-sales-study/

Neville, B. A., Bell, S. J., & Mengüç, B. (2005). Corporate
reputation, stakeholders and the social performance-
financial performance relationship. *European J. of
Marketing*, *39*(9–10), 1184–1198. https://doi.org/
10.1108/03090560510610798

pandas documentation — pandas 1.2.4 documentation.
(n.d.). Retrieved April 18, 2021, from https://
pandas.pydata.org/pandas-docs/stable/index.html

Pang, B., Lee, L., & Vaithyanathan, S. (2002). *Thumbs up?
Sentiment Classification using Machine Learning
Techniques*. https://doi.org/10.3115/1118693.1118704

Prabhu, J., & Stewart, D. W. (2001). Signaling strategies in
competitive interaction: Building reputations and
hiding the truth. *J. of Marketing Research*, *38*(1), 62–
72. https://doi.org/10.1509/jmkr.38.1.62.18826

Roy, P. K., Chowdhary, S. S., & Bhatia, R. (2020). A
Machine Learning approach for automation of Resume
Recommendation system. *Procedia Computer Science*,
167, 2318–2327. https://doi.org/10.1016/j.procs
.2020.03.284

Ruth, J. A., & York, A. (2004). Framing information to
enhance corporate reputation: The impact of message
source, information type, and reference point. *J. of
Business Research*, *57*(1), 14–20. https://doi.org/
10.1016/S0148-2963(02)00270-9

Sainju, B. (2020). Job Satisfaction and Employee Turnover
Determinants in Fortune 50 Companies: Insights from
Employee Reviews from Indeed.com. *Decision Support
Systems*, *Sept. 2020*, 113582. https://doi.org/10.1016/j.
dss.2021.113582

Schaarschmidt, M., & Walsh, G. (2020). Social media-
driven antecedents and consequences of employees'
awareness of their impact on corporate reputation. *J. of
Business Research*, *117* (Nov. 2018), 718–726. https://
doi.org/10.1016/j.jbusres.2018.11.027

Sivertzen, A. M., Nilsen, E. R., & Olafsen, A. H. (2013).
Employer branding: Employer attractiveness and the
use of social media. *J. of Product and Brand
Management*, *22*(7), 473–483. https://doi.org/10.1108/
JPBM-09-2013-0393

Spangler, S., Chen, Y., Proctor, L., Lelescu, A., Behal, A.,
He, B., Griffin, T. D., Liu, A., Wade, B., & Davis, T.
(2007). *COBRA - Mining Web for COrporate Brand
and Reputation Analysis*. https://doi.org/10.110
9/WI.2007.32

Stockman, S., Van Hoye, G., & da Motta Veiga, S. (2020).
Negative word-of-mouth and applicant attraction: The
role of employer brand equity. *J. of Vocational

Behavior*, *118* (Dec. 2019), 103368. https://doi.org/1
0.1016/j.jvb.2019.103368

Sürücü, L., & Maslakçi, A. (2020). Validity and Reliability
in Quantitative Research. *Business & Management
Studies: An Int. J.*, *8*(3), 2694–2726. https://doi.org/
10.15295/bmij.v8i3.1540

T. Al-Otaibi, S. (2012). A survey of job recommender
systems. *Int. J. of the Physical Sciences*, *7*(29).
https://doi.org/10.5897/ijps12.482

*TextBlob: Simplified Text Processing — TextBlob 0.16.0
documentation*. (n.d.). Retrieved June 17, 2021, from
https://textblob.readthedocs.io/en/dev/

Thangaraj, M., & Sivakami, M. (2018). Text classification
techniques: A literature review. *Interdisciplinary J. of
Information, Knowledge, and Management*, *13*, 117–
135. https://doi.org/10.28945/4066

Tischer, S., & Hildebrandt, L. (2014). Linking corporate
reputation and shareholder value using the publication of
reputation rankings. *J. of Business Research*, *67*(5), 1007
–1017.https://doi.org/10.1016/j.jbusres.2013.08.007

Wang, D. H. M., Yu, T. H. K., & Chiang, C. H. (2016).
Exploring the value relevance of corporate reputation: A
fuzzy-set qualitative comparative analysis. *J. of Business
Research*, *69*(4), 1329–1332. https://doi.org/10.1016
/j.jbusres.2015.10.101

Wayne, J. H., & Casper, W. J. (2012). Why does firm
reputation in human resource policies influence college
students? The mechanisms underlying job pursuit
intentions. *Human Resource Management*, *51*(1), 121–
142. https://doi.org/10.1002/hrm.21461

WordNet Interface. (n.d.). Retrieved June 17, 2021, from
https://www.nltk.org/howto/wordnet.html

Xie, C., Bagozzi, R. P., & Meland, K. V. (2015). The
impact of reputation and identity congruence on
employer brand attractiveness. *Marketing Intelligence
and Planning*, *33*(2), 124–146.
https://doi.org/10.1108/MIP-03-2014-0051

Xiong, L., King, C., & Piehler, R. (2013). "That's not my
job": Exploring the employee perspective in the
development of brand ambassadors. *Int. J. of
Hospitality Management*, *35*, 348–359. https://doi.
org/10.1016/j.ijhm.2013.07.009

Yi, X., Allan, J., & Croft, W. B. (2007). Matching resumes
and jobs based on relevance models. *Proc. of the 30th
Annual Int. ACM SIGIR Conference on Research and
Development in Information Retrieval, SIGIR '07*, 809–
810. https://doi.org/10.1145/1277741.1277920