

**Projektbericht aus dem KI-Praktikum der Veranstaltung
Management Support Systems M I (Artificial Intelligence)
an der Universität Osnabrück**

**Konzeption und prototypische Implementierung
eines Systems zur personalisierten Selektion von
Stellenausschreibungen auf Basis eines Lebenslaufes
mit Hilfe von Case-Based Reasoning**

Wintersemester 2014/2015

Autoren

**Alexandre Witte
Meike Werner**

Betreuer

**Prof. Dr.-Ing. Bodo Rieger
Axel Benjamins**

Konzeption und prototypische Implementierung eines Systems zur personalisierten Selektion von Stellenausschreibungen auf Basis eines Lebenslaufes mit Hilfe von Case-Based Reasoning

Alexandre Witte¹, Meike Werner²

¹ Universität Osnabrück, alwitte@uni-osnabrueck.de

² Universität Osnabrück, meike-werner@web.de

Abstract

Bewerber begegnen bei der Stellensuche im Internet steigender Komplexität in Hinblick auf die Auswahl geeigneter Stellen. Der Status Quo ist eine manuelle Suche mit Filtern auf elektronischen Jobbörsen. Dies führt nicht immer zu befriedigenden Resultaten und ist außerdem zeitintensiv. Dieser Beitrag stellt ein Konzept für ein System vor, welches in einer bestehenden Stellenbörse eine personalisierte Selektion durchführt. Diese Selektion wird durch Case-Based Reasoning realisiert, wobei sowohl Lebensläufe als auch Stellenanzeigen mittels Text Mining aufbereitet werden. Das System lernt mit zunehmender Benutzung und soll schneller relevante Ergebnisse liefern, als es mit einer manuellen Stellensuche möglich wäre. Durch den Einbezug des Lebenslaufes eines Bewerbers wird eine personalisierte Selektion der Stellenausschreibungen angestrebt. Die Verbindung des Lebenslaufes mit den Stellenausschreibungen ist als Teil des Konzeptes prototypisch implementiert worden.

1 Motivation und Problemstellung

Organisationen unterstützen durch Informationstechnologie (IT) nahezu alle Prozesse, unter anderem in den Bereichen Finanzen, Controlling, Marketing oder Produktion. Auch im Bereich Personalmanagement (engl. Human Resource Management) ist ein Anstieg der Verwendung von Software für verschiedene Aufgaben zu verzeichnen. Am häufigsten werden dabei Anwendungen der IT im Bereich Stellenausschreibungen, im Bewerbungsablauf oder im Auswahlprozess für vakante Stellen verwendet (Chapman und Webster 2003).

Deutsche Großunternehmen veröffentlichten 2011 ca. 87 % ihrer vakanten Stellen auf eigenen Unternehmens-Websites und ca. 61 % auf Online-Stellenbörsen (Bildat und Lau 2012). Aus zunehmend elektronisch unterstützter Rekrutierung ergeben sich einerseits Vorteile, wie eine höhere Reichweite für Bewerber und Unternehmen, geringere Aufwendung für Werbung, besserer Datenzugang und -

verfügbarkeit sowie geringere Kommunikationsaufwendungen (Holm 2010). Der gesamte Bewerbungsprozess kann also transparenter und effizienter gestaltet werden. Andererseits ergeben sich auch Nachteile. So können, resultierend aus der erhöhten Reichweite und für Bewerber kostengünstigen Bewerbungsart, das Bewerbungsaufkommen und der damit einhergehende Bearbeitungsaufwand für Unternehmen wesentlich steigen (Dachrodt et al. 2014). Des Weiteren bestehen kultur-, regions- und branchenabhängige Unterschiede in der Terminologie von Unternehmen und Bewerbern. Bewerber müssen ihre Qualifikationen und Erfahrungen mit den Formulierungen von Jobbörsen oder Unternehmen ausdrücken, um beurteilen zu können, welche Stellenanzeigen relevant sind. Vor der gleichen Herausforderung stehen Unternehmen, die beurteilen müssen, ob die in einer Bewerbung angegebenen Qualifikationen denen ihres Anforderungskatalogs entsprechen. Im Ergebnis kann es sein, dass relevante Stellenausschreibungen aufgrund falsch gewählter Filter nicht gefunden werden oder dass geeignete Bewerber nicht als solche erkannt werden (Bildat und Lau 2012).

Aus diesen Ausführungen folgt, dass die steigende Anzahl zugänglicher Stellenausschreibungen bei Bewerbern einen Anstieg des zeitlichen Aufwands verursacht, geeignete Angebote zu finden. Außerdem müssen sowohl Unternehmen als auch Bewerber Qualifikationen mit Anforderungen vergleichen, um Stellen passend zu besetzen bzw. auf eine Stelle besetzt zu werden. Zur Unterstützung der Bewerber ist es sinnvoll, Teile des Bewerbungsprozesses, konkret die Stellensuche, zu vereinfachen.

