

Florian Becker

Information Retrieval: Automatisierte Dokumentenanalyse in M&A Due Diligence

Eine der zentralen Herausforderungen der Due Diligence besteht darin, dass unter hohem Zeitdruck und überwiegend manuell grosse Mengen relevanter Dokumente auf Risiken geprüft werden müssen. Insbesondere in der Legal Due Diligence scheint daher der Einsatz KI-basierter, automatisierter Prüfverfahren erfolversprechend. Die auf Machine-Learning-Modellen basierenden Contract-Review-Systeme können Effizienz und Genauigkeit sowie den Umfang der in der finalen Risikobewertung berücksichtigten Datenmenge erhöhen. Die Softwarelösungen von Kira Systems und Levertton sind technisch fortgeschritten und werden erfolgreich in der M&A-Praxis angewandt.

Beitragsart: LegalTech

Region: Schweiz; EU

Rechtsgebiete: Artificial Intelligence & Recht; LegalTech

Zitiervorschlag: Florian Becker, Information Retrieval: Automatisierte Dokumentenanalyse in M&A Due Diligence, in: Jusletter IT 12. November 2020

Inhaltsübersicht

1. Problemstellung der Due Diligence
2. Anwendbarkeit automatisierter Prüfverfahren
3. Technische Umsetzung in der Legal Due Diligence
4. Führende Contract-Analytics-Anbieter
 - 4.1. Kira Systems
 - 4.2. Leverton
5. Fazit

1. Problemstellung der Due Diligence

[1] Die Due Diligence bildet einen zentralen Bestandteil der Informationsphase bei Unternehmenstransaktionen. Im Rahmen der Buy-Side Due Diligence prüft und untersucht der Käufer das Kaufobjekt, um sich ein möglichst gutes und umfassendes Bild über das zu übernehmende Unternehmen zu verschaffen. Ziel ist, die mit der Übernahme einhergehenden Chancen und Risiken adäquat zu erfassen. Die so geschaffene Informationsbasis dient der Reduktion der Informationsasymmetrie zwischen Käufer und Verkäufer sowie der Erhöhung der Entscheidungsqualität. Die Due Diligence ist somit im Kern eine Sorgfaltsmassnahme, um das eigene Risiko zu begrenzen. Die im Rahmen der Due-Diligence-Prüfung erlangten Erkenntnisse können erhebliche Auswirkungen auf Kaufpreis, Garantieübernahmen, Festlegung von Rücktrittsrechten und die Transaktionssicherheit im Allgemeinen haben.¹

[2] Eine der zentralen Herausforderungen der Due Diligence besteht darin, dass unter hohem Zeitdruck und während für beide Parteien viel auf dem Spiel steht eine Vielzahl an relevanten Dokumenten auf Risiken geprüft werden muss. Das Volumen an Informationen ist meist so hoch (mehrere zehntausend Dateien und Verträge), dass die Durchsicht in aller Regel stichpunktartig erfolgen muss und eine grosse Anzahl an Dokumenten gar nicht erst gesichtet werden kann. Dies birgt die Gefahr, dass wichtige Informationen ausser Acht gelassen werden. Wie das Beispiel der Übernahme von Autonomy durch HP zeigt (USD 8.8 Mrd. nachträgliche Abschreibung), ist das daraus resultierende Risiko teils beträchtlich.^{2, 3}

[3] Der Einsatz automatisierter Prüfverfahren unter Verwendung von Künstlicher Intelligenz zur Analyse grosser Datenmengen hat hohes Potenzial, diesen Prozess zu verbessern. Es können sowohl weit mehr Dokumente und Daten in den Analyseprozess miteinbezogen als auch Genauigkeit und Effizienz erhöht werden.

2. Anwendbarkeit automatisierter Prüfverfahren

[4] Die traditionelle Due Diligence lässt sich in drei Teilbereiche unterteilen: Financial Due Diligence (FDD), Legal Due Diligence (LDD) und Commercial Due Diligence (CDD).

¹ GOLDBERG & GODWIN, *Your merger: Will it really add value?*, 2001, 32; SCHALAST & RAETTIG, *Grundlagen des M&A Geschäftes. Strategie – Recht – Steuern* (2. Aufl.), 2019, 34, 109 f., 114.

