

# **Automatische Extraktion von Daten aus Online Social Networks (OSN) und Eingliederung der Daten in das Five Factor Model (FFM)**



**Hochschule  
für Oekonomie & Management**  
University of Applied Sciences

©FOM Hochschule für Oekonomie & Management

2014

**Tobias Albrecht, Florian Grundl & Florian Hofsäss**

## Zusammenfassung:

Diese Arbeit befasst sich mit der automatischen Extraktion von Nutzerdaten aus Online Social Networks (OSNs) mit anschließender Eingliederung der Daten in das Persönlichkeitsmodell Five Factor Model. Der Fokus der Arbeit liegt auf der Extraktion der Daten und dem zur Verfügung stellen der extrahierten Informationen in Form von granularen Methoden. Anhand dieser Methoden ist es dem Benutzer möglich die für seine Gruppe/Organisation relevanten Eigenschaften mit der individuell angegebenen Gewichtung, in den Vergleich der Persönlichkeiten einfließen zu lassen.

## Einleitung:

Personalabteilungen (= Human Resource / HR) bauen zunehmend mehr auf die Unterstützung von elektronisch gestützten Systemen, um einerseits eine große Masse von Bewerbern zu bewältigen und andererseits, um an festen Kriterien Bewerber selektieren zu können. Darüber hinaus versuchen viele Unternehmen, abseits von Bewerbungsunterlagen weitere Informationen aus Social Networks zu gewinnen, um weitere Erkenntnisse über den Bewerber zu erhalten. Jedoch konzentrieren sich die meisten dieser elektronisch gestützten Systeme auf den Vergleich von fachlichen Anforderungen zu den mitgebrachten Qualifikationen des Bewerbers.<sup>1</sup>

Vor dem Hintergrund, "dass die Zusammensetzung einer Gruppe Auswirkungen auf deren Leistung haben kann."<sup>2</sup>, spielen also nicht nur fachliche Eignung im Vergleich zu fachlichen Anforderungen (P-Job Fit) eine wichtige Rolle. Daher kann es für viele Organisationen und Gruppen interessant sein, eine Gegenüberstellung von persönlichen Eigenschaften des Bewerbers mit den Eigenschaften der Gruppe beziehungsweise der Organisation durchzuführen, um die Effizienz innerhalb eines Systems möglichst hoch zu halten.

Um diese Möglichkeit zu schaffen, wurde für diese Arbeit folgende Perspektive gewählt:

- Externe Perspektive: Die Arbeit ist aus der Sicht des Softwareentwicklers geschrieben, so dass die Gewichtung und Auswahl der diversen Eigenschaften individuell den Ansprüchen des Nutzers überlassen werden kann und Unternehmens/Gruppenunabhängig auf Bewerber angewandt werden kann
- Das Ziel der Arbeit ist eine Gegenüberstellung von persönlichen Eigenschaften eines Bewerbers durch OSN-Daten mit den Eigenschaften einer Gruppe / einer Organisation. Diese Untersuchungsebene nennt man Meso- bzw. Makroebene, da hier die Mikroebene, welche eine Person-Job fit widerspiegelt, nicht betrachtet wird.<sup>3</sup>
- Für die Eingliederung der verschiedenen Eigenschaftsattribute wird ein Persönlichkeitsmodell herangezogen, um die Eigenschaften diverser Bewerber zu vereinheitlichen und somit vergleichbar machen zu können. Als Persönlichkeitsmodell eignet sich das Five-Factor Modell

---

<sup>1</sup> Vgl (Büttner, 2014)

<sup>2</sup> Vgl.(Witte, 2007)

<sup>3</sup> (Büttner, 2014)

sehr gut, da es durch mehrere Skalen und Messtypen wie den Mini Makers von Saucier (1994)<sup>4</sup>, des NEO-FFI von Borkenau/Ostendorf (1993)<sup>5</sup> und den BFMM-D<sup>6</sup> validiert wurde und Messverfahren ermöglichen.

## Traits-Ansätze

Traits-Modelle sind die derzeit vorherrschenden Modelle, um Persönlichkeiten zu beschreiben. Zur Beschreibung werden im Rahmen solcher Modelle Traits, oder zu Deutsch Eigenschaften beziehungsweise Wesenszüge, verwendet. Hierbei besteht die grundlegend Annahme darin, dass Wesenszüge bei Personen stets individuell stark ausgeprägt sind und die Wesenszüge eher konstante als häufig wechselnde Merkmale sind. Gemessen werden Wesenszüge durch ähnliches oder unterschiedliches Verhalten einer Person im Vergleich zu anderen Personen unter gleichen Bedingungen <sup>7</sup> McCrae/Costa beschreiben Traits in Ihrer Arbeit wie folgt :"[...] as dimensions of individual differences in tendencies to show consistent patterns of thoughts, feelings, and actions"

Es wird auch deutlich gemacht, dass diese Merkmale in wesentlichem Zusammenhang zu den Umfeldbedingungen stehen und somit nicht als Konstanten eines Eigenschaftsprofils angesehen werden können. Beispiele für solche relevanten Umfeldbedingungen können beispielsweise die Erwartungshaltungen Dritter, Umfeld, Situation, Stimmung oder Stressfaktor sein.

Nach Allport lassen sich Traits auch als zentrale Einheit der Persönlichkeit bezeichnen, anhand derer unter oben aufgeführten Bedingungen, individuelle Unterschiede im Verhalten beschrieben werden können.<sup>8</sup>

Da Allport Traits mit der menschlichen Persönlichkeit so direkt verknüpft, "[...] no one [...] doubts that underlying the conduct of a mature person there are characteristic dispositions or traits"<sup>9</sup> folgt daraus, unter der Annahme deren Existenz, die Frage, auf wie viele solcher Traits sich die menschliche Persönlichkeit verteilt und wie diese Traits gemessen werden können.