2 Zielsetzung

In diesem Beitrag wird ein Konzept vorgestellt, welches das Ziel hat, dem Bewerber mit möglichst geringem Aufwand nur die relevanten Ergebnisse bei der Jobsuche anzuzeigen.

In der Regel bewirbt sich eine Person auf mehrere Stellen, d.h. sie bearbeitet verschiedene Aufgaben der grundlegend gleichen Art. Sie sammelt durch Wiederholung des Vorgangs Erfahrungen und lernt schneller die für sie relevanten Stellenausschreibungen zu identifizieren. Dieser Vorgang kann durch Techniken des maschinellen Lernens abgebildet werden. Insbesondere die Methode des Case-Based Reasoning (CBR) ist für die hier gestellte Problembeschreibung geeignet, da sie auf folgende Annahmen aufbaut (Beierle und Kern-Isberner 2014):

1. Ähnliche Probleme haben ähnliche Lösungen.
2. Probleme sind unterschiedlich, aber der Typ der Aufgabenstellung wiederholt sich.

Für die hier vorliegende Fallbasis aus Textdokumenten ist CBR besonders von Vorteil, da sie schwach strukturierte Konzepte beinhaltet, für die allgemeine Regeln schwer zu bilden sind (Beierle und Kern-Isberner 2014).

Für die Wissensakquisition aus den Dokumenten wird Text Mining, eine Unterform des Data Mining angewandt. Übergeordnet wird Data Mining auch mit dem Begriff Wissensfindung in Daten (Knowledge Discovery in Databases, KDD) bezeichnet (Beierle und Kern-Isberner 2014). Während Data Mining generell Daten untersucht, ist Text Mining auf Text spezialisiert, der in der Regel semi- oder unstrukturiert ist und aufgrund seiner Eigenschaft als Buchstabenfolge schwerer zu analysieren ist. Mittels Text Mining können Texte strukturiert und Informationen extrahiert werden. Dazu werden

hauptsächlich statistische oder mustererkennende (pattern recognition) Verfahren, wie bspw. Differenz- oder Clusteranalysen verwendet (Heyer et al. 2012).

Nachdem in Kapitel 1 das Thema eingeleitet und in Kapitel 2 das Ziel des Beitrags definiert wurde, beginnt Kapitel 3 mit der Darstellung des Konzeptes. Im darauf folgenden Kapitel 4 wird auf Basis der getroffenen Annahmen und verfügbaren Daten ein Prototyp vorgestellt. Kapitel 5 zeigt die Limitationen des Konzeptes auf, bevor in Kapitel 6 das Fazit den Beitrag abschließt.

3 Konzept

Das Konzept basiert auf einem für die Zwecke des Konzeptes modifizierten Case-Based Reasoning (CBR)-Cycle (Aamodt und Plaza 1994). Dabei werden die Eingabedaten für das System mittels Text Mining und Natural Language Processing (NLP) (Chowdhury 2003) aufbereitet und die Ähnlichkeit zwischen den Fällen wird mittels k-Nearest-Neighbor (k-NN)-Algorithmus ermittelt. Das gesamte Konzept ist in Abbildung 1 dargestellt und nachfolgend im Detail erläutert.

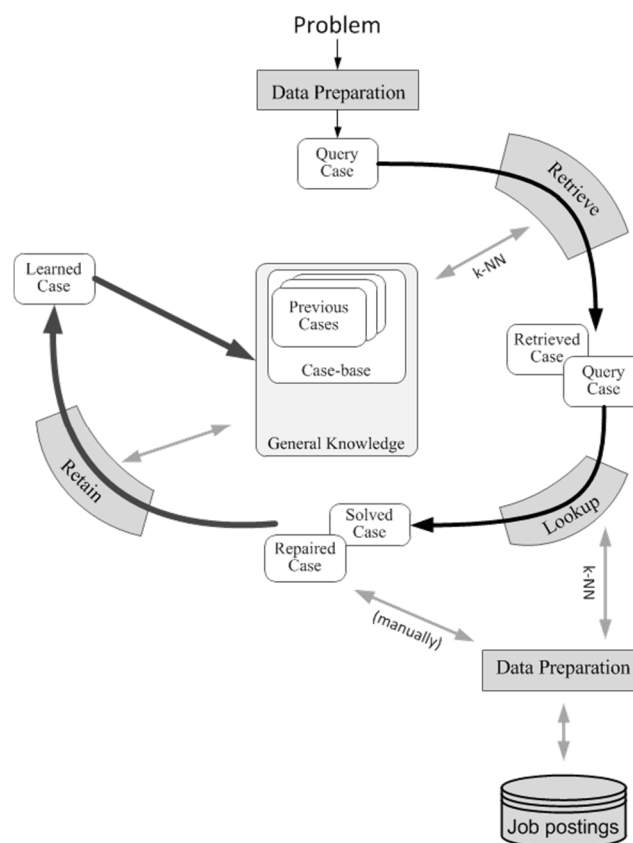


Abbildung 1: CBR-Cycle (Quelle: eigene Darstellung in Anlehnung an Aamodt und Plaza 1994)

