

Vergleichende Analyse von Unternehmenswerten in Online Stellenanzeigen mittels NLP / LIWC

Marcel Herold, Marc R.H. Roedenbeck

Technische Hochschule Wildau, Hochschulring 1, 15745 Wildau

Abstract

Die Analyse von Unternehmenswerten wird schon seit Jahrzehnten durchgeführt. Es gibt verschiedene Modelle, wie zum Beispiel den Ansatz von Hofstede oder das „Universal Value Structure Modell“ von Schwartz, welche die Unternehmenswerte darstellen. Aktuelle Studien empfehlen Unternehmenswerte unter der Nutzung von Natural Language Processing und definierten Wortlisten bei größeren Textdaten zu untersuchen. Ziel der vorliegenden Analyse war zu identifizieren, ob die deutschen Übersetzungen der bisher englischsprachigen Ansätze im Kontext von Online-Stellenanzeigen funktional anwendbar sind. Daher wurden ca. 151.000 online-Stellenanzeigen von ca. 29.000 Unternehmen mittels drei Wortlisten analysiert. Der Umfang der verwendeten Listen lag zwischen 126 und 944 deutschen Wörtern. Die Ergebnisse zeigen, dass a) alle angewendeten Wortlisten grundsätzlich funktional sind, b) dass einer der Listen jedoch herausragende Ergebnisse aufweist, welcher jedoch c) um fehlende Begriffe der beiden anderen Ansätze erweitert werden kann. Somit können Unternehmen ein eigenes Werteprofil Ihrer externen Online-Kommunikation ermitteln und dieses mit den internen Wertvorstellungen bzw. dem intern erhobenen realen Werteprofil abgleichen.

1. Einleitung

Kultur, als ein integraler Bestandteil organisationaler Handlungen, oder der Einfluss von Kultur auf die Funktionalität sozialer Organisationen, ist seit den Werken von Mead (1934) und Weber (1930) ein kontinuierlich diskutiertes Thema. Neben den sichtbaren Artefakten und den Basisannahmen sind die Unternehmenswerte dabei von zentraler Bedeutung (Schein, 1985). Klassische Kulturstudien in der Organisations- und Managementforschung sind auf die Erhebung kultureller Werte einer Personengruppe (Ogbonna & Wilkinson, 2003), eines Unternehmens (Wambugu, 2014), mehrerer Unternehmen im vergleichenden Sinne (Dastmalchian et al., 2000) oder Länder-übergreifend und vergleichend (van Muijen & Koopman, 1994) ausgerichtet. Diese erheben die Werte, verkörpert durch organisationale Handlungen, aus Sicht der beteiligten Einzelpersonen und nutzen dazu entweder psychometrische Fragebögen (Carter et al., 2012), qualitative Interview-Methoden (Sosnovskikh, 2016) oder ein Mixed-Method Design (Agbényiga, 2011).

Die Werte werden durch organisationale Handlungen jedoch nicht nur nach innen, sondern zudem auch nach außen kommuniziert. Die Selbstdarstellung auf Recruiting-Websites und den dazugehörigen Online-Stellenanzeigen (Online Job Advertisements, OJA) (Braddy et al., 2006) zeigt potentiellen Bewerbern bereits zu Beginn der ‚Candidate Journey‘ im Rahmen der Informationssuche (Verhoeven, 2016), worauf sich diese einlassen. Bewerber haben so die Chance einen Abgleich zwischen den eigenen Werte-

Vorstellungen und den Werte-Darstellungen des Unternehmens durchzuführen (Vandenbergh, 1999).

Aktuelle Studien (Pandey & Pandey, 2019; Ponizovskiy et al., 2020) schlagen vor, Textdaten (u.a. Briefe, Reports) mittels Methoden des Natural Language Processing auf die verwendete Sprache und deren Passung zu Wertemodellen zu untersuchen. Auf Basis der dabei benutzten kulturellen Begriffssammlungen (Dictionary / Vocabulary) werden Häufigkeitszählungen durchgeführt (Language Inquiry and Word Count) (Pennebaker & King, 1999), um anschließend Aussagen zu den Kulturdimensionen der Begriffssammlungen tätigen zu können. Da die obigen Analysen bisher nur auf eine jeweilige englischsprachige Begriffssammlung zurückgriffen, stellt sich die zentrale Frage, inwieweit eine deutsche Übersetzung funktional anwendbar ist, um Wertemodelle in deutschen OJA erheben zu können?

2. Theorie

Im Folgenden sollen zentrale Modelle der Unternehmenswerte und aktuelle Forschung zur Kultur-Analyse von OJA vorgestellt werden.

2.1 Unternehmenswerte

Der wohl bekannteste Ansatz für einen weltweiten Wertevergleich mit Unternehmensbezug ist das Modell von Hofstede (Hofstede, 1980, 1991). Er unterschied dabei in 3 bimodale Wertekategorien Individualistisch vs. Kollektivistisch, Maskulin vs. Feminin, Einschränkung vs. Freizügigkeit und 3 graduelle Wertekategorien der Machtdistanz, Unsicherheitsvermeidung und Langzeitorientierung. Gegen diese 6 Wertekategorien wurde z.B. argumentiert, dass es sich nur um ein westliches Wertesystem, der um Hofstede rangierenden Wissenschaftler handelt und andere Kulturen nicht hinreichend berücksichtigt werden (Baskerville, 2003). Ebenfalls wird methodisch kritisiert, dass die Modellentwicklung auf Basis eines Unternehmens (IBM) erfolgte, dann aber auf globaler Ebene angewendet wird (Venaik & Brewer, 2013).