## Von Traits zu Five-Factor-Modell

Die erste Frage die es zu Beginn der Forschung auf diesem Gebiet zu beantworten gab war, wieviele solcher Traits es denn gäbe. Dieser Frage ging Franziska Baumgarten (1933)<sup>10</sup> als Erste nach, nachdem Klages (1926) das erstmals den lexikalischen Ansatz formulierte. Baumgarten sammelte im Rahmen einer selbst erstellten Liste systematisch "persönlichkeitsbeschreibende Wörter" <sup>11</sup>. Ohne große

---

<sup>4</sup> (Saucier, 1994)

<sup>5</sup> (Ostendorf, 1993)

<sup>6</sup> (Ing Weller, 2009)

<sup>7</sup> (Angleitner, Ostendorf, & John, 1990)

<sup>8</sup> (Faillant, 2006)

<sup>9</sup> (Allport, 1937)

<sup>10</sup> (Baumgarten, 1933)

<sup>11</sup> (Faillant, 2006) S. 104

Beachtung dieser Arbeit griffen Allport, der sich wie oben ersichtlich bereits früher schon intensiv mit der Thematik der Traits auseinandersetzte, und Odbert diese Idee auf und erstellten im Folgenden eine Ordnung und Klassifikationen von Begriffen, die später als empirische Basis diente <sup>12</sup>

Nachdem Allport und Odbert sich des "Webster's New International Dictionary" bedienten, um entsprechende Traits zu filtern, kristallisierten sich nach mehreren Clusterebenen 4 große Kategorien heraus:

- **neutrale Persönlichkeitseigenschaften (ängstlich, schüchtern, mutig, aggressiv)**
- **Stimmungen, Aktivitäten und Gemütszustände (wütend, freudig, mitteilungsbedürftig)**
- **wertende Begriffe (nutzlos, exzellent)**
- **Adjektive zur Beschreibung von physischen Qualitäten und Talenten (handwerklich)**

Um die bis hierhin 4.500 neutralen Persönlichkeitseigenschaften zu einer brauchbaren und verwendbaren Anzahl von Merkmalen zu reduzieren bedürfte es eines weiteren Clusterings von Begriffen.

Gemäß seiner Aussage: "[...] the most potent method of attacking the tangle is to work out correlation coefficients between the inconveniently multitudinous variables abounding in the subject and to seek some smaller number of behind the scenes or underlying variables, known as factors"<sup>13</sup> konnte er anhand von mathematischen Kriterien 35 Traits herausarbeiten.

Es folgten weitere, voneinander unabhängige, empirische Studien, welche als Basis die 35 Traits von Catell verwendeten, die wiederholt fünf Faktoren als Ergebnis präsentierten.

Autor(en)	Faktor I	Faktor II	Faktor III	Faktor IV	Faktor V
Fiske (1949)	Confident Self-Expression	Social Adaptability	Conformity	Emotional Control	Inquiring Intellect
Tupes & Christal (1961)	Surgency	Agreeableness	Dependability	Emotional Stability	Culture
Norman (1963)	Surgency	Agreeableness	Conscientiousness	Emotional Stability	Culture
Digman & Takemoto-Chock (1981)	Extraversion	Friendly Compliance	Will to Achieve	Ego Strength (Anxiety)	Intellect
Goldberg (1981, 1989)	Surgency	Agreeableness	Conscientiousness	Emotional Stability	Intellect
<b>McCrae &amp; Costa (1985)</b>	<b>Extraversion</b>	<b>Agreeableness</b>	<b>Conscientiousness</b>	<b>Neuroticism</b>	<b>Openness to Experience</b>
De Raad et al. (1988)	Extraversion	Agreeableness vs. Cold-Heartedness	Conscientiousness	Emotional Instability	Culture

14

<sup>12</sup> (Fritz, Angleitner, & John, 2004)

<sup>13</sup> (Cattell, 1946) S. 272 & (Winter & Barenbaum, 2003)S. 12.

<sup>14</sup> (Angleitner, Ostendorf, & John, 1990), S. 72.

Diese Faktoren werden üblicherweise wie folgt benannt und nummeriert.

- (I) Begeisterungsfähigkeit / Extraversion ( Surgency / Extraversion),
- (II) Verträglichkeit (Agreeableness),
- (III) Gewissenhaftigkeit / Verlässlichkeit (Conscientiousness/Dependability),
- (IV) emotionale Stabilität / Neurotizismus (emotional Stability / Neuroticism).
- Faktor V wurde sowohl durch Kultur (z.b De Raad et.al (1988)),
  - durch den Verstand (z.b. Digman & Takemoto-Chock (1981) ; Peabody & Goldberg (1989))
  - als Offenheit (z.b. McCrae & Costa (1985))

bezeichnet.

Bis dato haben sich die Beschreibungen von John McCrae und Paul Costa durchgesetzt. Auch die deutsche Ausarbeitung für das Inventar zur Messung der Big Five von Ostendorf und Borkenau baut auf das amerikanische Modell von McCrae und Costa auf. Folglich sind diese fünf ausgearbeiteten Faktoren von McCrae und Costa die am häufigst verwendeten in der Literatur.<sup>15</sup>

Dementsprechend wird auch im Rahmen dieser Arbeit Bezug auf diese fünf Faktoren genommen

Alle der fünf Faktoren können auf einer Skala gemessen werden, wobei die niedrigste Skalenzahl genau das Gegenteil wie eine hohe Skalenzahl über den Getesteten aussagt.

## Extroversion

Die Extroversion beschreibt in welchem Maß der Befragte eine offene Haltung nach außen aufweist. Menschen mit einer hohen Punktzahl auf dieser Skala beschreiben sich selbst als aufgeschlossen, aktiv, kontaktfreudig, gesellig und optimistisch. Sie sind eher direkt, fröhlich und heiter als in sich gekehrt und distanziert. Extrovertierte Personen fühlen sich vor allem in Gruppen und gesellschaftlichen Events wohl und haben meist viele Freunde.

