

// ZEW-Kurzexpertise 20-05 · 20.04.2020

Jan Kinne/ Miriam Krüger/ David Lenz/ Georg Licht/ Peter Winker

Corona-Pandemie betrifft Unternehmen unterschiedlich

Tagesaktuelle Webseiten-Analyse
zur Reaktion von Unternehmen auf
die Corona-Pandemie in Deutschland



INHALT

ZUSAMMENFASSUNG	3
1. Datengrundlange	4
2. Schlagwortsuche und Textanalyse	4
3. Ergebnisse	6
3.1. Betroffene Unternehmen	6
3.2. Art der Betroffenheit	10
4. FAZIT	14

ZUSAMMENFASSUNG

Wir analysieren die Webseiten von circa 1,1 Millionen deutschen Unternehmen zwei Mal wöchentlich auf die Nennung des Wortes „Corona“ und entsprechender Synonyme. Anschließend werden die identifizierten Textstellen mittels KI-Textanalysemodellen ausgewertet, um den Kontext der Nennung zu erfassen. Dieses Vorgehen erlaubt es uns, die Auswirkungen der Corona-Pandemie auf Unternehmen in Deutschland zeitnah und umfassend abzuschätzen.

Seit dem ersten Abzug vom 19.03.2020 konnten wir so eine Erhöhung der betroffenen Unternehmen um circa einen Prozentpunkt pro Woche beobachten (von 10% auf 14% am 15.04.2020). Wir stellen außerdem deutliche sektorale und alters- sowie größenbezogene Unterschiede fest. Ältere und insbesondere größere Unternehmen berichten deutlich häufiger über die Pandemie. Welchen Einfluss hier die gegebenenfalls größeren Digitalkapazitäten (z.B. dezidierte Medienabteilung) und eine folglich aktuellere Webseite bei größeren Unternehmen haben, wird noch untersucht. Regional beobachten wir sowohl eine ungleiche Verteilung der aktuellen Fallzahlen (betroffene Unternehmen) auf Kreisebene, als auch ein ungleiches Wachstum. Hierbei ist ein West-Ost-Gefälle mit höheren Fallzahlen im Westen Deutschlands und ein stärkeres Wachstum der Fallzahlen im Süd-Westen der Bundesrepublik zu beobachten. Auch weisen Stadtkreise insgesamt höhere Werte auf als Landkreise. Welchen Einfluss die regionale Zusammensetzung der Unternehmenspopulation auf dieses räumliche Muster hat ist Gegenstand aktueller Analysen.

Außerdem deuten unsere Versuche mit fortschrittlichen, KI-basierten Textanalysemodellen (*istari.ai webAI Engine*) darauf hin, dass Unternehmen auf ihren Webseiten über ihre spezifischen Reaktionen auf die Corona-Pandemie berichten. So können wir aufzeigen, dass die gefundenen „Corona“-Referenzen in verschiedene Themenfelder eingeteilt werden können, deren Relevanz nach Branchen und Regionen variieren. Corona-Referenzen mit „Problem“-Kontext (Betriebs- und Ladengeschäftsschließungen, Absagen von Veranstaltungen, Kurzarbeit etc.), „Adaption“-Kontext (Anpassung Geschäftszeiten, neue Hygienevorschriften, Homeoffice etc.) und „Kein Problem“-Kontext (keine Beeinträchtigungen durch die Pandemie) treten regional gehäuft auf.

1. Datengrundlage

Als Datengrundlage dieser Untersuchung dienen 1,1 Millionen Webadressen (URLs) von Unternehmen in Deutschland, die dem Mannheimer Unternehmenspanel (Stand Ende 2019) entnommen wurden. Für jedes dieser Unternehmen sind darüber hinaus grundlegende Unternehmenseigenschaften, wie der Unternehmensstandort, die Mitarbeiteranzahl und die Branche bekannt. Aus vorangehender Forschung ist bekannt, dass etwa die Hälfte der Unternehmen in Deutschland über eine eigene Webseite verfügen [1], wobei insbesondere größere Unternehmen über eine gute bis sehr gute Abdeckung verfügen (ab 25 Mitarbeitern 84%; ab 250 Mitarbeitern 97%). Darüber hinaus gibt es branchenspezifische Unterschiede (niedrige Abdeckung z.B. bei landwirtschaftlichen Betrieben; hohe Abdeckung z.B. im Maschinenbau).

Die Webseiten der Unternehmen wurden über einen von dem Startup istari.ai entwickelten Ansatz abgefragt und heruntergeladen, pro Unternehmen maximal fünf Webpages (eine Webseite besteht meist aus mehreren (Sub-)Webpages) heruntergeladen. Die Auswahl dieser Subwebpages erfolgt dabei nicht zufällig, sondern folgt einer einfachen Heuristik: Zuerst werden Subwebpages ausgewählt, die wahrscheinlich in deutscher Sprache verfasst sind und darüber hinaus über die kürzeste URL verfügen. Letzteres führt dazu, dass meist jene Subwebpages mit allgemeineren und aktuelleren („top-level“) Informationen priorisiert heruntergeladen werden. Beispielsweise würde so „zew.de/aktuelles“ vor „zew.de/aktuelles/2019/februar/“ heruntergeladen. Insgesamt konnten 81% (1,11 Mio.) der ursprünglichen 1,36 Mio. Webseiten erfolgreich abgefragt und heruntergeladen werden. Die fehlgeschlagenen Versuche sind auf nicht mehr aktuelle und permanent oder temporär deaktivierte Webseiten zurückzuführen.