Ein weiterer Ansatz der vergleichenden Kulturanalyse ist das Universal Value Structure Modell (Schwartz, 1992). Dieses gliedert sich in vier übergeordneten Kategorien, welche insgesamt 10 Wertedimensionen beinhalten: „Conversation“ („Tradition“, „Conformity“, „Security“), „Self-Enhancement“ („Power“, „Achievement“), „Openness to Change“ („Hedonism“, „Stimulation“, „Self-Direction“) und „Self-Transcendence“ („Universalism“, „Benevolence“). Durch zahlreiche Studien mit 200 Samples in über 60 Ländern konnte sowohl das Modell, als auch das dazugehörige psychometrische Messinstrument bestätigt werden (Bardi & Schwartz, 2003; Schwartz, 1992, 1994; Schwartz et al., 2001; Schwartz & Bardi, 2001; Schwartz & Sagiv, 1995). Dieses wurde im Unternehmenskontext angewendet (Rosario et al., 2014), wobei die obige Kritik des westlichen Wertesystems an diesem generalistischen Ansatz nicht greift (Schwartz, 2006). Gleichwohl aber wurde die dahinterliegende Annahme, dass es eine „fictitious middle or mean person“ gäbe, deren Wertekanon man statistisch erfassen könnte, kritisiert (Witte et al., 2020).

Ein im Rahmen der vergleichenden Unternehmens-Kulturanalyse entwickelter Ansatz ist dagegen das „Competing Value Framework“ (Quinn & Rohrbaugh, 1983). In diesem Modell werden 4 verschiedene Unternehmenswerte-Archetypen gegenübergestellt: „Adhocracy“ (Unternehmertum, Kreativität, Wachstum), „Market“ (Ergebnis-, Wettbewerb, Kundenorientierung), „Hierarchy“ (Kontrolle, Effizienz, Effektivität) und „Clan“ (Personen, Kommunikation, Moral). In einer Studie wird die Analyse im Kontext von Colleges

und Universitäten durchgeführt (K. S. Cameron & Freeman, 1991) sowie ein vergleichender Review empirischer Forschung erstellt (Zammuto & Krakower, 1991). Spezifische Studien setzen das Werte-Framework in den Zusammenhang mit der Qualität des Arbeitslebens (Goodman et al., 2001). Größter Kritikpunkt, den die Entwickler des Modells nach langjähriger Forschung zusammengestellt haben, ist die Reliabilität und Validität der verschiedenen Messinstrumente (K. Cameron et al., 2022), wenngleich mehrere Studien für die unterschiedlichen psychometrischen Verfahren positive Kennzahlen ausweisen - für eine Übersicht (vgl. Choi et al., 2010).

2.2 Kultur-Analyse von Textdaten

Zum aktuellen Zeitpunkt sind vier Studien besonders hervorstechend und behandeln die Erstellung und Anwendung von Wortsammlungen zu Analyse von Kultur-Wertemodellen in Textdaten.

Die erste Studie (Christen et al., 2016) verfolgt einen grundsätzlichen Aufschlag für das Forschungsgebiet im Sinne des lexikalischen Ansatzes (Allport & Odbert, 1936). Dabei extrahierten die Autoren 460 Worte mit Wertebezug und zu jedem Wort wurde eine untergeordnete Wortliste über semantische Nähe, Synonyme und Thesauri kreiert. Insgesamt standen so 3749 englische und 4775 deutsche Worte zur Verfügung. Über eine zweistufige Iteration, als Kombination aus „Machine Learning“ und „Expert Rating“, wurden abschließend 78 sprachübergreifende Werte-Gruppen mit 247 Werte-Worten identifiziert. Auf Basis einer empirischen Studie wurde damit eine Werte-Landkarte im Sinne einer Ähnlichkeit der Begriffsverwendung erzeugt. Die finale Begriffssammlung enthält viele Eigenschaften oder Zustände wie „Schönheit“ oder „Autonomie“.

Die zweite Studie (Chapman et al., 2018) verwendet ebenfalls einen lexikalischen Ansatz. Das Forschungsteam extrahierte 422 Adjektive, welche die Werte einer Organisationskultur beschreiben können. Diese Worte wurden in einer Pilotstudie von 60 Teilnehmer*Innen in Bezug zur Verwendung im Rahmen von Gesprächen über eine Organisation auf einer 5er-Likert-Skala bewertet. Die anschließende Hauptstudie mit 343 Teilnehmer*Innen erforderte die Einwertung der auf 437 erweiterten Liste in Bezug zur Beschreibung des eigenen Arbeitgebers mit einer 7er Likert-Skala. Eine parallele Principal-Komponenten-Analyse mit rotierten Faktoren ergab 9 Dimensionen zur Beschreibung einer Organisation: Innovativ, Dominant, Pace (Geschwindigkeit), Friendly (Freundlichkeit), Prestigious (Bekanntheit), Trendy, Corporate Social Responsibility, Traditional, und Diverse. Die Begriffe mit den größten (positiven & negativen) Faktorladungen wurden als Lexical Organizational Culture Scale (LOCS) publiziert, jedoch noch nicht für die Analyse von Stellenanzeigen angewendet. Die finale Wortliste wurde nicht auf Textdaten angewendet, sondern die Ergebnisse der Skalen mit den Ergebnissen der Organizational Personality Scale (OPS), der Symbolic Scale (SbS), und der Corporate Character Scale (CCS) verglichen. Es besteht kein Bezug zu den oben vorgestellten Werte-Modellen.