Menschen mit niedrigen Punktzahlen auf der Skala der Extroversion gelten eher als in sich gekehrte Menschen die sich in kleinen Gruppen oder alleine wohler fühlen. Daher haben Sie meist einen eher kleinen Freundeskreis. Sie werden auch als weniger sozial beschrieben, da Sie den Kontakt zu Mitmenschen eher scheuen als Leute mit hohen Punktwerten. Folglich sind Adjektive wie zurückhaltend, unabhängig und ausgeglichen häufig eine zutreffende Beschreibung. Entgegen vieler Meinung sind introvertierte Menschen nicht unglücklich alleine oder in kleinen Gruppen zu sein, es ist vielmehr ihr Wunsch.<sup>16</sup>

---

<sup>15</sup> (Faullant, 2006), S. 112

<sup>16</sup> (Suda & Fouts, 1980)

## Verträglichkeit

Dieser zweite Faktor misst unter anderen wie schnell und vor Allem wie gut sich der Betroffene mit Menschen in seinem Umfeld versteht und eine Beziehung zu diesem aufbaut. Rui-Ping Zhang und Li Tsingang formulieren das in Ihrer Ausarbeitung zu den Faktoren der Big Five wie folgt:  
"Specifically, compared with individuals low on agreeableness, highly agreeable individuals may have higher levels of well-being because they can form and maintain positive and close relationships with others"<sup>17</sup>

Befragte mit einer hohen Punktzahl haben es dem zur Folge einfach neue Kontakte zu knüpfen und sich konfliktlos in eine neue Gruppe einzufügen und dort akzeptiert zu werden, da Sie als verständnisvoll, wohlwollend und mitfühlend gelten und stets bemüht sind anderen zu helfen. Darüber hinaus zeichnen Sie sich durch Vertrauen, Kooperativität und Nachgiebigkeit aus. Menschen mit niedrigen Werten auf dieser Skala haben es häufig schwerer sich in einer Gruppe einzufügen, da "verträgliche Menschen beliebter sein dürften als unkooperative"<sup>18</sup>. Nicht zu verwechseln ist dieser Faktor mit dem Vertreten von eigenen Meinung. Das Vertreten einer eigenen, von dem Rest der Gruppe abweichenden Meinung, lässt in der Regel keinen Schluss auf den Skalenwert dieses Faktors zu, da die Unstimmigkeiten nicht immer der Persönlichkeit des Betroffenen geschuldet sind, sondern dessen Ansicht zu einem speziellen Thema.

## Gewissenhaftigkeit

Die dritte Dimension Gewissenhaftigkeit hat mit dem Faktor der Verträglichkeit wohl die breite Akzeptanz von Mitmenschen gemeinsam. Sowohl gewissenhafte wie auch verträgliche Menschen, werden schneller und eher in einer Gruppe akzeptiert als schlampige und destruktiv eingestellte Menschen.

Der Faktor der Gewissenhaftigkeit beschreibt Menschen mit hohen Skalenwerte stets als gut organisiert, ambitioniert, fleißig, strebsam, ausdauernd und willensstark. Desweiteren für Mitmenschen angenehme und wichtige Eigenschaften könnten Zuverlässigkeit, Pünktlichkeit, Ordentlichkeit und Genauigkeit sein.<sup>19</sup>

Vorallem für Arbeitgeber könnte dieser Faktor von Wichtigkeit sein, da dieser Faktor den direktesten Schluss auf die Ernsthaftigkeit im Rahmen der Ausübung eigener Arbeit zulassen und auch Schlüsse auf akademische Leistungen und den beruflichen Werdegang zulassen.

---

<sup>17</sup> (Zhang & Tsingan, 2013)S. 3

<sup>18</sup> (Faillant, 2006), S. 113

<sup>19</sup> (Digman, 2004): S. 75

Zusammen mit Verträglich ist dieser Faktor einer der fünf Faktoren, welcher eine Wertung in der Definition aufweist. Denn niedrige Punktzahlen auf der Skala beide Faktoren werden eher als unerwünscht eingeschätzt, als das bei den anderen drei Faktoren der Fall ist. Dies könnte der Grund sein, so die Vermutung von McCrae und John, warum diese Faktoren lange bei Untersuchungen ignoriert wurden: " [...], but in fact, both represent objectively observable dimensions of individual differences"<sup>20</sup>

## Neurotizismus

Der Faktor Neurotizismus bildet den Grad der emotionalen Reaktion eines Menschen auf eine bestimmte Situation ab. Menschen mit einem hohen Wert auf der Skala sind schnell nervös und eher pessimistisch gestimmt. Selbst schätzen sie sich als sehr emotional, ähnlich, unsicher, aber auch als sehr mitfühlend ein. Von emotionalen Situation sind sie schneller ergriffen, erschüttert oder überwältigt als das bei Menschen mit niedrigen Skalenwerten der Fall ist. Menschen mit niedriger Ausprägung in diesem Bereich gehen eher analytisch, sachlich und gelassen mit der Situation um und distanzieren sich von der emotionalen Ebene der Situation. Die Stärke der Ausprägung ist bei den jeweiligen Menschen vor allem in Stresssituationen bemerkbar. Während Menschen mit hohen Werten eher schwierig mit der Situation klarkommen und sich davon überwältigen lassen ist es Menschen mit niedriger Ausprägung möglich strukturiert an diese Situation heranzutreten und diese möglichst effektiv zu lösen.<sup>21</sup>

Neben Extroversion gilt dieser Faktor als einer der meist akzeptieren Faktoren wie auch Digman in seiner Arbeit schildert: „There is general agreemant that Dimension I is Extraversion and that IV represents the [...] Dimension IV is usually refered to as Neuroticism vs Emotional Stability. Heren, then, are the original Eysenck „Big Two“[...] „<sup>22</sup>

## Offenheit für Erfahrungen

Die Skala Offenheit für Erfahrungen misst die Tendenz sich in neuen Aktivitäten zu Engagieren und bestehende Verfahren und Handhabungen mit neuen Ideen oder Erfahrungen zu verändern.<sup>23</sup>

Menschen mit einer hohen Ausprägung in diesem Faktor sind also dementsprechend eher Revoluzzer während Menschen mit geringer Ausprägung sich eher an gewohntes anpassen und sich eher mit bereits etabliertem zufriedengeben als eigene Verbesserungsvorschläge einbringen.

Unter dem Strich soll bis hierher deutlich werden, dass im Rahmen dieser fünf Faktoren viele Persönlichkeitszüge erfasst werden, die für HR sehr interessant sein können, um die Teamperformance hoch zu halten oder zu erhöhen.<sup>24</sup>

---

<sup>20</sup> (McCrae & John, 1992) S. 97

<sup>21</sup> (Faullant, 2006), S114

<sup>22</sup> (Digman J. , 1990) S. 422

<sup>23</sup> (Furnham, Dissou, Sloan, & Chamorro-Premuzic, 2007)S. 6

<sup>24</sup> (Witte, 2007) S. 1

Ein valider PG-Fit oder ein valider PO-Fit sind daher ein großer Vorteil um schon im Vorfeld selektieren zu können, welche Bewerber in die engere Auswahl und somit in die Gruppe rund um die ausgeschriebene Position sowie in die Organisation passen.