2. Schlagwortsuche und Textanalyse

Anschließend wurden die heruntergeladenen Webseiten auf Variationen des Begriffs „Corona Pandemie“ und dessen Synonyme (z.B. „SARS-CoV2“) durchsucht. Im Falle eines Treffers wurden die betroffenen Abschnitte der Webseite (HTML-Markups) markiert. Dieser einfache Ansatz erlaubt eine erste Abschätzung des Anteils von Unternehmen, die auf ihrer Webseite über die Corona-Pandemie berichten.

Darüber hinaus können die identifizierten Webseitenabschnitte mittels computerbasierter Textanalyse weiter unterteilt werden, beziehungsweise der Kontext der Nennung des „Corona“-Begriffs erfasst werden. Hierfür wurde ein von istari.ai speziell angepasstes Machine Learning basiertes Textanalysemodell verwendet. In einem ersten, arbeitsaufwändigen Schritt wurde dafür der Inhalt (z.B. „Verzögerungen in der Produktion“, „Mitarbeiterinformationen“, „Betriebsschließungen“ usw.)

einer zufälligen Auswahl von zuvor identifizierten „Corona“-Webseitenabschnitte gesichtet und manuell klassifiziert. Jeder Webseitenabschnitt wurde dabei in eine der folgenden Klassen eingeteilt:

- **Problem:** Das Unternehmen berichtet über Probleme im Zusammenhang mit der Corona-Pandemie. Dies umfasst insbesondere Schließungen von Geschäften, Absagen und Verschiebungen von Veranstaltungen, Berichte über Lieferengpässe, Kurzarbeit und ähnliches.
- **Kein Problem:** Das Unternehmen berichtet darüber, dass es von der Corona-Pandemie nicht betroffen ist, beziehungsweise, dass diese keine Auswirkungen auf das Geschäft haben.
- **Adaption:** Das Unternehmen berichtet davon, dass es sich an die neuen Umstände anpasst. Dies beinhaltet Maßnahmen wie neue Hygienevorschriften, geänderte Öffnungszeiten, Home Office und dergleichen.
- **Information:** Das Unternehmen berichtet allgemein über die Corona-Pandemie. Dies beinhaltet allgemeine Informationen wie die aktuelle Verbreitung, Krankheitssymptome, Nachrichten rund um Corona oder die Wiedergabe von behördlichen Verordnungen.
- **Unklar:** In dieser Gruppe werden Texte zusammengefasst, die sich nicht klar zuordnen lassen, da es sich entweder um Artefakte oder Falschklassifikationen handelt oder aus anderen Gründen nicht klar ersichtlich wird, was der Kontext der „Corona“-Nennung ist.

Diese „gelabelten“ Daten wurden anschließend als Trainingsdaten für ein Transfer-Lernen basiertes Sprachmodell [2] genutzt. Während dieses Trainings „lernt“ das Modell, Unterschiede zwischen den Texten zu erkennen und ist anschließend in der Lage, automatisch jeden der identifizierten Webseitenabschnitte in eine der vorher definierten Gruppen einzuteilen. Damit wird eine inhaltliche Bewertung jedes identifizierten Webseitenabschnittes möglich, was ein aktuelles Bild zur Auswirkung der Corona-Pandemie auf Unternehmen in Deutschland zulässt.

3. Ergebnisse

3.1. Betroffene Unternehmen

Abbildung 1 zeigt den Anteil an Unternehmen, die auf ihrer Webseite mindestens eine Referenz zur Corona-Pandemie aufweisen. Es ist ein ansteigender Trend zu beobachten, der sich über die vier abgebildeten Wochen abschwächt und dann auf einem Niveau von circa 14% verharret.

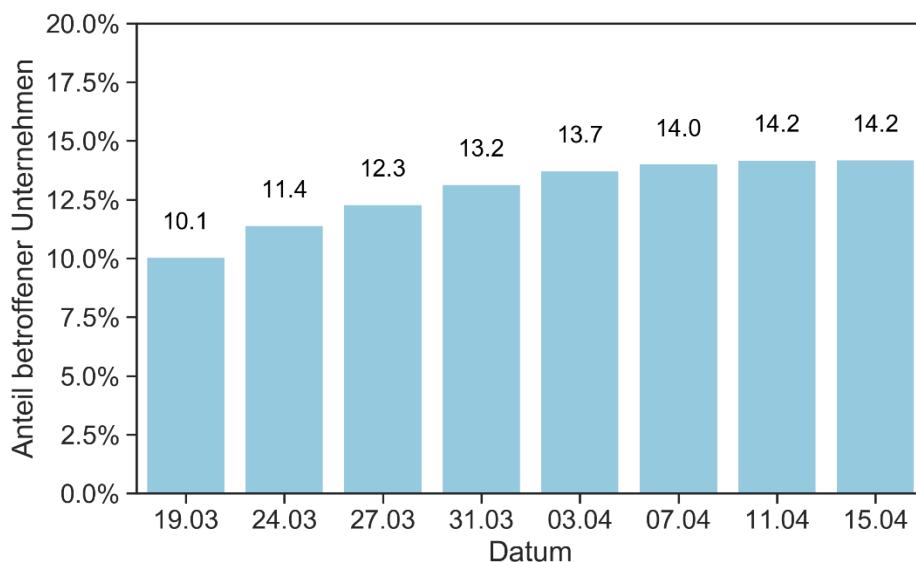


Abbildung 1: Anteil Unternehmen mit mindestens einer „Corona“-Webreferenz.