² CIESIELSKI, *How Autonomy Fooled Hewlett-Packard*, <http://fortune.com/2016/12/14/hewlett-packard-autonmy/>, 2016, Abschn. 1.

³ RIEN, *The Art of AI: The Impact of Artificial Intelligence on the Merger & Acquisition Strategy*, 2018, 22 f.; SCHALAST & RAETTIG, *Grundlagen des M&A Geschäftes. Strategie – Recht – Steuern* (2. Aufl.), 2019, 110.

[5] Die FDD beinhaltet die Analyse der finanziellen Situation der Unternehmung und dient der Bestimmung des zukünftigen Ertragspotentials. Die Datengrundlage bilden insbesondere vergangene Jahresabschlüsse, qualitative Interviews (Management) sowie die Arbeitsunterlagen der Wirtschaftsprüfer. Der Fokus liegt im Gegensatz zum Audit nicht auf dem Verifizieren der finanziellen Ergebnisse, sondern darauf, zu klären, woher sie stammen.⁴

[6] Bei der LDD müssen die Eigentumsverhältnisse überprüft und potenzielle (versteckte) Verbindlichkeiten identifiziert werden. Sie dient als Grundlage für den Kaufvertrag des Transaktionsobjekts. Von besonderer Bedeutung sind Change-of-Control-Klauseln und andere Übertragungshindernisse. Die Informationsbasis bilden primär die vorhandenen Verträge, wobei je nach Vertragstyp unterschiedliche Punkte zu prüfen sind. Im Kaufvertrag wird insbesondere festgelegt, wer für nach Abwicklung der Transaktion auftretende Schäden einstehen muss.⁵

[7] Bei der CDD oder auch Business Due Diligence wird evaluiert, ob das Übernahmobjekt von strategischem Interesse ist. Sie findet meist in der Vorbereitungsphase statt und wird vom Käufer selbst durchgeführt. Ziel ist, die zukünftige Performance des zusammenschlossenen Unternehmens zu prognostizieren. Es wird analysiert, inwieweit sich ergebende Wettbewerbsvorteile (z.B. grösserer Marktanteil) und Synergien (sich ergänzende Produkte, Economies of Scale) einen Unternehmenskauf rechtfertigen würden. Informationen werden durch qualitative Interviews und Rückgriff auf öffentlich zugängliche Quellen gewonnen.⁶

[8] Um automatisierte Prüftechniken unter Verwendung von *Natural Language Processing* (NLP) sinnvoll einsetzen zu können, müssen bestimmte Bedingungen erfüllt sein. Sie eignen sich am besten für die Analyse grosser Mengen an Daten und Dokumenten, die weitgehend standardisiert sind und deren Art des Inhalts in etwa bekannt ist.⁷ Wie effektiv die Verfahren genutzt werden können, hängt im Bereich der Due Diligence zudem von der Art der Transaktion, den zu prüfenden Informationen und den jeweils relevanten Fragestellungen ab.⁸

[9] Bei der CDD ist die Anwendung aktuell kaum möglich, da die Aufgabestellung meist sehr individuell und spezifisch gestaltet ist. In der FDD gibt es zwar standardisierte und wiederkehrende Aufgaben, diese können aber mit regelbasierten Analysetools gelöst werden, da sie überwiegend statisch sind und nicht stetig angepasst werden müssen. Insofern sind nicht zwingend Systeme mit Künstlicher Intelligenz erforderlich und aktuell im Markt auch kaum zu finden. Die meisten Anbieter von KI-Analysetools im Transaktionsbereich sind LegalTechs, die sich auf *Contract Review Automation* in der LDD spezialisiert haben.⁹

3. Technische Umsetzung in der Legal Due Diligence

[10] Die manuelle Überprüfung der Verträge nimmt bei der LDD in der Praxis bis zu 90 % der Zeit des gesamten Prozesses in Anspruch. Meist werden Junior-Lawyers oder Anwaltsgehilfen

⁴ HOWSON, Due Diligence: The Critical Stage in Mergers and Acquisitions, 2003, 45–49.