Beispielsweise ist der extrovertierte, vertriebsorientierte und lebensfroher 25-jährige trotz fachlicher Eignung, in den Reihen der sehr konservativen, in sich gekehrten und sehr gewissenhaft arbeitenden Angestellten nicht für die Stelle geeignet, obwohl er alle fachlichen Kriterien aufweist, die für diese ausgeschriebene Stelle notwendig sind.

Offen ist bis hierhin lediglich noch, wie die Daten in Kombination mit OSNs erhoben werden können.

## Literature Review „Social Media Datenextraktion“

Um die Frage beantworten zu können wie man für die Nachwuchsgewinnung relevanten Daten aus Online Sozialnetzwerken extrahieren kann ist es notwendig neben den Extraktionstechniken auch die derzeitigen Maßnahmen/Möglichkeiten des E-Recruiting in sozialen Netzwerken zu beleuchten.

Wenn Unternehmen bereits aktiv Soziale Netzwerke für die Nachwuchsgewinnung anerkennen und diese in Ihren Recruiting Prozessen und Konzepten nutzen ist die Integration einer automatisierten Datenextraktion mit evtl. elektronischer Auswertung leichter umsetzbar.

E-Recruiting ist die Abkürzung für Electronic Recruiting und steht für die automatisierte computergestützte bzw. internetbasierte Nachwuchsgewinnung<sup>25</sup>. Ziel dieser Art von Personalbeschaffung ist das ökonomische Prinzip des Minimalprinzips. Der benötigte Personalbedarf soll mit möglichst geringem Mitteleinsatz gedeckt werden. Die eingesetzten Systeme sollen die Rekrutierung eines fachlich, persönlich passenden Bewerber mit einem möglichst guten Aufwand-Ertragsverhältnis ermöglichen.<sup>26</sup>

## Aktueller Stand Social Media Recruiting

Der für diese Arbeit relevante Teil des E-Recruiting ist der Teilbereich „Social Media Recruiting“. Hierbei handelt es sich um die Bewerbergewinnung mithilfe von Online Sozialen Netzwerken (Facebook), Kurznachrichtendienste/Blogs (Twitter) oder Business Online Netzwerken (XING, LinkIN).

Da sich diese Wissenschaftliche Arbeit mit dem Sozialen Netzwerk Facebook beschäftigt werden die derzeitigen Möglichkeiten auf dieser Plattform kurz vorgestellt.

Die einfachste Möglichkeit E-Recruiting auf Facebook zu betreiben ist das Erstellen einer Firmenseite. Auf diesen Facebookseiten haben Firmen die Möglichkeit sich der Nutzergemeinde zu präsentieren,

---

<sup>25</sup> <http://www.personalwirtschaft.de/de/html/lexikon/27/E-Recruitment>, 22.08.2014

<sup>26</sup> (Schulten, Mertens, & Horx, 2012), S.330f

neue Produkte vorstellen oder mit potenziellen Kunden direkt in Kontakt zu treten. Die Basis Version einer solchen Seite ist kostenlos. Möchte das Unternehmen Werbung bei anderen Nutzern einblenden lassen oder mit Analyse Werkzeugen das Nutzerverhalten nachvollziehen können ist eine Gebühr an das Soziale Netzwerk zu entrichten.

Bei den angebotenen Analysen handelt es sich um Nutzerstromanalysen. Hierbei wird das Nutzerverhalten auf den Firmenseiten analysiert. Wie viele Fans hat das Unternehmen, wie oft bekomme es ein „gefällt mir“ von seinen Fans zu seinen Beiträgen, welche Inhalte der Seite werden wie oft und in welcher Reihenfolge betrachtet. Auch externe Verlinkungen auf die Unternehmenseite oder der Download von Broschüren lässt sich somit nachvollziehen.

Als spezielle Untergruppe der Facebookpages kann man die sogenannten Karrierepages betrachten. Dabei handelt es sich um Firmenseiten in sozialen Netzwerken welche ausschließlich das Thema Nachwuchsgewinnung, die Weiterbildungsangebote und aktuelle freie Stellen im Unternehmen behandeln.<sup>27</sup>

Die Anzahl dieser Karrierepages ist in den vergangenen Jahren von 100 Stück im Jahr 2011 auf über 340 im Jahr 2013 angestiegen<sup>28</sup>. Die Anzahl an interessierten Nutzern welche dem Angebot mit „gefällt mir“ folgen hat sich mehr als verzehnfacht.

Um sich aus der Masse der Karriere- bzw. Firmenpages abzugrenzen und dem künftigen Bewerber besonders attraktiv zu erscheinen verwenden Unternehmen das Konzept des „Employer Branding“.

Hierbei wird eine spezielle Untermarke (z.B. SIEMENS Karriere, Karriere Bundeswehr) des Unternehmens die vor allem als Arbeitgeber auftritt geplant, weiterentwickelt und kontrolliert. Die Differenzierung von anderen Arbeitgebern und die externe Darstellung einer Unternehmenskultur stehen hierbei im Mittelpunkt.<sup>29</sup>

Das Social Media Recruiting vor allem bei Firmen in der Informationstechnischen Branche auf Interesse stößt und auch schon Konzepte ausgearbeitet und angewandt werden zeigt eine Studie des Branchenverbandes BITKOM aus dem Mai 2014. Onlineplattformen wie Facebook, Twitter verwenden bereits 69% der Unternehmen und ein Viertel der Verbandsmitglieder setzt bereits auf Employerbranding.