Der dritte Ansatz (Pandey & Pandey, 2019) verfolgt die Zusammenstellung einer Wortliste auf Basis einer deduktiven und induktiven Prozedur, ohne jedoch den oben beschriebenen lexikalischen Ansatz zu berücksichtigen. Deduktiv wurden bereits entwickelte Fragebögen zur Messung der organisationalen Werte verwendet. Aus dem Denison Organizational Culture Survey (Denison & Neale, 2000), dem Organizational Culture Assessment Instrument (Cameron & Quinn, 1999), und dem Organizational Culture Profile (O'Reilly et al., 1991) wurden zunächst 6 Faktoren extrahiert, die in den meisten der von

den Autoren zitierten empirischen Studien verwendet wurden: Competitiveness (Wettbewerbsfähigkeit), Control- and coordination-oriented (Kontrolle), Customer-oriented (Kundenorientierung), Human-resource-oriented (Personenbezogen), Innovation- and learning-oriented (Innovationsbezogen), Team-oriented (Teambezogen). Anschließend wurden sowohl Einzelworte als auch Wortphrasen (Bi- / Tri-Gramme) aus den Fragebögen extrahiert und ergänzt durch Synonymsuche und Thesauri. Induktiv wählten die Autoren die 50 ‚most admired‘, 60 ‚industry champions‘ und 58 ‚contender‘ der Fortune500 Unternehmen und extrahierten Nomen und Verben von Briefen an die Anteilseigner aus dem Jahre 2012. Mit Ihrer Wortliste untersuchten Sie 300 Unternehmen der Fortune500 anhand von Abschlussberichten und Briefen an die Anteilseigner aus dem Jahr 2014. Durch die Nutzung des Organizational Culture Assessment Instruments, welches auf dem Competing Value Framework basiert, ist ein klarer Bezug zu einem der obigen Modelle gegeben.

Der vierte Ansatz (Ponizovskiy et al., 2020) hat einen eindeutigen theoretischen Bezug und greift auf das Universal Value Structure Model und die dazugehörigen Fragebögen (271 Worte mit Wertebezug), zusätzlich aber auf die erste Studie von Christen und Kollegen (2016) sowie eine Liste der 20.000 meist genutzten Unigramme (Brants & Franz, 2006) zurück. Daraus extrahierten die Autoren mittels Expertenrating 2570 Worte für den Pretest. Diesen führten die Autoren mit 4 unterschiedlichen Text-Datenquellen (COCA, Personal Blogs, Political Blogs, Essays) und dabei jeweils der Hälfte der zur Verfügung stehenden Datenpunkte durch. Nach erfolgter Dimensionsreduktion (EFA, PA) wiesen Sie 1068 Worte in den 10 Dimensionen des Universal Value Structure Model aus und analysierten die zweite Hälfte der 4 Datensätze sowie einen Facebook Status Datensatz zur Modellvalidierung.

Im Folgenden werden nur die Ansätze zwei bis vier für die weitere Analyse zusammengestellt, da der erste Ansatz mit seinen 78 Wertekategorien und Eigenschafts- bzw. Zustandsworten keine Möglichkeit bietet, Textdaten auszuwerten. Gleichzeitig ist er Forschungsgrundlage für das dritte Modell gewesen. Tabelle 1 stellt die wesentlichen Charakteristiken der einzelnen Ansätze nach der Anzahl der Domains, Kategorien und maximal verwendete N-Gramme gegenüber.

Tabelle 1 - Gegenüberstellung der Ansätze zur Messung von Kultur mittels Wortlisten

Chapman et al 2018	Pandey & Pandey 2019	Ponizovskiy et al 2020 (10)
Domains (9): Innovativ, Trendy, Pace, Prestigious, Dominant, Traditional, Friendly, Corporate Social Responsibility, Diverse	Domains (6): Innovation- and learning-oriented, Competitiveness, Customer-oriented, Control- and coordination-oriented, Human-resource-oriented, Team-oriented	Domains (10): Self-Direction, Hedonism, Stimulation, Achievement, Power, Tradition, Conformity, Security, Benevolence, Universalism
Kategorien (15): 6 positive & negative = 12 3 rein positive	Kategorien (6): 6 rein positive	Kategorien (14): 10 rein positive 4 aggregierte
Max. n-gram: 4	Max. n-gram: 5	Max. n-gram: 3

Forschungsfragen

Ausgehend von der eingangs aufgestellten Forschungsfrage, ob deutsche Übersetzung der englischsprachigen Ansätze für die Analyse von deutschen OJAs funktional anwendbar sind, können weitere Subfragen definiert werden:

1. Werden alle Kategorien der einzelnen Ansätze identifiziert?
2. Welcher Ansatz funktioniert im Kontext von OJAs am besten?
3. Besteht ein Optimierungspotential für den best-funktionierenden Ansatz?