⁵ HOWSON, Due Diligence: The Critical Stage in Mergers and Acquisitions, 2003, 68, 71; SCHALAST & RAETTIG, Grundlagen des M&A Geschäftes. Strategie – Recht – Steuern (2. Aufl.), 2019, 127f.

⁶ HOWSON, Due Diligence: The Critical Stage in Mergers and Acquisitions, 2003, 77 f.; SCHALAST & RAETTIG, Grundlagen des M&A Geschäftes. Strategie – Recht – Steuern (2. Aufl.), 2019, 122f.

⁷ DALE, Industry Watch Law and Word Order: NLP in Legal Tech, 2018, 214.

⁸ SCHALAST & RAETTIG, Grundlagen des M&A Geschäftes. Strategie – Recht – Steuern (2. Aufl.), 2019, 99.

⁹ NIKOLAIDIS, Die Zukunft von Artificial Intelligence bei M&A-Transaktionen, 2018, 28, 35.

eingesetzt, um die Berichte und Vertragsbeziehungen zu prüfen. Dabei müssen alle für unterschiedlichste Aspekte der Transaktion relevante Passagen identifiziert werden. Sofern die Dokumente elektronisch verfügbar sind, wird häufig mit einer Kombination aus Keyword-Suche und manueller Überprüfung gearbeitet. Zunächst werden innerhalb der gesamten Datensammlung die wichtigsten Dokumente identifiziert und anschliessend nach den benötigten Informationen in den jeweiligen Verträgen gesucht. Um Zeit zu sparen, wird dabei stichpunktartig vorgegangen und grosse Teile ausgelassen. In der Konsequenz wird nur ein geringer Anteil des zur Verfügung stehenden Materials berücksichtigt und bei den übrigen Unterlagen darauf vertraut, dass alles in Ordnung ist.¹⁰

[11] Mittels Automated Contract Review können tausende Verträge automatisch auf bestimmte Kriterien hin überprüft werden. Voraussetzung ist, dass die Dokumente elektronisch lesbar sind. Sofern sie ursprünglich mit Textverarbeitungsprogrammen (z.B. MS Word) erstellt wurden, können sie mit Natural Language Processing Software direkt verarbeitet werden. Elektronische Formate wie Scans sowie alle Dokumente in Papierform müssen mittels Natural Image Processing maschinell lesbar gemacht werden. Die verwendete Technik nennt sich *Optical Character Recognition* (OCR) oder auch Texterkennung und erzielt bei Verwendung rekurrenter neuraler Netzwerke (Deep Learning) die besten Ergebnisse. Sind alle Dokumente elektronisch lesbar gemacht worden, kann der virtuelle Datenraum als Ausgangspunkt für die Analyse genutzt werden.¹¹

[12] Bis auf einige wenige regelbasierte Anwendungen (für Due Diligence z.B. Fuzzy-C-Means), die auf Wenn-Dann-Algorithmen basieren, benutzen für Volume Contract Review nahezu alle LegalTech-Anbieter Systeme, die auf Machine Learning beruhen.¹² Bei Machine-Learning-Algorithmen kann zwischen Supervised und Unsupervised Learning unterschieden werden. Bei der Analyse rechtlicher Unterlagen und Verträge dient *Supervised Learning* der Ausführung von Klassifizierungs-Aufgaben und der Durchführung von Regressionen. Mit *Unsupervised Learning* können Muster erkannt und die Struktur der Dokumente nachvollzogen werden. Durch Clustering können bestimmte Textabschnitte, Paragraphen und Klauseln gruppiert werden.¹³

[13] Der technische Prozess der *Information Extraction* (IE) beim Contract Review umfasst die Zerlegung des Vertrags in einzelne Bestimmungen, Klauseln oder Sätze, um diese anschliessend individuell zu prüfen und entweder Informationen zu extrahieren oder diese mit einer Benchmark zu vergleichen. Dadurch kann auf Anomalien (z.B. ungewöhnliche Vertragselemente) und statistische Ausreisser geschlossen werden, die dann von Rechtsanwälten näher analysiert werden können. Durch Benchmarking und Cross-Validation kann auch das Fehlen üblicherweise vorhandener Bestimmungen erkannt werden. Die Ergebnisse werden abschliessend tabellarisch dargestellt, um dem Verwender ein gutes Gesamtbild der aggregierten Daten zu vermitteln. Diagramme erleichtern zusätzlich den dokumentenübergreifenden Vergleich.¹⁴