Abbildung 2 zeigt, dass insbesondere größere (und ältere) Unternehmen mit mehr als 250 Mitarbeitern über die Corona-Pandemie berichten und in den vergangenen Wochen auch den höchsten Zuwachs zu verzeichnen hatten. Inwieweit dies die tatsächliche Betroffenheit der Unternehmen oder deren größere Digitalkapazitäten (z.B. dezidierte Medienabteilung) und eine folglich aktuellere Webseite widerspiegelt, wird noch untersucht.

Corona-Pandemie betrifft Unternehmen unterschiedlich

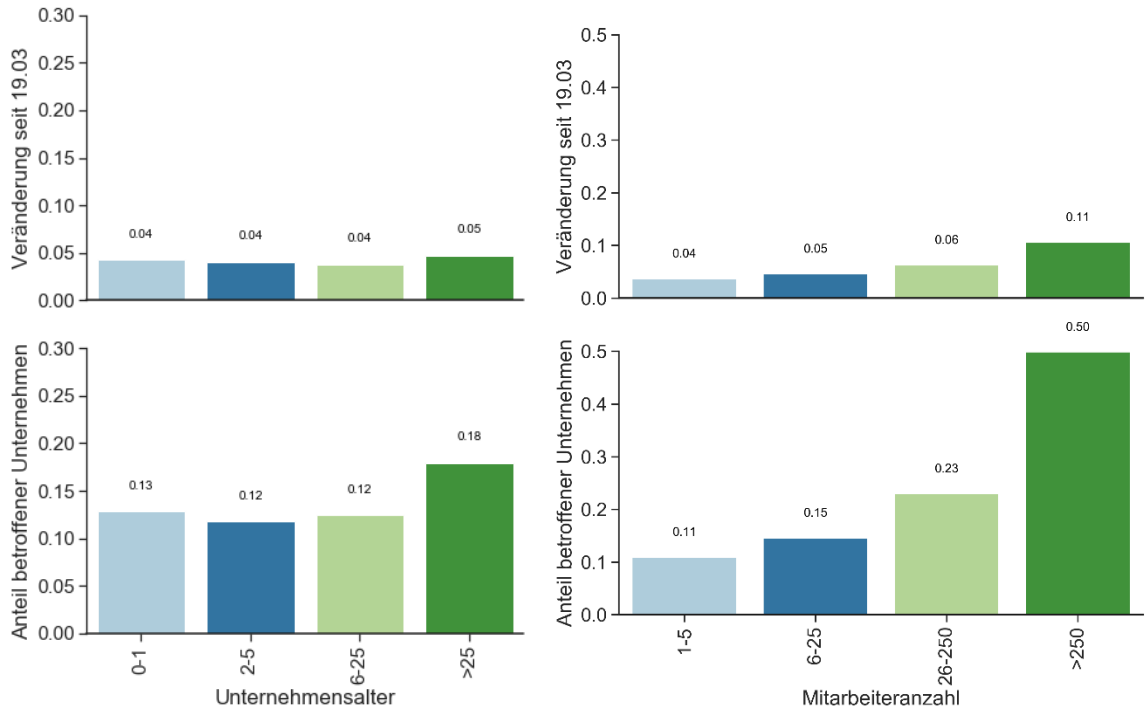


Abbildung 2: Anteil Unternehmen mit mindestens einer „Corona“-Webreferenz nach Unternehmensalter und -größe.

Abbildung 3 zeigt, dass es auch große sektorale Unterschiede gibt. Insbesondere die öffentliche Verwaltung, aber auch Vereine, Unterhaltungsbetriebe, das Gesundheitswesen und Bildungs- und Erziehungseinrichtungen berichten über die Corona-Pandemie. Nur geringe Anteile an Meldungen finden sich hingegen für das Baugewerbe, den Großhandel, die Landwirtschaft, Logistikunternehmen und das verarbeitende Gewerbe.