3. Forschungsmethode

3.1 Datenset

Um die Forschungsfragen untersuchen zu können, ist es notwendig einen großen Datensatz von deutschsprachigen OJA zur Verfügung zu haben. Während OJA im Internet auf den Recruiting-Webseiten der Unternehmen oder auf Jobportalen frei verfügbar sind, müssen sie für die Analyse erst extrahiert werden. Eine Form für die Informationsextraktion ist das Web Scraping. Es kann verwendet werden, um Daten und Informationen aus dem Internet abzurufen (Saurkar et al., 2018), dabei sollen semi-strukturierte Elemente mit Hilfe der Extrahierung in besser strukturierte Datensätze gespeichert werden (Vargiu & Urru, 2012). Mehrere Forscher haben Web Scraping-Techniken bereits für OJA angewandt (De Mauro et al., 2018; Gurcan & Cagiltay, 2019).

Um eine große Anzahl von OJAs zu extrahieren, wurde ein Browser-Plug-in von webscraper.io verwendet. Die zu analysierenden Daten wurden von einem großen deutschen Jobportal im Zeitraum von Januar bis Mai 2022 gescraped (N1 = 151k). Diese OJAs verteilen sich auf eine Menge an Unternehmen (N2=29.039), mit minimal 1 und maximal 4118 OJAs. Der Mittelwert beträgt 5,19 (SD 35,88) und der Median 2 OJA pro Unternehmen.

3.2 Analysemethode

Bevor der Datensatz analysiert werden kann, muss dieser aufbereitet werden. Dazu wurden folgende Pre-processing Schritte durchgeführt. Zuerst wurden alle Elemente der Stellenanzeige (Titel, Beschreibung des Unternehmens/ der Stelle/ des Angebots) zu einem Text verbunden, da sie beim scrapen als verschiedene Datenfelder extrahiert wurden. Im zweiten Schritt wurden alle Sonderzeichen aus dem Textfeld entfernt und drittens wurden alle Großbuchstaben in Kleinbuchstaben umgewandelt. Viertens wurde der Datensatz tokenisiert, das bedeutet, dass der Text in Einzelwörter zerlegt wird. Im letzten Schritt werden n-Gramme in Abhängigkeit der Notwendigkeit vom jeweiligen Dictionary gebildet.

Um die Unternehmenswerte aus den OJAs extrahieren zu können, wurde das Language Inquiry and Word Count (LIWC) Verfahren (Pennebaker & King, 1999) angewendet. Dabei werden die Worte von drei der vier oben dargestellten Ansätze ins Deutsche übersetzt und die Häufigkeit über die OJAs gezählt. Anschließend werden die Ergebnisse zu allen OJAs über die Unternehmen gruppiert und Mittelwerte über alle Wertedimensionen je Unternehmen gebildet. Damit liegen Mittelwerte über die in den OJAs dargestellte Unternehmenswerte auf Basis beschreibender Adjektive vor. Durch dieses Verfahren hat die Anzahl der OJAs pro Unternehmen keinen Einfluss auf das Ergebnis.

4. Ergebnisse

Die Tabelle 2 stellt die deskriptiven Daten der drei Ansätze gegenüber. Alle drei Wörter-

bücher kommen aus dem Englischen und wurden ins Deutsche übersetzt. Die unterschiedliche Anzahl der Begriffe zwischen den Sprachen liegt daran, dass es für einen Begriff im Deutschen z.B. zwei verschiedene Begriffe im Englischen geben kann. Bei der Anzahl der Wörter lässt sich erkennen, dass es bei Ponizovskiy und Kollegen (2020) insgesamt 918 deutsche Wörter gibt, wohingegen Chapman und Kollegen (2018) 123 und Pandey und Kollegen (2019) 160 deutsche Wörter haben. Durch die hohe Anzahl an Wörtern bei Ponizovskiy und Kollegen ist der Mittelwert pro Dimension deutlich höher als bei den anderen beiden Ansätzen. Beim Minimum fällt auf, dass bei Chapman und Kollegen eine Dimension nur mit einem Wort besetzt ist. Bereits aus dieser Gegenüberstellung ist es wahrscheinlich, dass der Ansatz von Ponizovskiy und Kollegen am besten funktionieren wird.

Tabelle 2 Deskriptive Beschreibung der drei Ansätze

	Chapman et al	Pandey et al	Ponizovskiy et al (Einzeldimension)
Kategorien / Dimensionen	15	6	10
Worte im Dictionary (EN)	135	172	1.068
Worte im Dictionary (DE)	123	160	918
Mittelwert	8,20	26,67	91,80
Standardabweichung	7,02	2,98	17,16
Minimum	1	22	70
Maximum	20	31	124