---

<sup>27</sup> <https://de-de.facebook.com/business/overview>, 22.08.2014

<sup>28</sup> Knabenreich H. Vergesst Facebook! Von Vertrauen, Feingefühl und der Interaktion auf Facebook-Karriereseiten 13.10.2013, <http://personalmarketing2null.de/2013/10/13/vergesst-facebook-von-vertrauen-feingefuehl-interaktion-facebook-karriereseiten/>, 27.04.2014

<sup>29</sup> (Sponheuer, 2010). S.268ff

## Entwicklung Karriere-Pages

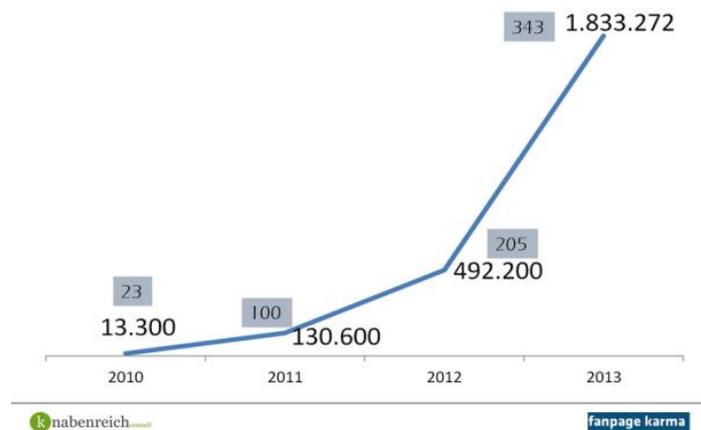


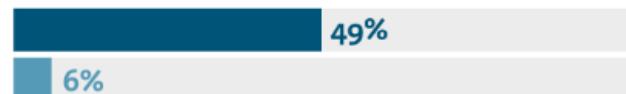
Abbildung 1: Knabenreich, H. Vergesst Facebook! Von Vertrauen, Feingefühl und der Interaktion auf Facebook-Karriereseiten  
13.10.2013, <http://personalmarketing2null.de/2013/10/13/vergesst-facebook-von-vertrauen-feingefuehl-interaktion-facebook-karriereseiten/>

### Young Professionals werden über Online-Kanäle angesprochen <sup>1</sup>

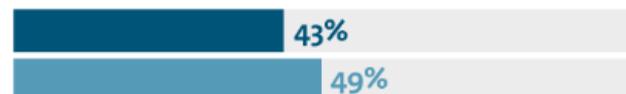
Verstärkter Einsatz von Online-Rekrutierungskanälen



Kreativere Anzeigen für Young Professionals



Neue Ansprache-Formate eingesetzt (z.B. Recruiting Videos)



Verstärkte Investitionen in das Employer Branding



**Bitkom**  
Research GmbH

<sup>1</sup> Basis: Alle befragten Unternehmen, die sich in den letzten 12 Monaten damit beschäftigt haben, inländische oder ausländische Young Professionals zu rekrutieren  
<sup>2</sup> Basis: Alle befragten Unternehmen, die in den letzten 5 Jahren die Rekrutierungsstrategie bewusst verändert haben  
Quelle: BITKOM Research 2014

Abbildung 2: Rekrutierung von Young Professionals, 22.08.2014,  
[http://www.bitkom.org/files/documents/Grafik\\_PK\\_Young\\_Professionals\\_20\\_05\\_2014.jpg](http://www.bitkom.org/files/documents/Grafik_PK_Young_Professionals_20_05_2014.jpg)

Da Unternehmen diverse Online Medien z.B. Soziale Netzwerke bereits zum Informieren und rekrutieren von potentiellen Bewerbern nutzen ist es nur auch folgerichtig, dass versucht wird über diese Werkzeuge auch Informationen vom Bewerber selbst zu bekommen.

Als günstige und schnelle Lösung verwenden hier Arbeitgeber vor allem die gängigen Suchmaschinen wie google oder yahoo. Hierbei kann mit dem Namen des Bewerbers in Kombination mit Wohnort oder anderen persönlichen Daten mit einfachen Mitteln Ergebnisse erzielt werden. Will das

Unternehmen eine professionellere Lösungen so gibt es Bewerbertrackingsoftware welche auch auf gängige Suchmaschinen zugreifen und Auswertungen erstellen kann. Beispielhaft kann man die Toolsammlung von Zoho anführen<sup>30</sup>.

Bei sozialen Netzwerken und Web 2.0 Lösungen ist es ungemein schwieriger. Zeng et al 2013 S13 beschreibt die Kommunikation in sozialen Netzwerken als „dialogorientierten, verteilten Modus der Inhaltsgenerierung, Verbreitung und Kommunikation“. Er sieht einen Paradigmen Wechsel in der Art der Online Kommunikation und Kollaboration. Die Inhalte werden nicht nur generiert sondern geteilt, kommentiert, geliked um eigene Standpunkte ergänzt. Zudem stehen die Informationen durch die weltweite Vernetzung und größer werdender Bandbreiten immer mehr Menschen zur Verfügung.<sup>31</sup>

Durch die wachsende Verbreitung von Mobiltelefonen mit Internetzugang und entsprechender Netzausbaumaßnahmen der Provider werden die Menschen zukünftig auch Unterwegs ständig auf soziale Netzwerke und Onlineinhalte zugreifen und mitwirken können. Hierbei werden auch die Lokations bezogenen Informationen und Dienste der Mobilgeräte stärker in die Nutzung von sozialen Netzwerken einfließen und für zusätzliche Datenmengen sorgen<sup>32</sup>.

## Herausforderungen der Social Media Datenextraktion

Da es sich bei den Informationen in den Datenbanken der Social Media Anbietern um eine unsortierte Menge an Daten handelt ist eine herkömmliche Analyse nur schwer möglich.

Ähnlich wie in anderen Fällen von sehr großen Datenmengen wir hier eine spezielle computer basierte algorithmische „Big Data“ Auswertung notwendig.

In Fall von sozialen Netzwerken wurde hierbei von Lev Manovich der Begriff „Big Social Data“ verwendet. Er sieht in der „Big Data“ Analyse die Möglichkeit aus Online Social Networks die Meinungen, Ideen und Erwartungen von Millionen Menschen auswerten zu können. Der nutzergenerierte Inhalt (z.B. Videos, Blogeinträge, Likes, geteilter Inhalt) sind dafür vorhanden. Um einen Zugriff auf diese Daten zu gewährleisten müssen die Anbieter jedoch auch offene und dokumentierte API (Application Programmable Interfaces) zur Verfügung stellen.<sup>33</sup>

Da es sich bei den zu erwartenden Daten um eine große Anzahl an komplexer und hoch dynamischer Art handelt können diese nicht einfach von traditionellen Datenbankwerkzeugen und Visualisierungslösungen ausgewertet werden. Es ist zu empfehlen einen disziplinären Ansatz bei der Auswertung zu verfolgen. Einerseits müssen die mathematisch, naturwissenschaftlichen Forscher (Informatiker, Statistiker, Mathematiker) an der technischen Lösung arbeiten andererseits kann eine Auswertung ohne die Fachbereiche der Sozialwissenschaften (Soziologie, Kommunikationswissenschaft) und Wirtschaftswissenschaften (Betriebswirtschaft) nicht erfolgen. Die extrahierten Informationen weisen sowohl Eigenschaften von strukturierten als auch unstrukturierten Daten auf.