¹⁰ RIEN, *The Art of AI: The Impact of Artificial Intelligence on the Merger & Acquisition Strategy*, 2018, 22 f.; ROEGEST, HUDEK & McNULTY, *A Dataset and an Examination of Identifying Passages for Due Diligence*, 2018, 465 f.

¹¹ DABASS & DABASS, *Scope of Artificial Intelligence in Law*, 2018, S. 5; SCHALAST & RAETTIG, *Grundlagen des M&A Geschäftes. Strategie – Recht – Steuern* (2. Aufl.), 2019, 116.

¹² DALE, *Industry Watch Law and Word Order: NLP in Legal Tech*, 2018, 214.

¹³ DABASS & DABASS, *Scope of Artificial Intelligence in Law*, 2018, 4 f.

¹⁴ DABASS & DABASS, *Scope of Artificial Intelligence in Law*, 2018, 5; DALE, *Industry Watch Law and Word Order: NLP in Legal Tech*, 2018, 214; ROEGEST et al., *A Dataset and an Examination of Identifying Passages for Due Diligence*, 2018, 466.

4. Führende Contract-Analytics-Anbieter

[14] Die beiden LegalTechs Kira Systems und Leverton gehören zu den international führenden Anbietern von Contract Review Automation Software im Bereich der LDD. Nachfolgend werden Geschäftsmodell, Produktangebot und eingesetzte Machine-Learning-Technologie näher betrachtet.

4.1. Kira Systems

[15] Kira Systems ist ein 2011 in Toronto, Kanada, gegründetes LegalTech, das sich auf automatisierte Vertragsanalyse und Due Diligence spezialisiert hat. Laut eigenen Angaben verwendet die Mehrheit der grössten Rechtsanwaltskanzleien der Welt ihre Technologie. Das Unternehmen zählt aktuell über 200 Mitarbeiter und konnte 2018 in einer Series A Finanzierungsrunde USD 50 Mio. für eine Minderheitsbeteiligung durch Insight Venture Partners erzielen. Die Gründer Noah Waisberg und Alexander Hudek haben ihren Hintergrund in Rechtswissenschaften, Computer Science und Bioinformatik.¹⁵

[16] Kira klassifiziert das Due Diligence Problem als *Information Retrieval* (IR) Aufgabe, bei der zunächst alle relevanten Bestimmungen in den Verträgen identifiziert werden müssen, um anschliessend auf deren Basis eine Risikobewertung vornehmen zu können. Die Algorithmen der Software operieren auf Satzebene, was bedeutet, dass Dokumente als eine Sequenz von Sätzen verstanden werden. Mithilfe von *Conditional Random Fields* (CRFs) wird jeder einzelne Satz als relevant (*responsive*) oder nicht relevant markiert. Den Kira Computer Scientists Roegiest, Hudek & McNulty zufolge liefert diese Methode deutlich bessere Ergebnisse als andere sequenzbasierte (Hidden Markov Models) und auch nicht-sequenzbasierte Machine-Learning-Modelle (logistische Regression) sowie hybride Ansätze (Hidden Markov Models kombiniert mit SVM classifiers).¹⁶ Zur Evaluation der Performance wurden die Metriken Präzision, Recall (Sensitivität) und der F1 Score (gewichteter Durchschnitt aus Präzision und Recall) verwendet. CRFs erreichten einen F1 Score von 0.88 auf Satzebene und 0.92 auf Annotationsebene. Kira wurde mit einem Datenset in englischer Sprache trainiert, das aus etwa 4'200 manuell annotierten Rechtsdokumenten (hauptsächlich Kreditverträgen) und über 15 Millionen einzelnen Sätzen besteht. Mit diesem Datenset wurden auch alternative Verfahren gegengeprüft.¹⁷