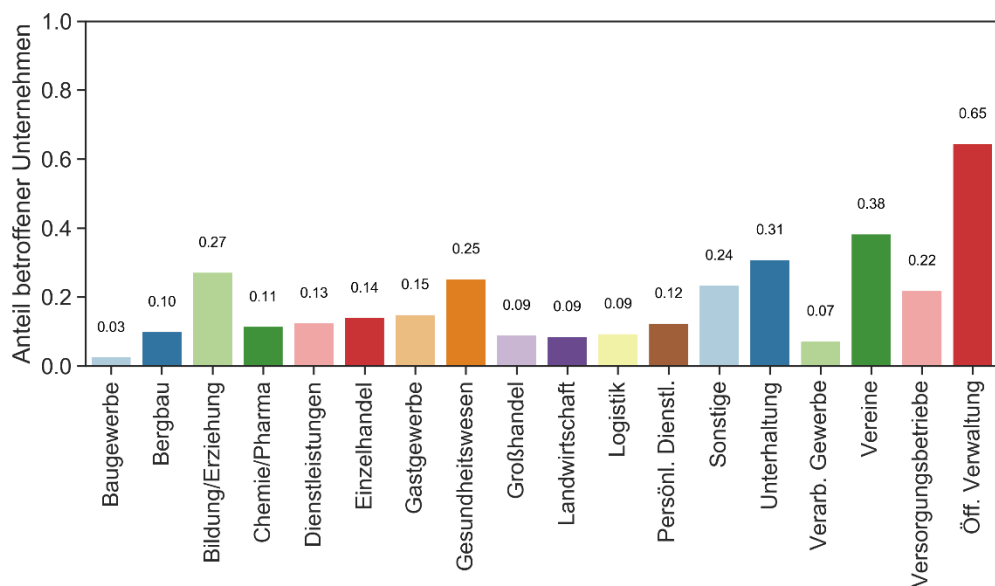


Abbildung 3: Anteil Unternehmen mit mindestens einer „Corona“-Webreferenz nach Branchen.

Corona-Pandemie betrifft Unternehmen unterschiedlich

In Abbildung 4 ist zu sehen, dass alle Branchen über die vergangenen vier Wochen eine ähnliche Zunahme an „Corona“-Referenzen zu verzeichnen haben. Allerdings starteten einige Branchen, wie zum Beispiel die öffentliche Verwaltung, auf einem bereits zu Beginn unserer Messungen sehr hohen Niveau und verzeichnen somit ein eher geringes relatives Wachstum. Andere Branchen, wie zum Beispiel der Großhandel, starten auf einem sehr niedrigen Niveau, zeigen dann aber ein sehr hohes relatives Wachstum. So verdoppelt sich die Anzahl betroffener Großhandelsunternehmen über den gezeigten Zeitraum beinahe.

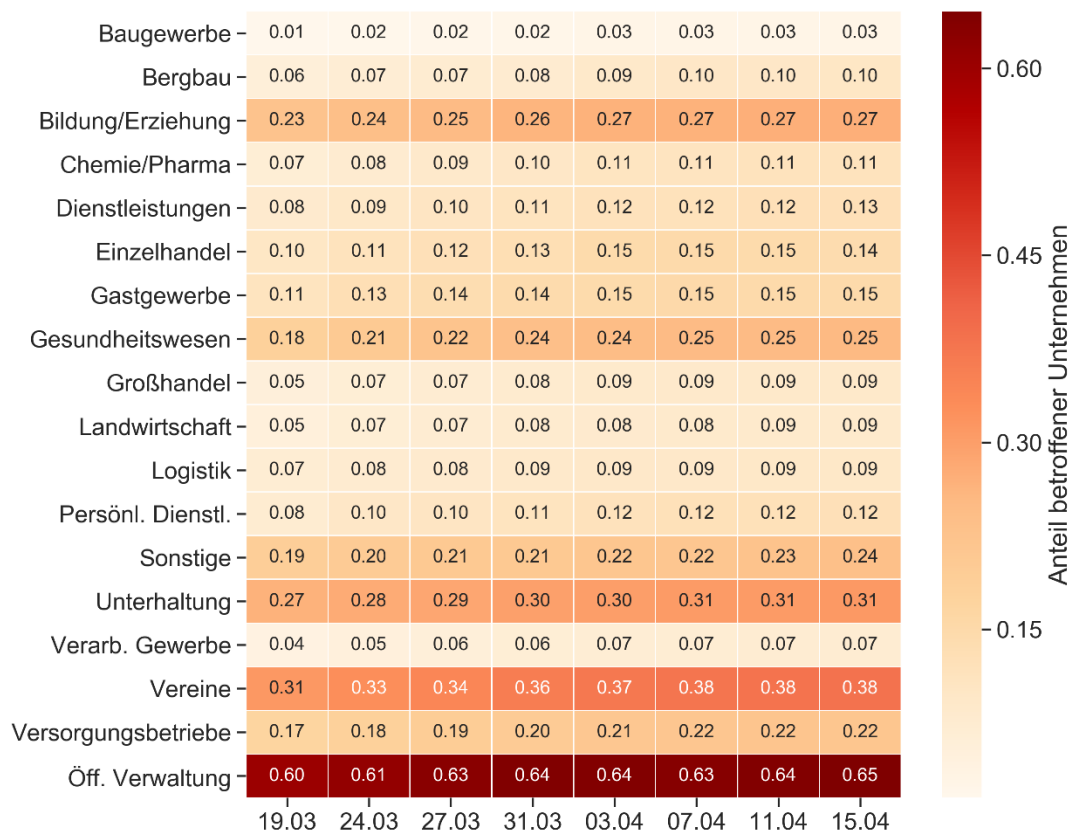


Abbildung 4: Anteile und Veränderung Unternehmen mit mindestens einer „Corona“-Webreferenz nach Branchen.