---

<sup>30</sup> <https://www.zoho.com/recruit/features.html>, 22.08.2014

<sup>31</sup> (Zeng, Chen, Li, & Lusch, 2010):13–16

<sup>32</sup> (Dang-Xuan & Stieglitz, 2014)

<sup>33</sup> (Manovich, 2012)

Strukturierte Daten	Unstrukturierte Daten
<ul style="list-style-type: none"> <li>• Daten in Nutzerprofilen</li> <li>• Räumlich/zeitliche Daten</li> <li>• Thematische Daten z.B. „Likes“; „Follower“; „Inhalt Teilen“</li> <li>• Daten aus gekoppelten Anwendungen/Apps z.B. Facebookspiele</li> </ul>	<ul style="list-style-type: none"> <li>• Nutzergenerierter Text z.B. Kommentare, „Pinnwand-posts“</li> <li>• Audio/visuelles Material</li> </ul>

*Tabelle 1: Aufteilung Strukturierte und unstrukturierte Daten*

Neben der unterschiedlichen Datencharakteristika ist vor allem der ständige Wandel und die Dynamik der Kommunikationsstile in sozialen Netzwerken ein Problem und kaum voraussehbar. Die Einführung von Hashtags, Emoticons, Abkürzungen ala ROFL können als Beispiele hierfür genannt werden.

Im Rahmen von „Big Social Data“ muss ebenfalls das Thema Datenschutz betrachtet werden. Welche Informationen benötige ich wirklich und welche sind nicht relevant und können gelöscht werden. Die zur Verarbeitung gespeicherten Daten müssen gegebenenfalls anonymisiert und immer gegen Fremdzugriff gesichert abgespeichert werden.

Ein weiteres Problem können die zuvor erwähnten API der Sozialen Netzwerke werden, da sich die Netzwerke designtechnisch immer wieder an neue Trends anpassen und um neue Funktionen ergänzt werden, um für die Nutzer interessant zu bleiben können die APIs Veränderungen unterlegen. Während an einem Tag noch Daten extrahiert werden konnten ist am darauf folgenden kein Abruf der Daten mehr möglich. Ein ähnliches Problem entsteht bei Textmining und Parsing von HTML-Code. Die Skripte können bei Designanpassungen zu falschen, inkonsistenten Datensätzen führen.

## Das Social Media Analytics Framework

Um bei der Extraktion und Analyse von Socialnetwork Daten eine vorgegebenen Baukasten zu haben wurde von Stefan Stieglitz und Linh Dang-Xuan das „Social Media Analytics Framework“ entworfen.

Hierbei muss man zunächst die verschiedenen Forschungsbereiche und Bezugsrahmen für welche das Framework angewandt werden kann festlegen. Beispielhaft wird hier das Innovationsmanagement, Reputationsmanagement oder Stakeholdermanagement genannt. Das Feedback von Kunden zu meinen Produkten, dessen Ideen zur Verbesserung und das Aufspüren von Trends, Marktlücken meines Produktes wahren Teile des Innovationsmanagements. Im Falle dieser Arbeit liegt der Fokus auf dem Thema Social Media Recruiting. Genauer betrachtet geht es um ein Recommender System welches mit Hilfe von sozialen Netzwerken des PO-Fit des Bewerbers sicherstellen soll. Als Grundlage dient Ricardo Büttners „Framework für Recommender Systems“ bei welchem hier die Makro/Meso Ebene betrachtet

des P-OE fits betrachtet wird.<sup>34</sup>

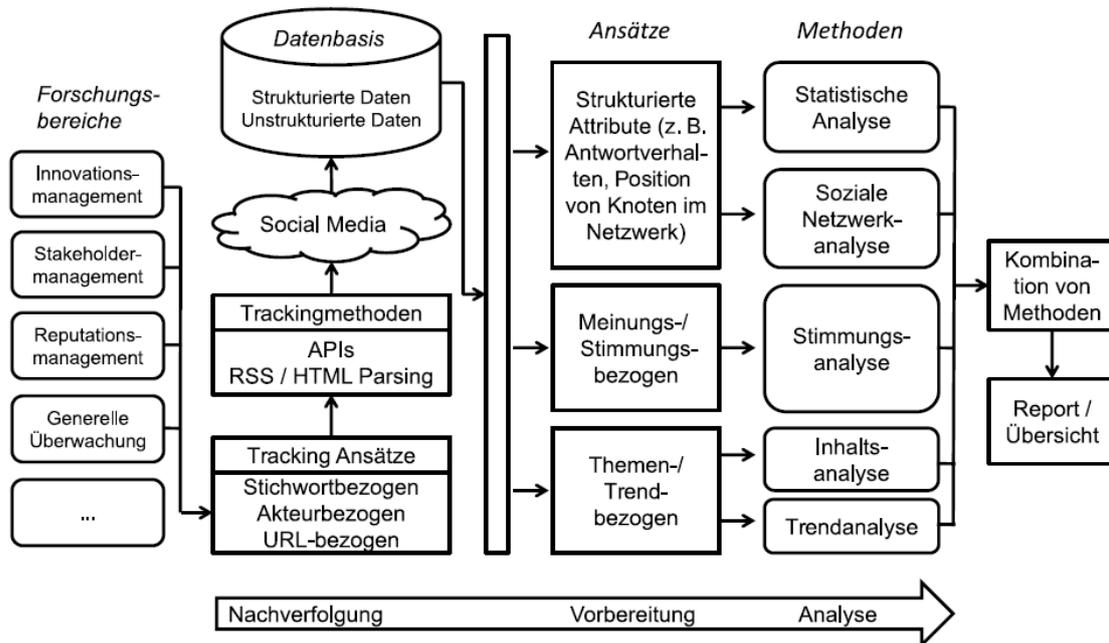


Abbildung 3: Social Media Analytics Framework nach Stieglitz, Duang-Xuan (2013)

Nach dem Festlegen des Forschungsbereiches wird der Tracking Ansatz gewählt werden. Es gibt stichwortbezogene, akteursbezogene und URL-bezogene Ansätze. Hierbei kommt es bei der Auswahl auf das jeweilige Soziale Netzwerk und dessen Gegebenheiten an. Ebenfalls vom sozialen Netzwerk sind die Trackingmethoden abhängig. Die großen Amerikanischen Netzwerke Facebook und Twitter verfügen über entsprechende API und können darüber ausgelesen werden. Andere kleinere Netzwerke wie XING stellen diese nicht zur Verfügung und ein man muss Textmining und HTML Parsing zur Datenextraktion benutzen.