Der Ausgangspunkt für das System ist der Eingang eines Problems, zu welchem eine Lösung zu finden ist. Der Lebenslauf des Arbeitssuchenden, welcher durch ein Web-Formular in das System geladen wird, stellt jenes Problem dar. Es wird von einem Lebenslauf in tabellarischer Form ausgegangen, in welchem die Ausbildung und Qualifikationen des Bewerbers aufgelistet sind. Somit würden bei-

spielsweise im Fall eines angehenden Anwalts entsprechende Jura-Studiengänge und Praktika in diversen Kanzleien Teil des Lebenslaufes sein. Die Phase *Data Preparation* wandelt die unstrukturierten bzw. semistrukturierten Daten mittels NLP-Methoden in einen strukturierten Datensatz um. Zunächst wird der gesamte Text in Kleinschreibung transformiert, um Unterschiede zwischen beispielsweise „Ein“ (Satzanfang) und „ein“ (Normalfall) zu eliminieren. Anschließend werden die Wörter auf ihren Wortstamm zurückgeführt (sog. *Stemming*), um eine Wortmenge zu bilden, welche korrekt quantifiziert werden kann. So werden beispielsweise konjugierte Verbformen in den Infinitiv transformiert und Pluralformen in den Singular umgewandelt. Im Folgenden werden die Sätze in einzelne Wörter aufgebrochen. Aus der resultierenden Wortliste werden Stopp- und Füllwörter, sowie Wörter, welche weniger als drei oder mehr als 50 Zeichen aufweisen, entfernt. Das Resultat ist ein Wortvektor, welcher alle für den Prozess gefilterten relevanten Wörter inklusive Quantifizierung enthält. Dieser Wortvektor bildet die Problembeschreibung des nun neu erzeugten *Query Case*. Die Phase *Data Preparation* ist damit abgeschlossen.

Es folgt die *Retrieve*-Phase des CBR-Cycles nach Aamodt und Plaza (1994). Die im vorherigen Schritt erzeugte Problembeschreibung des neuen Falles wird benutzt, um die Fallbasis nach einem Fall mit ähnlicher Problembeschreibung zu durchsuchen. Zur Ermittlung der Ähnlichkeit wurde, aufgrund der Form der zu vergleichenden Daten (quantifizierte Wortvektoren), eine numerische Abstandsermittlung gewählt. Das bedeutet, dass Fälle sich umso ähnlicher sind, desto mehr gleiche Wörter in ähnlicher Anzahl in der Problembeschreibung enthalten sind. Dieser Vergleich wird mittels des k-NN-Algorithmus durchgeführt, welcher das euklidische Abstandsmaß zwischen den Wortvektoren errechnet und den „nächsten Nachbarn“ (nearest Neighbor) zur neuen Problembeschreibung zurückliefert.

Abhängig vom Erfolg der vorhergegangenen Operationen wird entweder mit dem *Retrieved Case*, welcher den aus der Fallbasis gefunden Fall darstellt, oder mit dem ursprünglichen *Query Case* weiter verfahren. Die Fälle in der Fallbasis (Abbildung 2) weisen neben der Problembeschreibung zusätzlich eine Lösungskomponente auf, welche strukturell äquivalent zur Problembeschreibung gestaltet ist. Allerdings wurde der Wortvektor nicht aus dem Lebenslauf erstellt, sondern aus den Qualifikationsanforderungen einer Stellenausschreibung.