Die durchgeführten Analysen zeigen, dass alle drei Ansätze funktional für die Verwendung bei OJA sind. Der Ansatz von Chapman und Kollegen (2018) hat insgesamt 9 Dimensionen mit 15 Kategorien, da einige Dimensionen sowohl eine positive wie auch eine negative Ausprägung besitzen. 13 Kategorien konnten in unterschiedlicher Deutlichkeit identifiziert werden (vgl. Tabelle 3). Betrachtet man die maximale Anzahl der durchschnittlichen Worte pro Unternehmen (S1-Max), ist die Dimension „Dominant“ mit 13 Wörtern am stärksten. 9 Dimensionen verwenden durchschnittlich drei bis fünf Wörter. Ein Blick auf die Summe der durchschnittlich verwendeten Wörter über alle Unternehmen (S2-Summe) zeigt, dass die Kategorie „Trendy“ (9.047,20) und „Pace“ (4.821,39) im Datensatz am häufigsten verwendet wurden. Für die Kategorien „Innovative“, „Dominant“, „Friendly“, „Traditional“ und „CSR“ liegen Werte über 1.000 vor. Weniger häufig sind die Kategorien „Dominant negative“, „Friendly negative“, „Prestigious“ und „Diverse“. „Diverse negative“ und „Traditional negative“ sind kaum vorhanden und „Pace negative“ und „Trendy negative“ gar nicht. Die Auswertung der Anzahl der Unternehmen, die die Wörter der jeweiligen Kategorie verwendet haben (S3-Anzahl), zeigt eine unterschiedlich starke Nutzung. Bildet man das Verhältnis zwischen der Summe der durchschnittlich verwendeten Wörter über alle Unternehmen zu der Anzahl der Unternehmen (S4-Verhältnis), ist die stärkste Dimension „Dominant“, das heißt im Durchschnitt wurde 0,801 Wörter pro Stellenanzeige je Unternehmen verwendet. Neben den Kategorien „Pace negative“ und „Diverse negative“, welche gar nicht in den OJA verwendet werden, ist die kleinste Kategorie „Friendly negative“, welche durchschnittlich 0,480 Wörter pro Stellenanzeige je Unternehmen genutzt wird. S5-Prozent stellt das prozentuale Verhältnis zwischen den Unternehmen (S3-Anzahl), welche die Kategorie verwendet haben, zu allen Unternehmen im Datensatz dar. Die Dimension „Trendy“ wird von ca. 41,3 Prozent der Unternehmen verwendet. Die Kategorien „Pace negative“ und „Divers negative“ werden gar nicht verwendet und „Trendy negative“ ist die Verwendung ebenfalls sehr gering.

Tabelle 3 - Chapman und Kollegen (2018)

Dimension	S1-Max	S2-Summe	S3-Anzahl	S4-Verhältnis	S5-Prozent
Innovative	6,60	4.399,46	6.109	0,720	0,210
Dominant	13,00	2.824,53	3.527	0,801	0,121
Dominant_n	4,00	144,11	229	0,629	0,008
Pace	5,00	4.821,39	6.379	0,756	0,220
Pace_n	0,00	0,00	0	0,000	0,000
Friendly	4,00	2.502,98	3.956	0,633	0,136
Friendly_n	3,00	499,83	1.041	0,480	0,036
Prestigious	3,00	526,31	847	0,621	0,029
Trendy	4,00	9.047,20	12.000	0,754	0,413
Trendy_n	1,00	1,50	2	0,750	0,000
CSR	3,00	1.523,41	2.553	0,597	0,088
Traditional	4,17	1.234,08	1.872	0,659	0,064
Traditional_n	2,00	43,83	65	0,674	0,002
Diverse	2,00	581,73	1.001	0,581	0,034
Diverse_n	0,00	0,00	0	0,000	0,000

Anmerkung: S1-Max = Maximale Anzahl der durchschnittlichen Wörter je Unternehmen; S2-Summe = Summe der durchschnittlich verwendeten Wörter über alle Unternehmen hinweg; S3-Anzahl = Anzahl der Unternehmen, die Wörter der jeweiligen Kategorie verwendet haben, S4-Verhältnis = S2-Summe / S3-Anzahl; S5-Prozent = Prozent der Unternehmen, die Wörter der jeweiligen Kategorie verwendet haben (Verhältnis Spalte 3 zu Gesamtunternehmen)

Tabelle 4 ist analog zu Tabelle 3 aufgebaut und stellt den Ansatz von Pandey und Kollegen (2019) dar. Die Autoren haben bei ihrem Ansatz insgesamt sechs Dimensionen identifiziert. Die Dimension „Competitiveness“ hat die stärksten Ausprägungen von allen Dimensionen. Die maximale Anzahl der durchschnittlichen Wörter (S1-Max) beträgt 15,0 bei dieser Dimension. Den geringsten Wert hat die Dimension „HR oriented“ mit maximal durchschnittlich 4,0 Wörtern. Die anderen vier Dimensionen liegen zwischen 7,5 und 9,0 Wörtern. Bei S2-Summe, S3-Anzahl und S4-Verhältnis hat ebenfalls die Dimension „Competitiveness“ die höchsten Werte und die Dimension „HR oriented“ die geringsten. In Abgrenzung zu Tabelle 2 ist in Spalte 4 ein Wert über 1 vorhanden.

Tabelle 4 - Pandey und Kollegen (2019)

Dimension	S1-Max	S2-Summe	S3-Anzahl	S4-Verhältnis	S5-Prozent
Competitiveness	15	47.561,6	24.114	1,972	0,830
ControlCoordination	8,8	5.786,6	7.884	0,734	0,271
Customer	14	23.430,5	14.981	1,564	0,516
HR	4	1.376,7	1.841	0,748	0,063
InnovationLearning	9	11.464,6	12.182	0,941	0,420
Team	7,5	12.738,8	14.100	0,903	0,486

Anmerkung: für die Spaltenbeschriftung siehe Anmerkung Tabelle 2

Der letzte Ansatz von Ponizovskiy und Kollegen (2020) ist der umfangreichste basierend auf den verwendeten Wörtern. Dieser weißt zum einen die aggregierten Dimensionen sowie die Einzeldimensionen auf, im Sinne der Vergleichbarkeit konzentrieren wir uns auf die Einzeldimensionen.