[17] Für Kunden ist Kira als Software sofort einsatzbereit, es ist also prinzipiell kein Training nötig. Mittels OCR wird automatisch in elektronisch lesbare Formate konvertiert. Kira enthält standardmässig über 1'000 integrierte Analysemethoden, um die geläufigsten Arten von Vertragsbestimmungen in den Dokumenten herauszufiltern. Der Verwender kann angeben, welche Bestimmungen für ihn besonders relevant sind und das System bei ungewöhnlichen Bestimmungen individuell auf deren Erkennung trainieren. Kira schliesst auf neue Bestimmungstypen auf Basis zuvor prozessierter Bestimmungen, bei denen Ähnlichkeiten durch automatische Abgleichsysteme entdeckt wurden. Bereits nach weniger als 20 Beispielen wird der neue Typ mit fast 90 % Genauigkeit erkannt. Je mehr Daten bzw. Verträge verarbeitet werden, umso besser ist das von Kira entwickelte Verständnis, nach welchen Vertragselementen gesucht werden soll. Grosse

¹⁵ Kira Inc, <http://kirasystems.com/how-it-works/due-diligence/> (abgerufen am 28. April 2020).

¹⁶ ROEGIEST et al., A Dataset and an Examination of Identifying Passages for Due Diligence, 2018, 465, 474.

¹⁷ ROEGIEST et al., A Dataset and an Examination of Identifying Passages for Due Diligence, 2018, 466–473.

Anwaltskanzleien haben so die Möglichkeit, mit der Zeit mit Kira ihre eigenen Due Diligence Analysemodelle zu bauen.¹⁸

[18] Während der Datenanalyse wird mittels Progress Tracking über den laufenden Fortschritt der Verarbeitung informiert. Die Ergebnisse werden in Tabellenform und mit Diagrammen präsentiert. Eine typische Tabelle ist nach den folgenden (und weiteren) Kriterien strukturiert: Document ID, Filename, Title, Parties, Date, Terms. Durch Auswählen der dargestellten Information wird man im System direkt zu der jeweiligen Stelle im jeweiligen Vertrag weitergeleitet, um so die Ergebnisse kontrollieren zu können (*Second-Level-Review*). Um einen Überblick über das Projekt und identifizierte Risiken zu erhalten, kann zum Dashboard (Cockpit-Modus) gewechselt werden.¹⁹

[19] Mehrere US-amerikanische Grosskanzleien berichten von ausgesprochen positiven Erfahrungen, vorwiegend aufgrund der individuellen Anpassbarkeit, einfachen Integrierung und Handhabung sowie Risikominderung durch das Vermeiden menschlicher Fehler. Bei individuellen Projekten wurde von Einsparungen von über USD 500'000 und 5'000 nicht abrechenbaren Arbeitsstunden berichtet (20–60 % Zeitersparnis).²⁰

4.2. Leverton

[20] Leverton ist neben Kira Systems eines der wenigen LegalTechs im Bereich Contract Analytics, das sich auch auf Due Diligence spezialisiert hat. Das Unternehmen wurde 2012 von einem fünfköpfigen Team um Emilio Matthaei in Berlin gegründet und unterhält Standorte in New York, London und Neu-Delhi. Vor der Übernahme durch MRI Software im Juli 2019 beschäftigte Leverton etwa 80 Mitarbeiter und konnte seine Series A Finanzierungsrunde mit über USD 10 Mio. Beteiligungskapital abschliessen. Neben der Tätigkeit in der LDD hat sich zunehmend auch der Einsatz immobilienrechtlicher Analysetools zum Kerngeschäft entwickelt.²¹

[21] Technisch operiert Leverton auf einem vergleichbaren Niveau wie Kira Systems, wobei Informationen zu den genauen Verfahrensweisen und verwendeten Machine-Learning-Modellen kaum zugänglich sind. Teilweise werden *Convolutional Neural Networks* (CNNs) eingesetzt. Leverton verwendet ein patentgeschütztes OCR-Verfahren und hat eine spezielle Methode für verbesserte *Named Entity Recognition* (NER) mit CNNs entwickelt. Diese Technik zur Identifizierung z.B. der in Verträgen genannten Vertragsparteien basiert auf der Kalkulation von Feature-Vektoren mithilfe der syntaktischen, orthografischen und räumlichen Eigenschaften des erkannten Wortes, denen eine bestimmte Position in einer Feature-Map (Netz aus einzelnen Zellen) zugewiesen wird. Je nach relativer und räumlicher Position (innerhalb des Dokuments und im Ver-