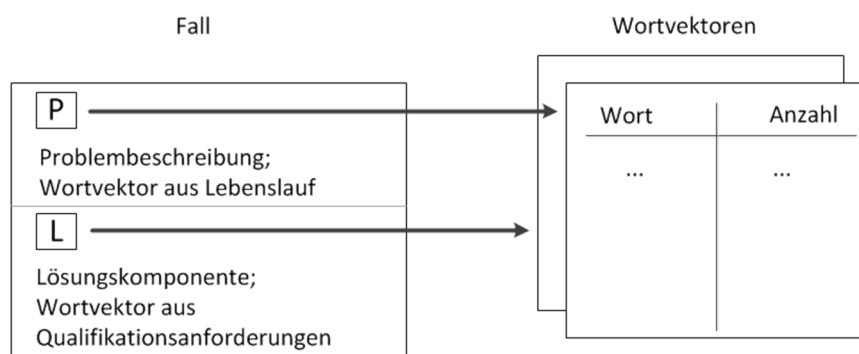


Abbildung 2: Fallbeschreibung und Wortvektoren (Quelle: eigene Darstellung)

Die Ermittlung einer Stellenausschreibung für den neuen Fall (den *Query Case*) ist der elementare Bestandteil der folgenden *Lookup*-Phase. An dieser Stelle löst sich das Konzept vom klassischen

CBR-Cycle nach Aamodt und Plaza (1994), in welchem die *Reuse*-Phase folgen würde. Während der *Reuse*-Phase würde die Lösungskomponente des aus der Fallbasis ermittelten Falles angepasst werden, um eine Lösung für den neuen Fall vorzuschlagen. Im vorliegenden Konzept wird die Lösungskomponente des gefundenen Falles nicht adaptiert, sondern lediglich vom k-NN-Algorithmus als Zielvorgabe genutzt, anhand derer nach ähnlichen Stellenausschreibungen in der Datenbank der aktuellen Stellenausschreibungen gesucht wird (*Lookup*). Falls während der *Retrieve*-Phase lediglich Fälle gefunden wurden, welche einen festzulegenden Schwellwert als Ähnlichkeitsmaß unterschreiten, wird die Problembeschreibung des *Query Case* verwendet. Hierzu werden analog zur *Data Preparation* am Beginn des Prozesses die vorhandenen, aktuellen Stellenausschreibungen in die Wortvektor-Form überführt, um eine Vergleichbarkeit mit der Problembeschreibung des *Query Case* herzustellen.

Es wird davon ausgegangen, dass die aktuellen Stellenausschreibungen in einer semistrukturierten Form vorliegen, d.h. dass das Objekt „Stellenausschreibung“ mehrere Attribute wie beispielsweise „Arbeitgeber“, „Ort“ oder „Gehaltsstufe“ etc. enthält, welche in Freitextform ausgeprägt sein können. Das Merkmal, welches eine relevante Bewerbung für den Bewerber wesentlich kennzeichnet, ist die Qualifikationsanforderung, welche wiederum in Freitextform vorliegt. Aufgrund dessen wird dieses Merkmal für alle aktuellen Stellenausschreibungen in die Form des oben beschriebenen Wortvektors transformiert. Abhängig vom Ausgang der *Retrieve*-Phase wird nun mittels des k-NN-Algorithmus entweder die Problembeschreibung des *Query Case* oder die Lösungskomponente des *Retrieved Case* mit den Qualifikationsanforderungen der Stellenausschreibungen verglichen. Im Gegensatz zur *Retrieve*-Phase im klassischen CBR-Cycle wird jedoch nicht ein einziger passender Fall, sondern eine Auswahl von Stellenbeschreibungen, welche bezüglich ihrer Ähnlichkeit zum Lebenslauf einen Schwellwert nicht unterschreiten, gesucht. Der Prozess ist in Abbildung 3 dargestellt.