Abbildung 5 verdeutlicht ein Ost-West-Gefälle mit Blick auf den Anteil regional betroffener Unternehmen in Deutschland (linke Karte). Die Anteile betroffener Unternehmen in Stadtkreisen sind meist höher als in Landkreisen, wobei dieser Unterschied insbesondere im Osten der Bundesrepublik auffällig ist. Auch bei der Veränderung über die vergangenen Wochen (rechte Karte), zeigen sich regionale Unterschiede. So sind besonders hohe Wachstumsraten in den westlichen Bundesländern Baden-Württemberg, Hessen, Rheinland-Pfalz, Nordrhein-Westfalen und Niedersachsen zu beobachten.

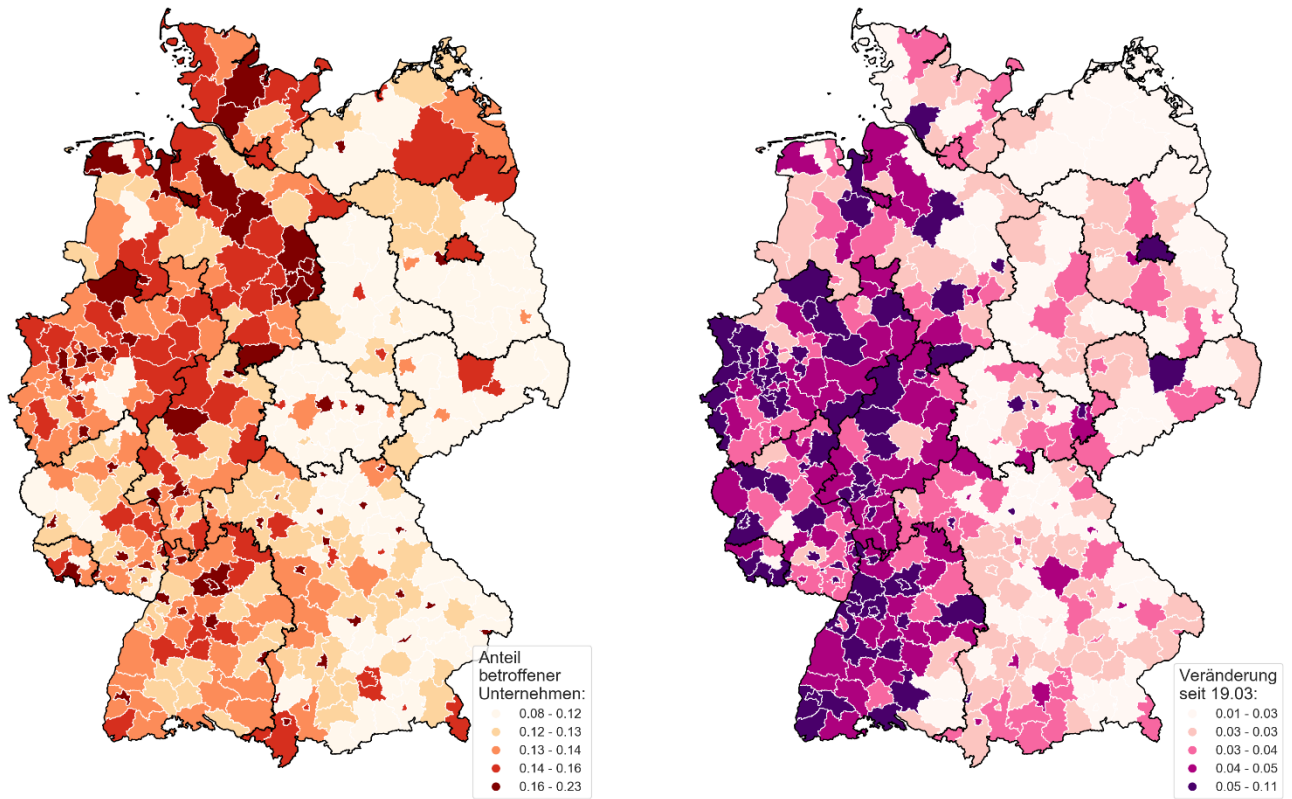


Abbildung 5: Anteil (links) und Veränderung (rechts) Unternehmen mit mindestens einer „Corona“-Webreferenz nach Kreisen.

3.2. Art der Betroffenheit

Abbildung 6 zeigt die häufigsten Wörter der von uns definierten „Corona“-Referenzklassen („Problem“, „Kein Problem“, „Adaption“ und „Information“, siehe Abschnitt 2) als Wortwolken. Als Grundlage dienen hierbei die insgesamt etwa 460.000 identifizierten und mit dem istari.ai Textanalysemodell klassifizierten „Corona“-Referenzen aus den Unternehmenswebseiten (Stand 24.03).