In Anschluss an die Extraktion müssen die strukturierten Daten Evaluert und bereinigt werden. Bei der Evaluierung wird geprüft ob die Daten Inkonsistenzen enthalten was auf einen fehlerhaften Export hindeuten kann. Eine Filterbasierte Entfernung von Spam sollte ebenfalls vorgenommen werden. Es folgt die Normalisierung der Daten und das Abspeichern in ein Datenbankmanagement System.

Aus der Sicht des Qualitätsmanagement sollten diese Daten dann anschließend ebenfalls auf Inkonsistenzen und Redundanzen geprüft werden.

Nach dem erfolgreichen Datenimport kann man bei den Daten von FACT Daten sprechen, diese gilt es im Anschluss zu analysieren. Hierfür kennt das Social Media Analytics Framework verschiedene Ansätze.

Unter dem Ansatz der Skstrukturierten Attribute werden die vorhadenen FACT Daten, wie die Anzahl Facebook Freunde, Anzahl Pots, Anzahl geteilter Inhalte ausgewertet. Die Analyse kann hierbei durch eine Statistische Anaylse erfolgen. Man versucht hierbei durch Mathematische Formeln die Gemeinschaft der Nutzer in Cluster mit bestimmten Eigenschaften einzuteilen. Der gesamte Datenbestand wird auf Gemeinsamkeiten durchsucht und statistisch bewertet so entstehen Verschiedene Nutzertypen. Vergleichbar ist hier der „Big Data“ Teilbereich einer

<sup>34</sup> (Büttner, 2014)

Kundensegmentierung. Facebooknutzer oder in diesem Fall der Bewerber werden in eine dieser statistischen Gruppen eingeordnet. In Bezug auf eine Five Factor Analyse wäre eine Clusterung im Persönlichkeitsfeld der Extoversion möglich. Die Anzahl der Pinwandeinträge, Anzahl an Freunden Des Nutzers könnte in Verhältnis zum Durchschnitt in verschiedene Nutzertypen aufgliedert werden.

Als zusätzliche Methode bietet sich die Soziale Netzwerk Analyse an. Diese beschäftigt sich mit den Beziehungen zwischen Nutzer, Organisationen, Firmen, Interessengruppen indem die Struktur der Verbindungen analysiert wird<sup>35</sup>.

Eine Zuordnung zu politischen Gruppen, das Finden von Meinungsführer sowie die Einteilung in Freundeskreisen/Cliquen/Fangruppen ist somit möglich. Die Mathematischen Werkzeuge hierzu sind vor allem in der Graphentheorie zu finden, aber auch die Clusteranalyse kann hierbei als relevant erachtet werden.

Um eine Inhaltanalyse durchzuführen sind Text Mining Methoden zu benutzen. Programme scannen hierbei den Text und versuchen mittels einer künstlichen Intelligenz den Sinn des Geschriebenen zu erschließen. Meist reagieren diese Systeme auf Schlüsselwörter und führen dann Rückschlüsse/Aktionen aus. Vergleichbar mit Text to speech Systemen oder Sprachsteuerungen im Automobilbau.

Soll eine Rückmeldung zu einem Thema ausgegeben werden, um festzustellen wie ein Personenkreis zu diesem steht, bietet das Social Media Analytics Framework die Stimmungsanalyse als Methode an. Diese setzt wie die Inhaltsanalyse auf Textmining, hierbei wird jedoch auf die positive, negative oder neutrale Aussage des Geposteten geachtet. Hierbei wird von der Sentimentausrichtung oder auch Opinion-Mining einer Aussage gesprochen. Diese kann klassisch mit Wörterbuch Analyse oder mittels selbstlernenden Systemen/Künstliche Intelligenz durchgeführt werden. Beim Thema P-OE fit würde sich die Wörterbuchmethode anbieten. Der nutzergenerierte Inhalt wird auf positive oder negative Wörter untersucht, auch eine Kombination mit neutralen Wörtern (z.B. „Arbeit“AND“krank“OR“Arbeit“AND“schlecht“) überprüft um z.B. die Gewissenhaftigkeit zu überprüfen.

Sollen zukünftige Aussagen getroffen werden, gibt es das noch junge Feld der Trendanalyse. Die statische Auswertung basiert hier laut Stefan Stieglitz „versteckten Markov-Modellen, bei denen die Themenbeobachtung bei solchen Modellen trainiert wird, die wiederum in einer Bibliothek zur Vorhersage eben dieses Themas gespeichert werden.“ In Falle des E-Recruiting ist es interessant wie sich die Anzahl an Studierenden und zukünftigen Absolventen entwickelt. Im Bereich P-OE fit sind diese Methoden aber nicht relevant.

Die Methoden sind für sich allein betrachtet sehr spezialisiert und liefern bei den praktischen Anwendungsfällen nur einen Bruchteil des gewünschten Ergebnisses aus. Um bestmögliche Aussagen treffen zu können, müssen die Methoden kombiniert eingesetzt werden. Darum muss vor Beginn der Auswertung ein Katalog zusammengestellt werden, welche Methoden einzusetzen sind.

Im Falle des P-OE fits könnten dies die Statistische Analyse, Soziale Netzwerk Analyse und Inhaltanalyse mit eingeschlossener Stimmungsanalyse sein.

Nach dem die Analyse angeschlossen wurde sind diese Ergebnisse auf Konsistenz zu evaluieren und anschließend in einem Report darzustellen. Es gibt verschiedene Möglichkeiten diese Darstellung zu realisieren. Eine texturelle Darstellung, eine graphische Darstellung mit Diagrammen, Graphen oder

---

<sup>35</sup> (Scott & Carrington, 2011)

eine Kombination aus Beidem bietet sich hierbei an.