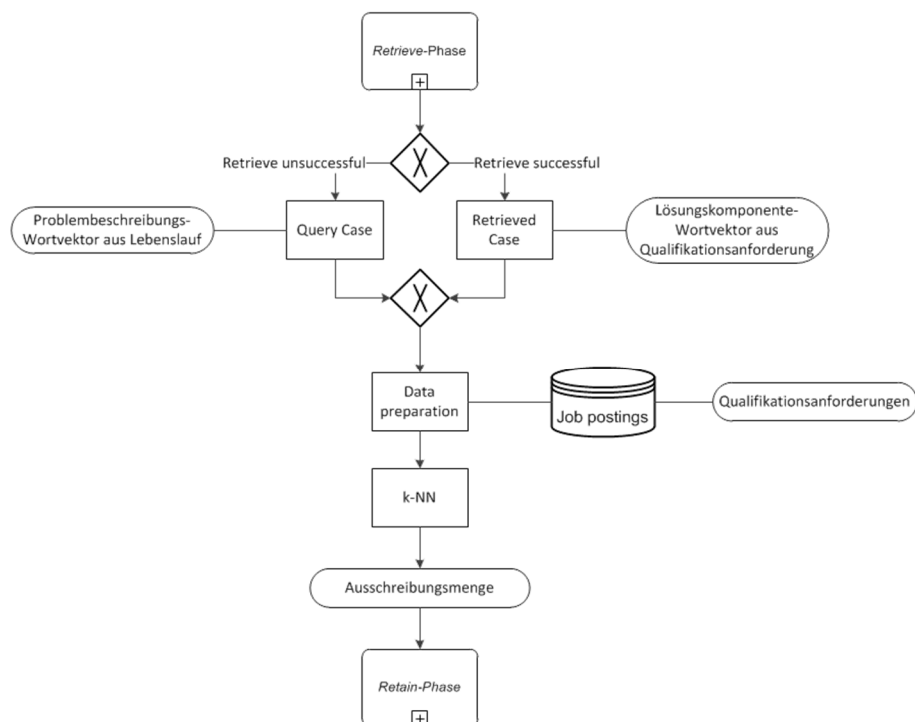


Abbildung 3: Ablauf der Lookup-Phase (Quelle: eigene Darstellung)

Die resultierende Menge an Stellenausschreibungen, welche das System als ausreichend relevant ermittelt hat, wird dem Bewerber präsentiert. Interessiert sich der Bewerber für eine der Ausschreibungen, kann er dieses dem System durch einen Klick auf eine entsprechende Schaltfläche mitteilen. Findet der Arbeitssuchende kein passendes Angebot, wird ihm die Möglichkeit geboten, manuell bzw. mit einer klassischen Suchmaske die aktuellen Stellenausschreibungen zu durchsuchen. Im Falle eines Sucherfolges auf diesem Wege, wird ebenfalls durch die Auswahl einer Schaltfläche festgehalten, welche Stellenausschreibung für den Bewerber interessant ist (*Repaired Case*). Dies ist der Ausgangspunkt für die *Retain*-Phase des CBR-Cycles. Das vorliegende Konzept verzichtet auf die *Revise*-Phase des klassischen CBR-Cycle, da die Fallkomponenten nicht bearbeitet oder angepasst, sondern lediglich zur Suche ähnlicher Fälle bzw. Wortvektoren genutzt werden.

In der *Retain*-Phase wird der *Query Case* um die Lösungskomponente ergänzt. Die Lösungskomponente besteht aus dem Wortvektor der Qualifikationsanforderung der ausgewählten Stellenbeschreibung. Es resultiert ein vollständiger Fall, der sogenannte *Learned Case*. Er wird der Fallbasis zugeführt, um die Wissensbasis des Systems zu erweitern. Lädt nun beispielsweise ein weiterer Bewerber einen Lebenslauf in das System, welches der Problemkomponente des *Learned Case* (oder eines anderen Falles der Fallbasis) ähnelt, so kann die Lösungskomponente dieses Falles in der *Lookup*-Phase verwendet werden. Dadurch wird die Lernfähigkeit des Systems erreicht.

4 Prototypische Implementierung

4.1 Datengrundlage

Für den Prototypen werden zwei Datenquellen verwendet: Lebensläufe und Stellenanzeigen. Die Lebensläufe liegen im PDF-Format vor. Sie sind in englischer Sprache verfasst und weisen in Hinblick auf Inhalt und Formatierung Unterschiede auf. Es lassen sich lediglich schwach strukturierte Konzepte erkennen, die Dateien sind also als unstrukturiert bzw. höchstens semistrukturiert einzustufen. Für die Implementierung des Prototyps wurden exemplarisch verschiedene Lebensläufe, welche über eine Suchmaschinen-Recherche im Internet gefunden wurden, ausgewählt.

Die zweite Datenquelle liegt in Form eines CSV-Dokuments vor. Dieses enthält englischsprachige Stellenangebote der Jobbörse von New York City. Der Datensatz beinhaltet alle vakanten Stellen zu Dezember 2014 und stellt diese anhand von 26 Merkmalen dar. Die Ausprägungen der Merkmale sind jedoch nicht durchgehend gleichmäßig aufgebaut. Dies bezieht sich auf die inhaltliche sowie formale Ausgestaltung. Inhaltliche Unterschiede sind bei Merkmalen zu finden, die ein persönliches Ermessen hinsichtlich Konkretisierung und Inhalt zulassen. Die Unterschiede in der Formatierung betreffen die Darstellung von Sonderzeichen und fehlende Ausprägungen von Merkmalen. Der Datensatz ist daher als semistrukturiert einzustufen.

Aufgrund der qualitativen Unterschiede unter den Stellenanzeigen wird für den Prototyp eine Auswahl relevanter Merkmale getroffen. Für die Bildung von Wortvektoren beim Text Mining während der *Data Preparation* werden nur textbasierte Merkmale verwendet, die inhaltlich in Bezug zu den Le-

bensläufen stehen. Dafür werden ausschließlich Qualifikationsanforderungen und Tätigkeitsbeschreibungen ausgewählt.