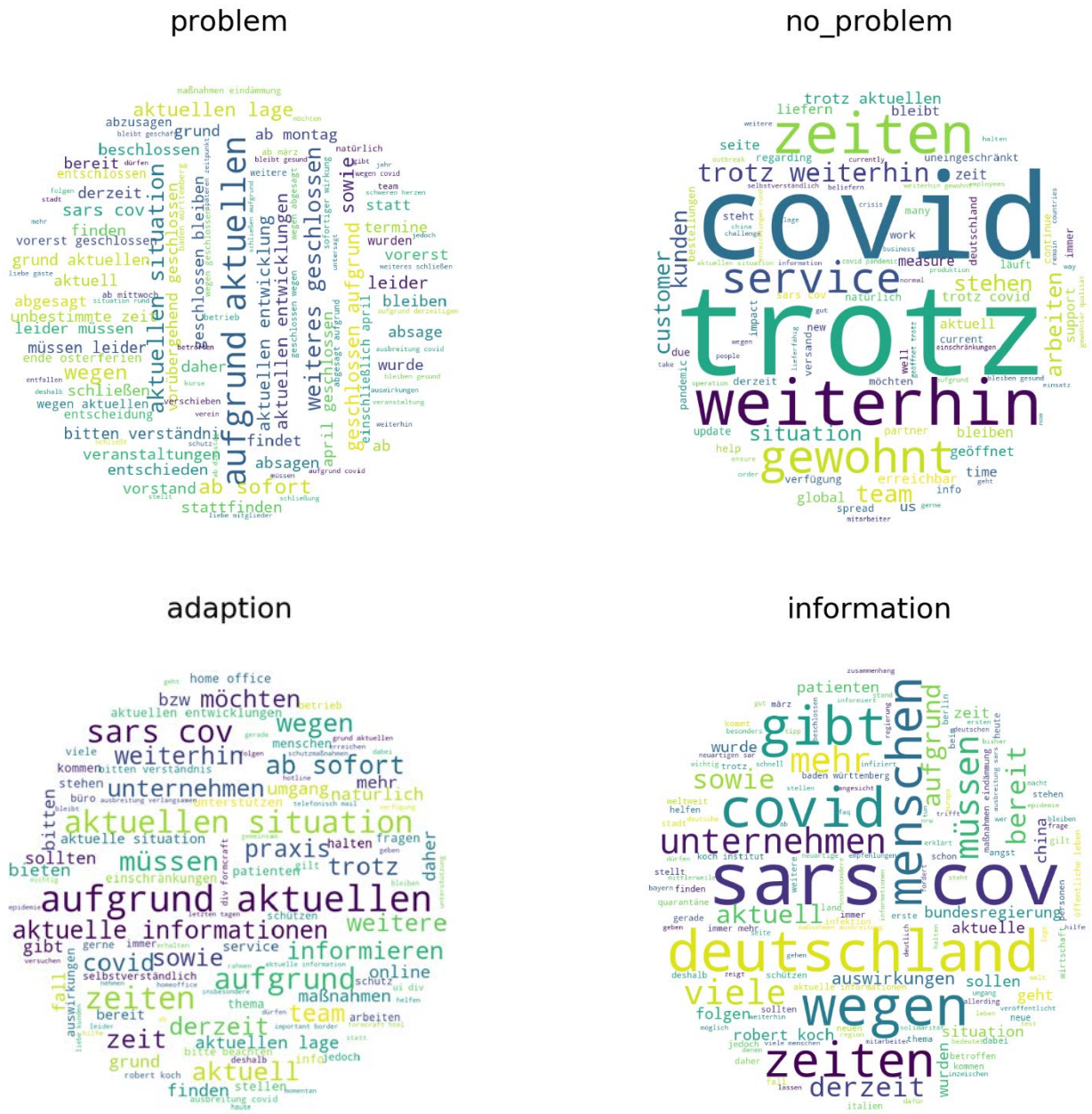


Abbildung 6: Wortwolken für vier „Corona“-Referenzklassen.

Abbildung 6 zeigt, dass Unternehmen mit mindestens einer „Corona“-Referenz je nach Branche unterschiedlich stark von den einzelnen Referenzklassen betroffen sind. So informiert beispielsweise die öffentliche Verwaltung sehr häufig über „Corona“-spezifische Themen, während das Gesundheitswesen vor allem über Anpassungen berichtet und die Unterhaltungsbranche insbesondere von schwerwiegenden Problemen. Keine, beziehungsweise geringe Probleme scheinen die Sektoren Chemie und Pharmaindustrie, Großhandel und (nicht-persönliche) Dienstleistungen zu haben. Anzumerken ist hier, dass ein Unternehmen von mehreren Referenzklassen betroffen sein kann, wenn mehrere „Corona“-Referenzen auf dessen Webseite gefunden wurden. So kann es auch vorkommen, dass ein Unternehmen beispielsweise sowohl von „Problem“ als auch „Kein Problem“ betroffen sein kann. Ein Beispiel hierfür wären Gärtnereien, die ihre (mittlerweile verbreitet geführte) Gastronomie schließen müssen, aber das Kerngeschäft (Verkauf von Pflanzen und Gartenbedarf) größtenteils uneingeschränkt fortführen können.