In Tabelle 5 ist zu erkennen, dass die Dimension „Achievement“ über alle Kennzahlen hinweg am stärksten ist, gefolgt von der Dimension „Self-Direction“. Hervorzuheben in S4-Verhältnis ist, dass die Dimension „Achievement“ durchschnittlich 4,124 Wörter pro OJA je Unternehmen verwendet wird. Bei der Dimension „Self-Direction“ ist es im Vergleich nur die Hälfte. Insgesamt haben 9 der 10 Dimensionen einen Wert über eins, was einer kontinuierlichen Nutzung der Dimension im Rahmen der einzelnen Stellenanzeigen entspricht, nur „Tradition“ liegt knapp darunter.

Tabelle 5 - Ponizovskiy und Kollegen (2020)

Dimension	S1-Max	S2-Summe	S3-Anzahl	S4-Verhältnis	S5-Prozent
Security	11,00	15.221,7	15.162	1,004	0,522
Conformity	17,80	74.923,1	27.189	2,756	0,936
Tradition	8,00	5.368,7	6.172	0,870	0,213
Benevolence	22,00	59.394,4	25.314	2,346	0,872
Universalism	19,00	46.023,7	23.685	1,943	0,816
SelfDirection	22,00	78.651,5	26.788	2,936	0,922
Stimulation	12,00	26.440,6	19.041	1,389	0,656
Hedonism	11,00	19.266,0	16.328	1,180	0,562
Achievement	23,50	115.802,9	28.077	4,124	0,967
Power	12,00	42.398,3	23.272	1,822	0,801

Anmerkung: ¹ aggregierte Dimensionen; für die Spaltenbeschriftung siehe Anmerkung Tabelle 2

In Tabelle 6 werden alle drei Ansätze gegenübergestellt. Hier ist zu erkennen, dass Ponizovskiy und Kollegen bei S4-Verhältnis einen Mittelwert von 2,037 Wörter pro Stellenanzeige je Unternehmen und die Spannweite zwischen 0,870 und 4,124 liegt. Verglichen zu Chapman und Kollegen (0,577 Wörter) und Pandey und Kollegen (1,144 Wörter) ist der Wert bei Ponizovskiy und Kollegen knapp zwei- bzw. dreimal so groß. Dies spiegelt sich auch bei S5-Prozent wieder, wo knapp 73 Prozent der Unternehmen Wörter von Ponizovskiy und Kollegen verwendet haben. Somit kann auch im Vergleich der Ergebnisse festgehalten werden, dass der Ansatz von Ponizovskiy und Kollegen tatsächlich am besten funktioniert.

Tabelle 6 - Vergleich der drei Ansätze

	Chapman et al		Pandey et al		Ponizovskiy et al	
	S4-Verhältnis	S5-Prozent	S4-Verhältnis	S5-Prozent	S4-Verhältnis	S5-Prozent
Mittelwert	0,577	0,091	1,144	0,431	2,037	0,727
Standardabweichung	0,248	0,116	0,507	0,257	1,023	0,239
Minimum	0,000	0,000	0,734	0,063	0,870	0,213
Maximum	0,801	0,413	1,972	0,830	4,124	0,967

In Tabelle 7 werden die Wörter aus dem Wörterbuch je Ansatz mit den gefundenen Wörtern in den OJAs dargestellt. Bei Chapman und Kollegen werden 110 der 126 Wörter (89,4%) gefunden. Bei Pandey und Kollegen liegt der Wert nur bei 56,9 Prozent. Der Wert bei Ponizovskiy und Kollegen liegt zwar nur bei 79,3 Prozent und ist geringer als bei Chapman und Kollegen, aber dies entspricht 728 Wörter. Folgend dem Ergebnis, dass der Ansatz von Ponizovskiy und Kollegen der aktuell performanteste ist, ist fraglich, wie viele Worte aus den anderen Ansätzen im Sinne einer Optimierung ggf. überführt werden könnten. Dies entspricht bei Chapman und Kollegen 76 und bei Pandey und Kollegen 149 Wörtern, welche nicht im Wörterbuch von Ponizovskiy und Kollegen vorhanden sind. Daher empfiehlt diese Forschungsarbeit ein zusammengeführtes Wörterbuch für den deutschsprachigen Raum mit insgesamt 1143 Begriffen.

Tabelle 7 - Übersicht der gefundenen Wörter in den OJAs

	Chapman et al	Pandey et al	Ponizovskiy et al
Worte im Dictionary (DE)	123	160	918
Worte in OJA identifiziert	110	91	784
Worte in OJA / Worte im Dictionary DE	89,4%	56,9%	85,4
Wörter bei Ponizovskiy et al hinzufügen	76	149	-

Diskussion

OJAs sind eine Form der Kommunikation mit potenziellen Bewerbern. Aus diesem Grund ist es wichtig, dass in dieser Kommunikationsform bereits die verkörperten Unternehmenswerte eingebracht werden. Diese können einen wesentlichen Einfluss auf die Entscheidung des Bewerbers für oder gegen ein Unternehmen haben. Damit Unternehmen aber in der Lage sind, ihre externe Kommunikation von Werten zu analysieren ist eine funktionale, deutschsprachige Wortliste nötig, deren Entwicklung Hauptbestandteil dieses Beitrages war.