Wird eine Auswahl passender Stellen für den Arbeitssuchenden generiert, sind weitere Merkmale aus der Datenbank relevant. Um die Arbeitsplätze angemessen zu präsentieren, sind neben Qualifikationsanforderung und Tätigkeitsbeschreibung weitere Angaben wie Abteilung, Tätigkeitsstätte, Arbeitszeit, Gehalt und der Adressat der Bewerbung erforderlich. Mittels NLP werden unstrukturierte Texte verarbeitet und im Output nur eine Auswahl von Merkmalen angezeigt. Aus diesem Grund kann auf eine Überarbeitung der ursprünglichen Daten verzichtet werden.

4.2 Implementierung

Um eine erste Evaluation des Konzeptes sowie der Eignung des k-NN-Algorithmus zur Ermittlung ähnlicher Wortvektoren zu ermöglichen, wurde ein Ausschnitt der *Lookup*-Phase (Abbildung 3) prototypisch implementiert. Hierzu wurde RapidMiner Studio 6.1 Enterprise Edition (<http://rapidminer.com>) genutzt. Der erstellte Prozess ist in Abbildung 4 dargestellt.

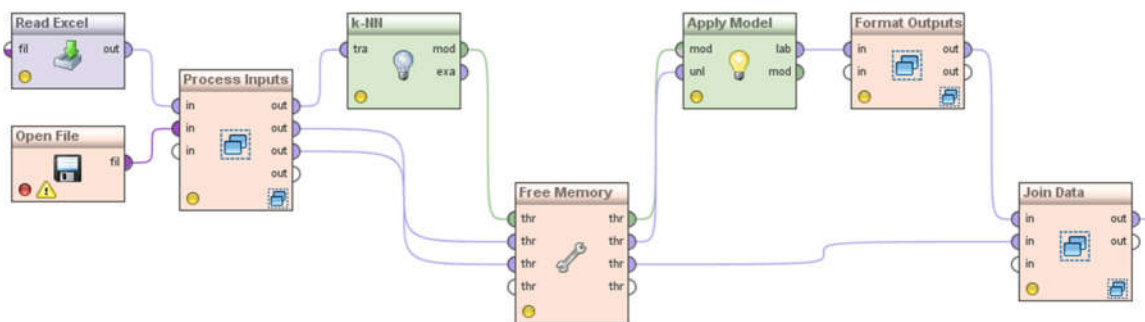


Abbildung 4: *Lookup*-Prozess in RapidMiner (Quelle: eigene Darstellung)

Der Prozessablauf erfolgt von links nach rechts. Zunächst werden die Eingabedaten eingelesen. Am Operator *Read Excel* liegen die Stellenangebotsdaten im CSV-Format an, am *Open File*-Operator ein Lebenslauf im PDF-Format. Die Daten durchlaufen zunächst den Teilprozess *Process Inputs*, in welchem sie in Wortvektoren transformiert werden. Der Vorgang ist in zwei Phasen gegliedert. Im ersten Schritt werden die relevanten Merkmale des Stellenangebot-Datensatzes ausgewählt sowie ein bezeichnendes Attribut (*Business Title*, der Titel der Stellenausschreibung) als *Label* gekennzeichnet. Die Kennzeichnung als *Label* bedeutet, dass dies das Attribut ist, welches der k-NN-Algorithmus für einen anderen Datensatz, welcher keinen *Business Title* aufweist, annähern soll. Anschließend wird mittels NLP-Methoden der Wortvektor aus den Merkmalen gebildet, welche die Qualifikationsanforderungen enthalten. Bei der Bearbeitung des Lebenslaufes wird analog vorgegangen, mit Ausnahme der Auswahl der Attribute. Es wird der gesamte Lebenslauf verwendet, im Anschluss werden lediglich Wörter aus dem Lebenslauf in den Wortvektor aufgenommen, welche mit einer bestimmten Anzahl (im Prototyp wurde der Schwellwert fünf verwendet) im Dokument vertreten sind. Hiermit soll eine entsprechende Relevanz sichergestellt werden.

Das Resultat bilden drei Datengruppen: Zunächst die Wortvektoren, welche aus den Stellenausschreibungen generiert wurden, der Wortvektor aus dem Lebenslauf sowie die bereinigte Menge der ursprünglichen Stellenausschreibungen (vgl. Kapitel 4.1). Letzteres wird zum Abschluss des Prozesses verwendet, um eine für den Bewerber sinnvolle Liste (inkl. Arbeitgeber, -ort, etc.) von Stellenausschreibungen erzeugen zu können.