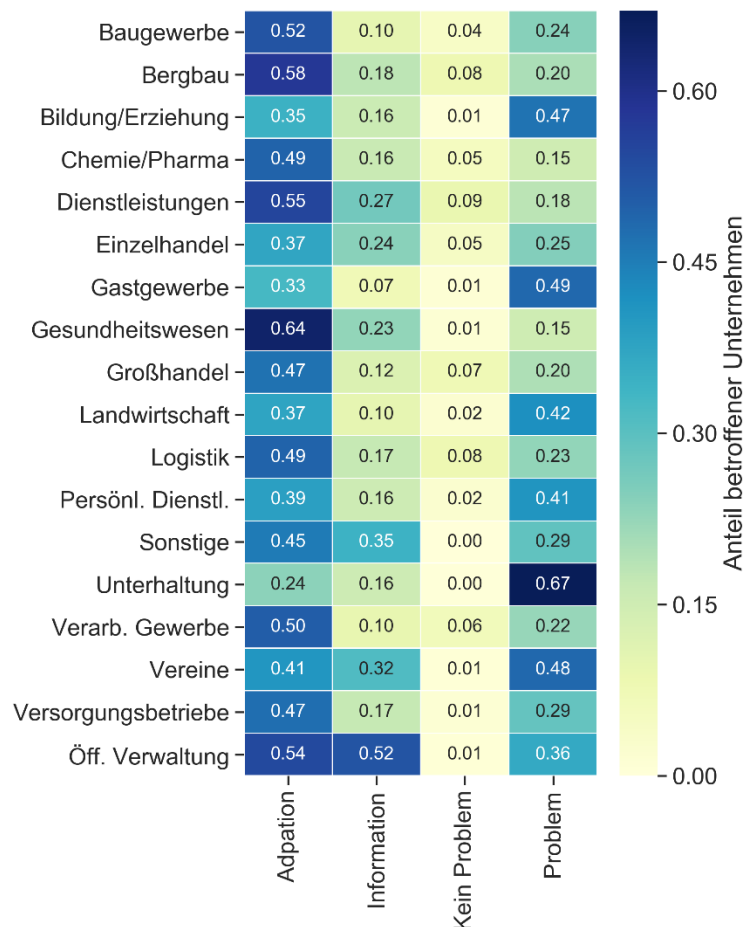


Abbildung 7: Anteil von Unternehmen mit mindestens einer „Corona“-Webreferenz, die der jeweiligen Klasse zugeordnet wurde, nach Branchen.

Corona-Pandemie betrifft Unternehmen unterschiedlich

Abbildung 8 zeigt, dass diese „Corona“-Referenzklassen regional unterschiedlich und teilweise (statistisch signifikant) gehäuft auftreten. Besonders auffällig ist dies bei den entgegengesetzten Klassen „Problem“ und „Kein Problem“. So weisen insbesondere die Ostdeutschen Bundesländer mehrere statistisch signifikante (99% Konfidenz) Häufungen von Kreisen mit niedrigen „Problem“ (blau schraffierte Coldspots), beziehungsweise hohen „Kein Problem“ (rot schraffierte Hotspots) Anteilen betroffener Unternehmen auf. In diesen Regionen berichten die Unternehmen auch vermehrt über Anpassungen an die Corona-Pandemie, wie den „Adaption“-Hotspots (rot schraffiert) zu entnehmen ist.