Die Autoren haben dabei eine zentrale Forschungsfrage und drei Subfragen untersucht. Die zentrale Frage, inwieweit Unternehmenswerte durch Wortlisten in OJAs identifiziert werden können, kann positiv bejaht werden. Tabelle 6 fasst die Funktionalität aller ins Deutsche übersetzten und angewendeten Ansätze entsprechend zusammen. Bezüglich der ersten Subfrage, ob alle Kategorien der unterschiedlichen Ansätze identifizierbar sind, kann festgehalten werden, dass nur zwei Kategorien aus dem Ansatz von Chapman und Kollegen nicht identifiziert wurden. Alle anderen Dimensionen der drei Ansätze konnten aus den OJA extrahiert werden. Daher konnte gezeigt werden, dass durch Wortlisten Wertemodell in Stellenanzeigen wiedergespiegelt werden. Die Subfrage zwei bezüglich des am besten funktionierenden Ansatz kann auf Basis deren Güte beantwortet werden. Diese variiert deutlich, weil a) die Wortverteilung über die Dimensionen sehr unterschiedlich ist (Tabelle 2). Bei Ponizovskiy und Kollegen liegt die Spannweite zwischen 70 und 124 je Dimension, wohingegen bei Pandey und Kollegen die Spannweite zwischen 22 und 31 liegt und bei Chapman und Kollegen zwischen 1 und 20. Gleichzeitig ist b) bei den Auswertungen zu sehen (Tabelle 6), dass der Ansatz von Ponizovskiy und Kollegen im S4-Verhältnis deutlich besser ist als die anderen beiden. Durchschnittlich werden bei diesem Ansatz 2,037 Wörter pro Stellenanzeige je Unternehmen verwendet, was einer kontinuierlichen Nutzung der Dimension über alle publizierten Stellenanzeigen verdeutlicht. Somit ist für weitere Forschungsarbeiten sowie angewandte Studien mit Unternehmen nur der Ansatz von Ponizovskiy empfehlenswert. Die dritte Subfrage, den besten Ansatz (hier also den von Ponizovskiy und Kollegen) ggf. verbessern zu können, kann mit Verweis auf Tabelle 7 ebenfalls bejaht werden. Die Wörter von Chapman und Kollegen sowie von Pandey und Kollegen, welche nicht im Wörterbuch von Ponizovskiy und Kollegen vorhanden waren, sollten dem Ansatz auf Basis eines Expertenratings unter Verwendung der Kategorien hinzugefügt werden. Dabei ist zu berücksichtigen, dass ggf. einzelne Begriffe auf mehrere Dimensionen laden könnten, so dass eine gewichtete Wortliste detaillierter erscheint, als eine explizite Zuordnung zu einer Dimension.

Auf Basis der Ergebnisse dieser Arbeit können Unternehmen nun das Werteprofil der externen Online-Kommunikation ermitteln und dieses mit den internen Wertvorstellungen bzw. dem intern erhobenen realen Werteprofil abgleichen. So haben sie die Chance, den Person-Organization-Fit zu verbessern, um eine Fluktuation durch inkongruente Kommunikation der Wertvorstellung zu verhindern.

Literaturverzeichnis

- Agbényiga, D. L. (2011). Organizational culture influence on service delivery: A mixed methods design in a child welfare setting. *Children and Youth Services Review, 33*(10), 1767–1778.
- Allport, G., & Odbert, H. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs, 47*(1), i–171. <https://doi.org/10.1037/h0093360>