Mittels der aus den Stellenausschreibungen generierten Wortvektoren wird der k -NN-Algorithmus trainiert (Operator k -NN), so dass das resultierende Modell in der Lage ist, Datensätzen entsprechend dem Trainingsdatensatz ein *Label* (Business Title, s.o.) zuzuordnen zu können. Dies erfolgt im *Apply Model*-Operator, welcher als Input-Daten das trainierte Modell sowie den Wortvektor des Lebenslaufes erhält.

Das Ergebnis ist eine Menge von Stellenausschreibungen, welchen jeweils ein *Confidence*-Wert zugeordnet ist. Dieser Wert ist umso höher, desto ähnlicher sich die Wortvektoren aus den Qualifikationsanforderungen der Stellenangebote und den Angaben im Lebenslauf sind. Im Teilprozess *Format Outputs* wird die Ergebnistabelle auf diejenigen Ergebnisse, deren zugeordnete *Confidence* einem bestimmten Schwellwert an Relevanz genügt, reduziert und überflüssige Attribute, welche durch die Berechnungen des k -NN-Operators entstanden sind, entfernt. Abschließend werden die resultierenden Stellenausschreibungen mit ihren ursprünglichen Merkmalen (Arbeitgeber, Ort, etc.) zusammengeführt.

Der Prozess soll durch ein Beispiel illustriert werden. Ein Arbeitssuchender mit einer akademischen Ausbildung im juristischen Bereich sei auf der Suche nach Beschäftigung bei der Stadt New York. Die Person lädt ihren Lebenslauf in das System. Hierdurch wird der in Abbildung 4 dargestellte Prozess angestoßen. Zunächst werden parallel sowohl die aktuellen Stellenausschreibungen sowie der hochgeladene Lebenslauf in Wortvektoren transformiert. Für den Bewerber beinhaltet die Wortmenge Wörter, welche mindestens fünfmal im Dokument vorkommen und keine Stopp- oder Füllwörter sind. In diesem Fall ist es wahrscheinlich, dass bspw. „law“, „lawyer“, „college“, „school“ usw. in der Wortmenge enthalten sind.

Die Wortvektoren der Qualifikationsanforderungen, welche ebenfalls um Stopp- und Füllwörter gekürzt wurden, werden genutzt, um den k -NN-Algorithmus zu trainieren. Beispielsweise sei ein Wortvektor aus der Qualitätsanforderung einer Stelle als Anwalt entstanden, welcher ebenfalls Wörter wie „law“, „lawyer“ usw. enthält sowie mittels des Merkmals *Business Title* als Stelle für einen Anwalt gekennzeichnet ist. Das Training mit diesem Datensatz lehrt den Algorithmus, ähnliche Wortvektoren in Zukunft ebenfalls mit dem *Label* „lawyer“ zu versehen.

Im *Apply Model*-Operator wird der trainierte k -NN-Algorithmus auf den Wortvektor des Lebenslaufes angewandt. Anhand der Ähnlichkeit der Wortmengen wird eine Auswahl an Stellenangeboten selektiert, welche im Anschluss mit den für die Anzeige relevanten Informationen wie Arbeitsort oder Arbeitgeber verbunden wird. Anschließend wird dem Bewerber die resultierende Menge von Stellenausschreibungen präsentiert.

5 Limitationen

Das vorgestellte Konzept beschreibt ein System zur Erleichterung des Bewerbungsprozesses für Bewerber. Für die Sicherstellung einer hohen Effizienz bei der Zuordnung von Stellenausschreibung zu Lebensläufen sind jedoch gewisse Bedingungen zu erfüllen.

Zunächst ist ein hoher Datenbestand an Stellenausschreibungen beim Arbeitgeber wünschenswert – je höher die zu untersuchende Datenmenge, desto wahrscheinlicher ist es, dass ein Angebot mit einer hohen Relevanz gefunden werden kann. Aus diesem Grund wurden als Datenbasis für den Prototyp die frei verfügbaren Stellenangebote der Stadt New York gewählt. Die öffentliche Verwaltung der Stadt sucht in vielfältigen Gebieten Arbeitskräfte, weswegen verschieden ausgeprägten Lebensläufen passende Stellenangebote zugeordnet werden können. Im Bereich der privaten Arbeitgeber ist das Konzept insofern problematisch, dass das Jobangebot für einzelne Unternehmen typischerweise relativ gering ausfällt. Ein möglicher Lösungsansatz für dieses Problem wäre die Nutzung von Jobportalen für die Erfassung von Stellenangeboten verschiedener Arbeitgeber. Die kombinierte Stellenangebotsbasis würde eine größere Datenmenge, auf welcher die Suche mittels des k-NN-Algorithmus stattfinden kann, darstellen. Im Fall einer Einzelbetrachtung der Arbeitgeber ist jedoch eine spezialisierte Fallbasis denkbar, was in einer effizienteren Stellenangebotszuordnung resultieren kann.