Corona-Pandemie betrifft Unternehmen unterschiedlich

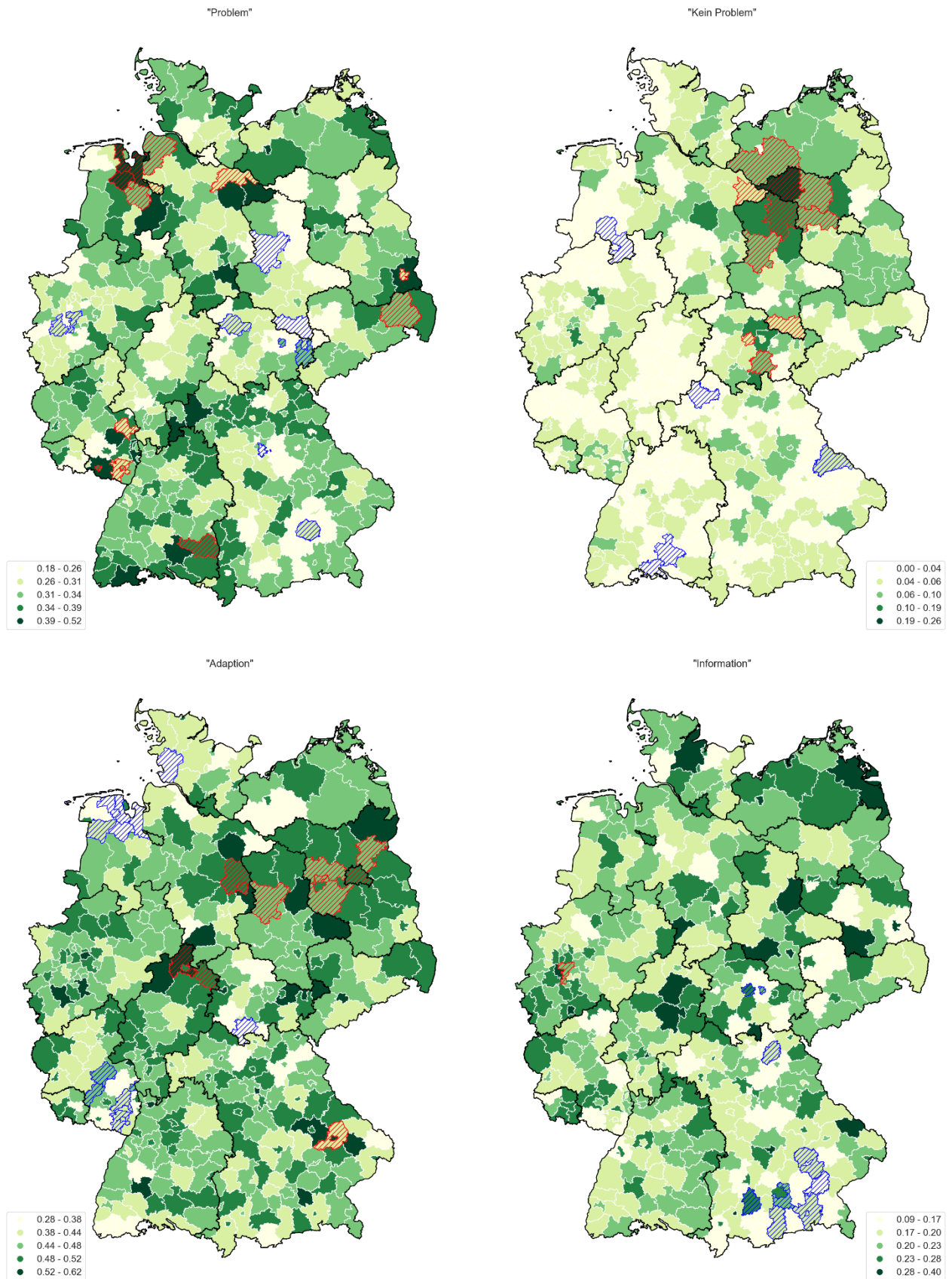


Abbildung 8: Anteil von regional betroffenen Unternehmen mit mindestens einer „Corona“-Webreferenz, die der jeweiligen Klasse zugeordnet wurde, nach Kreisen. „Hotspots“ rot, „Coldspots“ blau schraffiert.

Anmerkung: Schraffierte Regionen bilden gemeinsam mit den ihnen angrenzenden (teilweise nicht schraffierten) Regionen einen Hotspot beziehungsweise Coldspot. Beispielsweise bildet der schraffierte Landkreis Erding nahe München gemeinsam mit den angrenzenden Landkreisen Landshut, Freising, Mühldorf, Ebersberg und München einen „Problem- Coldspot“.

4. Fazit

Die Ergebnisse der vorliegenden Kurzexpertise zeigen, dass sich die Auswirkungen der Corona-Pandemie auf Deutschlands Unternehmen mittels KI-gestützter Webanalyse tagesaktuell und in hoher sektoraler wie regionaler Auflösung beobachten lassen. Die hierbei gewonnen Erkenntnisse zeigen den dynamischen Verlauf der Pandemie und die kommunizierten Reaktionen der betroffenen Unternehmen, welche sich regional teils deutlich unterscheiden. So konnte insbesondere ein starkes Ost-West-Gefälle festgestellt werden, sowohl bei der generellen Betroffenheit (Erwähnung der Pandemie auf der Unternehmenswebseite), als auch nach der Art der Betroffenheit fest. So berichten Unternehmen aus den Bundesländern Sachsen-Anhalt, Brandenburg und Mecklenburg-Vorpommern tendenziell eher von Anpassungen an die aktuelle Lage oder davon, dass sie wenig betroffen sind, während Unternehmen in westlichen Bundesländer eher von schwerwiegenderen Problemen berichten. Allerdings gibt es auch hier Unterschiede innerhalb der einzelnen Bundesländer. Im bundesweiten Vergleich kommunizieren vor allem Unternehmen aus städtischen Gebieten über die Pandemie, wobei unsere inhaltliche Auswertung hier keine klaren Unterschiede in der Art der Betroffenheit gezeigt hat. Für eine endgültige Bewertung der präsentierten Ergebnisse muss allerdings noch möglichen Verzerrungen nachgegangen werden, die sich aus regional unterschiedlichen Branchenstrukturen und unternehmensspezifischen Eigenschaften ergeben könnten.

Der hier präsentierte Ansatz ist ein erster Schritt hin zu einem tagesaktuellen Monitoring der deutschen Wirtschaft und könnte folglich ein nützliches Instrument bei der Steuerung einer koordinierten und evidenzbasierten Antwort auf die gegenwärtige Krise sein.

Referenzen

- [1] J. Kinne and J. Axenbeck, "Web Mining of Firm Websites : A Framework for Web Scraping and a Pilot Study for Germany," Mannheim, 18–033, 2018.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018.

AUTORENTEAM

Georg Licht

ZEW – Leibniz-Zentrum für europäische
Wirtschaftsforschung Mannheim GmbH
L 7, 1
68161 Mannheim
www.zew.de
georg.licht@zew.de

Jan Kinne*

ZEW – Leibniz-Zentrum für europäische
Wirtschaftsforschung Mannheim GmbH
L 7, 1
68161 Mannheim
Tel.: +49 (0)621 1235-297
jan.kinne@zew.de

*Ansprechpartner für Rückfragen