¹⁸ BETTS & JAEP, *The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life Into a Decades-Old Promise*, 2017, 224 f.; DALE, *Industry Watch Law and Word Order: NLP in Legal Tech*, 2018, 215; Kira Inc, <http://kirasystems.com/how-it-works/due-diligence/> (abgerufen am 28. April 2020).

¹⁹ Kira Inc, <http://kirasystems.com/how-it-works/due-diligence/> (abgerufen am 28. April 2020).

²⁰ BETTS & JAEP, *The Dawn of Fully Automated Contract Drafting: Machine Learning Breathes New Life Into a Decades-Old Promise*, 2017, 225 f.

²¹ DALE, *Industry Watch Law and Word Order: NLP in Legal Tech*, 2018, 215; HAUSER, *Aus in Deutschland: US-Softwareunternehmen kauft Leverton und schliesst Berliner Büro*, <http://www.juve.de/nachrichten/namenundnachrichten/2019/08/aus-in-deutschland-us-softwareunternehmen-kauft-leverton-und-schliesst-berliner-buero>, 2019, Abschn. 1.

gleich zu in der Nähe gelegenen anderen Wörtern) kann mit einer gewissen Wahrscheinlichkeit auf die Named Entities geschlossen werden.²²

[22] Die Nutzung der Software erfolgt auf einer SaaS basierten Online-Plattform. In 25 verschiedenen Sprachen können Dokumente per Drag-&-Drop-System hochgeladen und anschliessend per OCR in maschinell lesbare Formate konvertiert werden. Mit Deep-Learning-Algorithmen (unter Verwendung z.B. der im vorherigen Absatz erläuterten Verfahrensweise) werden die Verträge und Dokumente analysiert und die gesuchten Informationen extrahiert. In einem zweistufigen Verfahren wird zudem die Qualität der Ergebnisse geprüft und die Daten validiert. Ähnlich wie bei Kira Systems sind die dargestellten Informationen stets mit der originären Datenquelle im Dokument verknüpft, um dem Benutzer schnelles Überprüfen zu ermöglichen. Mittels Open API lassen sich die Analyseergebnisse exportieren bzw. in Systeme von Drittanbietern überführen.²³

[23] Leverton wird aktuell (Stand April 2020) von über 200 Unternehmen und 5'000 einzelnen Benutzern verwendet. In den USA wurden fünf Patente für die verwendeten Technologien angemeldet.²⁴

5. Fazit

[24] Eine der zentralen Herausforderungen der Due Diligence besteht darin, dass unter hohem Zeitdruck und überwiegend manuell grosse Mengen relevanter Dokumente auf Risiken geprüft werden müssen. Insbesondere bei der Vertragsanalyse in der Legal Due Diligence scheint daher der Einsatz KI-basierter, automatisierter Prüfverfahren erfolgversprechend. Die auf Machine-Learning-Modellen basierenden Contract-Review-Systeme können Effizienz und Genauigkeit sowie den Umfang der in der abschliessenden Risikobewertung berücksichtigten Datenmenge erhöhen. Kira Systems verwendet *Conditional Random Fields*, während Leverton teilweise mit *Convolutional Neural Networks* arbeitet. Beide Softwarelösungen sind technisch fortgeschritten und werden erfolgreich in der M&A-Praxis angewandt.

FLORIAN BECKER ist Student der Betriebswirtschaftslehre an der Universität St. Gallen (HSG).

²² U.S. Patent No. 16/582,351, 2020.

²³ Leverton Holdings LLC, <http://leverton.ai/product/#works> (abgerufen am 28. April 2020).

²⁴ Leverton Holdings LLC, <http://leverton.ai/product/#works> (abgerufen am 28. April 2020).