- Bardi, A., & Schwartz, S. H. (2003). Values and Behavior: Strength and Structure of Relations. *Personality and Psychological Bulletin*, *29*(10), 1207–1220.
- Baskerville, R. F. (2003). Hofstede never studied culture. *Accounting, Organizations and Society*, *28*, 1–14.
- Braddy, P. W., Meade, A. W., & Kroustalis, C. M. (2006). Organizational recruitment website effects on viewers' perceptions of organizational culture. *Journal of Business and Psychology*, *20*(4), 525–543.
- Brants, T., & Franz, A. (2006). *Web 1T 5-gram version 1*.
- Cameron, K., Quinn, R., DeGraff, J., & Thakor, A. (2022). *Competing Values Leadership*. Edward Elgar.
- Cameron, K. S., & Freeman, S. J. (1991). Cultural Congruence, Strength, and Type. *Research in Organizational Change and Development*, *5*, 23–58.
- Carter, N. T., Kotrba, L. M., Diab, D. L., Lin, B. C., Pui, S. Y., Lake, C. J., Gillespie, M. A., Zickar, M. J., & Chao, A. (2012). A Comparison of a Subjective and Statistical Method for Establishing Score Comparability in an Organizational Culture Survey. *Journal of Business Psychology*, *27*, 451–466.
- Chapman, D. S., Reeves, P., & Chapin, M. (2018). A Lexical Approach to Identifying Dimensions of Organizational Culture. *Frontiers in Psychology*, *9*(876), 1–16. <https://doi.org/doi:10.3389/fpsyg.2018.00876>
- Choi, Y., Seo, M., Scott, D., & Martin, J. J. (2010). Validation of the organizational culture assessment instrument: An application of the Korean version. *Journal of Sport Management*, *24*(2), 169–189.
- Christen, M., Narvaez, D., Tanner, C., & Ott, T. (2016). Using thesauruses as a heuristics for mapping values. *Cognitive Systems Research*, *40*, 59–74. <https://doi.org/10.1016/j.cogsys.2016.02.003>
- Dastmalchian, A., Lee, S., & Ng, I. (2000). The interplay between organizational and national cultures: A comparison of organizational practices in Canada and South Korea using the Competing Values Framework. *International Journal of Human Resource Management*, *11*(2), 388–412.
- De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, *54*(5), 807–817. <https://doi.org/10.1016/j.ipm.2017.05.004>
- Goodman, E. A., Zammuto, R. F., & Gifford, B. D. (2001). The competing values framework: Understanding the impact of organizational culture on the quality of work life. *Organization Development Journal*, *19*(3), 58–68.
- Gurcan, F., & Cagiltay, N. E. (2019). Big Data Software Engineering: Analysis of Knowledge Domains and Skill Sets Using LDA-Based Topic Modeling. *IEEE Access*, *7*, 82541–82552. <https://doi.org/10.1109/ACCESS.2019.2924075>
- Hofstede, G. (1980). *Culture's Consequences: International Differences in Work-Related Values*. Sage.
- Hofstede, G. (1991). *Cultures and Organizations*. McGraw-Hill.
- Mead, G. H. (1934). *Mind, Self, and Society*. University of Chicago Press.
- Ogbonna, E., & Wilkinson, B. (2003). The False Promise of Organizational Culture Change: A Case Study of Middle Managers in Grocery Retailing. *Journal of Management Studies*, *40*(5), 1151–1178.
- Pandey, S., & Pandey, S. K. (2019). Applying natural language processing capabilities in computerized textual analysis to measure organizational culture. *Organizational Research Methods*, *22*(3), 765–797.
- Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality Social Psychology*, *77*(6), 1296–1312.
- Ponizovskiy, V., Ardag, M., Grigoryan, L., Boyd, R. L., Dobewall, H., & Holtz, P. (2020). Development and Validation of the Personal Values Dictionary: A Theory-Driven Tool for Investigating References to Basic Human Values in Text. *European Journal of Personality*, *34*, 885–902.
- Quinn, R. E., & Rohrbaugh, J. (1983). A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. *Management Science*, *29*(3), 363–377.
- Rosario, G.-R. M., Carmen, D.-F. M., & Biagio, S. (2014). Values and corporate social initiative: An approach through Schwartz theory. *International Journal of Business and Society*, *15*(1), 19–48.
- Saurkar, A. V., Pathare, K. G., & Gode, S. A. (2018). An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*, *4*(4), 363–367.
- Schein, E. H. (1985). *Organizational Culture and Leadership*.
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Elsevier.
- Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues*, *50*(4), 19–45.
- Schwartz, S. H. (2006). Basic human values: Theory, measurement, and applications. *Revue Française de Sociologie*, *47*, 249–288.
- Schwartz, S. H., & Bardi, A. (2001). Value Hierarchies Across Cultures: Taking a Similarities Perspective. *Journal of Cross-Cultural Psychology*, *32*, 268–290.

- Schwartz, S. H., Melech, G., Lehmann, A., Burgess, S., & Harris, M. (2001). Extending the cross-cultural validity of the theory of basic human values with a different method of measurement. *Journal of Cross-Cultural Psychology, 32*, 519–542.
- Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology, 26*(January), 92–116.
- Sosnovskikh, S. (2016). Toyota Motor Corporation: Organizational Culture. *Philosophy Study, 6*(7), 442–454.
- van Muijen, J. J., & Koopman, P. L. (1994). The Influence of National Culture on Organizational Culture: A Comparative Study Between 10 Countries. *European Work & Organizational Psychologist, 4*(4), 367–380.
- Vandenberghe, C. (1999). Organizational culture, person-culture, and turnover: A replication in the health care industry. *Journal of Organizational Behavior, 20*, 175–184.
- Vargiu, E., & Urru, M. (2012). Exploiting web scraping in a collaborative filtering- based approach to web advertising. *Artificial Intelligence Research, 2*(1), p44. <https://doi.org/10.5430/air.v2n1p44>
- Venaik, S., & Brewer, P. (2013). Critical issues in the Hofstede and GLOBE national culture models. *International Marketing Review, 30*(5), 469–482.
- Verhoeven, T. (2016). *Candidate Experience*. SpringerGabler.
- Wambugu, L. W. (2014). Effects of Organizational Culture on Employee Performance (Case Study of Wart-sila—Kipevu li Power Plant). *European Journal of Business and Management, 6*(32), 80–92.
- Weber, M. (1930). *The Protestant Ethic and the Spirit of Capitalism*. Scribners.
- Witte, E. H., Stanciu, A., & Boehnke, K. (2020). A New Empirical Approach to Intercultural Comparisons of Value Preferences Based on Schwartz's Theory. *Frontiers in Psychology, 11*, 1723.
- Zammuto, R. F., & Krakower, J. Y. (1991). Quantitative and Qualitative studies of organizational culture. In R. W. Woodmann & W. A. Pasmore (Eds.), *Research in organizational change and development*. JAI Press.