Das Konzept berücksichtigt weiterhin lediglich die Auswahl des Stellenangebotes, welche der Bewerber getroffen hat. Das Ergebnis wird als erfolgreicher Fall in der Fallbasis abgelegt. Zu diesem Zeitpunkt ist jedoch unklar, ob die anschließende Bewerbung erfolgreich ist. Es ist denkbar, dass der Bewerber trotz passender Qualifikationen abgelehnt wird. Das Konzept ermittelt folglich die Eignung bestimmter Stellenangebote aus der Sicht des jeweiligen Bewerbers. Eine allgemeingültige Verbindung zwischen Qualifikationsangaben in Lebensläufen und Qualifikationsanforderungen in Stellenausschreibungen wird nicht hergestellt.

Als weitere Einschränkung ist anzumerken, dass lediglich mit Lebensläufen von Akademikern positive Ergebnisse beim Prototyp erzielt werden konnten. Mit den Lebensläufen anderer Arbeitnehmer konnte der Algorithmus, zumindest in den Datenbeständen der Stadt New York, keine relevanten Angebote herausfiltern, was in der Struktur und dem Inhalt der Lebensläufe einerseits, sowie in der generellen Eignung des k-NN-Algorithmus andererseits begründet liegen kann.

Der k-NN-Algorithmus ist durchaus mit anderen Klassifizierungsalgorithmen austauschbar. Beispielsweise könnten mittels künstlicher neuronaler Netze Lebensläufe klassifiziert werden, um anschließend eine Menge an relevanten Stellenangeboten auszugeben (Ertel 2011).

6 Fazit

In diesem Beitrag wurde ein Konzept für ein modifiziertes CBR-System für Online-Stellenbörsen anhand des Beispiels der Stellenbörse der Stadt New York entwickelt. Durch den Upload eines Lebenslaufs und personalisierter Selektion relevanter Stellenanzeigen sowie deren Präsentation für den Bewerber, kann die Effizienz der Stellensuche gesteigert werden. Die Zuhilfenahme von bereits bearbeiteten Fällen ermöglicht es, den Suchprozess effektiver zu gestalten.

Der entwickelte Prototyp analysiert die Datenquellen anhand von Text Mining und Klassifikation der Wortvektoren mit dem k-Nearest-Neighbor-Algorithmus. Die prototypische Implementierung umfasst folglich die *Lookup*-Phase als Teil des Konzeptes.

Die vom Prototyp erforderten hohen Voraussetzungen an eine Datenbank von Stellenanzeigen stellen grundsätzlich eine Hürde für die Implementierung des Konzeptes bei Arbeitgebern dar. Des Weiteren gilt es, Einschränkungen in der Anwendung aufzuheben. Dazu gehören der Umgang mit Fällen, die keinen Erfolg darstellen, die Kontrolle der Ablage erfolgreicher Fälle und die Schwächen bei der Analyse nicht akademischer Lebensläufe. Letztere Einschränkung könnte durch eine Anpassung der verwendeten Algorithmen erreicht werden. Es bleibt jedoch festzustellen, dass Datensätze mit hohem Informationsgehalt, also detaillierte Lebensläufe, generell besser analysiert werden können. Da nicht jede Stelle eine detaillierte Dokumentation der Qualifikationen erfordert, stellt die Verringerung der Unterschiede in der Qualität der Suchergebnisse eine besondere Herausforderung dar.

7 Literatur

- Aamodt A, Plaza E (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications* 7(1):39-59
- Beierle C, Kern-Isberner G (2014) Methoden wissensbasierter Systeme. 5. Auflage. Springer Fachmedien, Wiesbaden
- Bildat L, Lau D (2012) Mit den falschen Netzen fischen. *Personalwirtschaft* 39(5):50-52
- Chapman DS, Webster J (2003) The Use of Technologies in the Recruiting, Screening, and Selection Processes for Job Candidates. *International Journal of Selection and Assessment* 11(2/3):113-120
- Chowdhury G (2003) Natural language processing. *Annual Review of Information Science and Technology* 37: 51-89
- Dachrodt HG, Engelbert V, Koberski W, Dachrodt G (2014) Praxishandbuch Human Resources: Management – Arbeitsrecht – Betriebsverfassung. Springer Fachmedien, Wiesbaden
- Ertel W (2011) Introduction to Artificial Intelligence. Springer-Verlag, London
- Heyer G, Quasthoff U, Witting T (2012) Text Mining: Wissensrohstoff Text. W3L, Herdecke
- Holm AB (2010) The Effect of E-recruitment On the Recruitment Process: Evidence from Case Studies of Three Danish MNCs. *CEUR Workshop Proceedings* 16(570):91-111