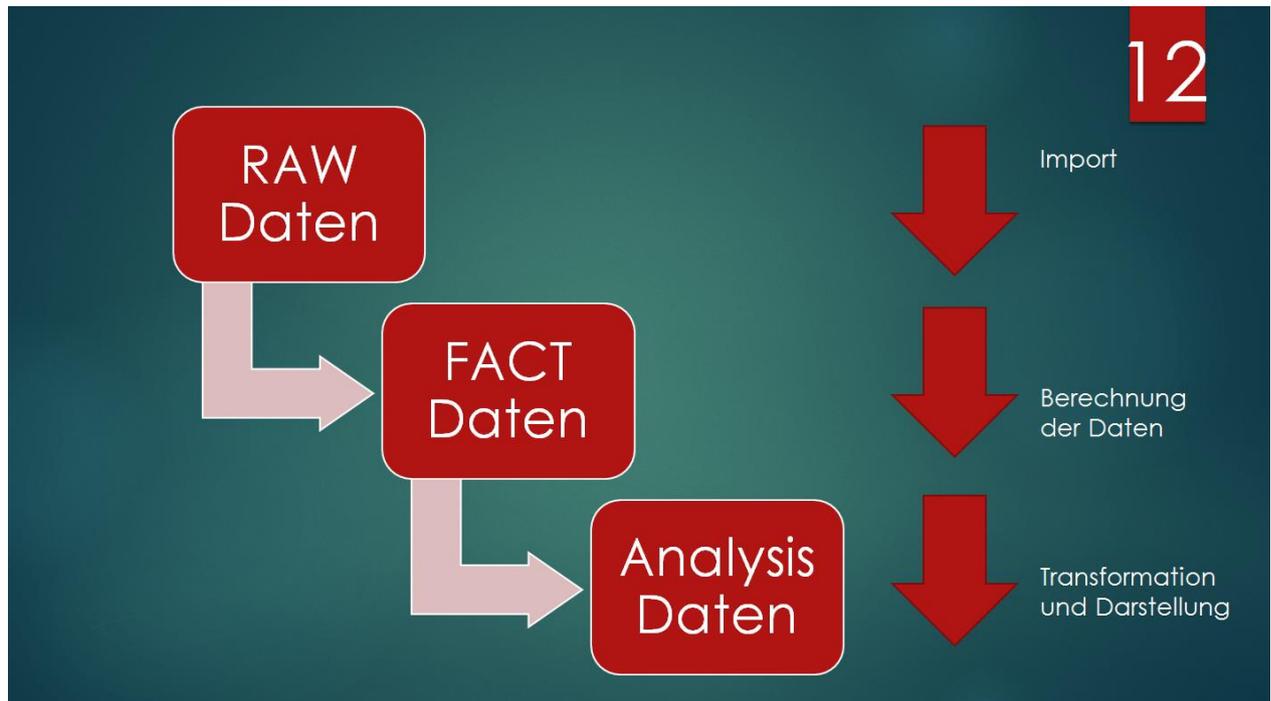


Abbildung 4: Datenverarbeitung bei P-OE fit

Abschließend ist zu konkludieren, dass die trotz der Komplexität der Datenextraktion und Evaluierung aus sozialen Netzwerken mit dem Social Network Analytics Framework ein nützliches Werkzeug zur Verfügung steht.

## Allgemeine technische Beschreibung der Anforderungen

Benötigt wird eine Applikation, welche in 5 einzelne Parts unterteilt werden kann. Diese Teile sind Benutzerdaten einlesen, statistische Daten erzeugen, statistische Daten exportieren, Five Factor Daten importieren und grafische Vergleichbarkeit von Personen ermöglichen.

### Benutzerdaten einlesen

Bereitstellung einer Internetseite, die die Autorisation zwischen Facebook und der Applikation übernimmt und eines Daemons, welche die Benutzerdaten in ein Datenbankschema importiert.

### Statistische Daten erzeugen

Ein Daemon, welcher neu importierte Benutzerdaten erkennt und aus diesen auswertbare Rangvariablen sowie metrisch und kategoriale Werte erzeugt.

### **Statistische Daten exportieren**

Ein Daemon, welcher statistische Daten in ein Verzeichnis ablegt und eine Möglichkeit zur späteren Validierung bereitstellt.

### **Five Factor Daten importieren**

Ein Daemon, welcher alle Dateien innerhalb eines Verzeichnisses in ein Datenbankschema importiert und diese auf Validität prüft.

### **Grafische Vergleichbarkeit von Personen**

Bereitstellung einer Internetseite, die nur mit einem Login einsehbar ist und eine Möglichkeit bietet Five Factor Daten grafisch miteinander zu vergleichen.

### **Projektplan / Milestones**

Durch die klar definierten Teilbereiche der Applikation und da diese aufeinander Aufsetzen werden diese als Milestones betrachtet und mit einer Deadline versehen.

Bezeichnung	Bis wann	Verantwortlicher
Einarbeitung Facebook API	01.06.2014	Tobias Albrecht
Entscheidung Setup / Programmiersprache / Framework	01.08.2014	Tobias Albrecht
Benutzerdaten einlesen	10.08.2014	Tobias Albrecht
Statistische Daten erzeugen	15.08.2014	Tobias Albrecht
Statistische Daten exportieren	17.08.2014	Tobias Albrecht
Five Factor Daten importieren	20.08.2014	Tobias Albrecht
Grafische Vergleichbarkeit von Personen	22.08.2014	Tobias Albrecht

## Einarbeitung Facebook API

Facebook bietet einen API Service zur Abfrage von Profildaten mittels OAuth2.0 an. Dieses Verfahren beschreibt, dass eine Applikation Zugriff auf Daten nur durch Einwilligung erhält.

Profildaten werden innerhalb von Facebook durch AccessTokens freigegeben.

Um Zugriff auf diese zu erhalten muss die Applikation einen Link zur API / Internetseite des Datenlieferanten generieren, bei der um die Erlaubnis auf AccessTokens gebeten wird. Klickt der Kunde auf den Link, so wird er um Zugriff, auf durch die Applikation definierten AccessTokens, gebeten und bei Zustimmung zurück mit einem gültigen Session Parameter geschickt.

Mit diesem Session Parameter können Profildaten für 2 Stunden ausgelesen werden.

Facebook bietet zur Realisierung dieses Verfahrens für verschiedene Programmiersprachen SDKs (Software Development Kits) an.

Wichtig: Während der Einarbeitungsphase wurden die AccessTokens mehrmals geändert ebenso wie eine neue Version des SDK Released.

Daher wird nicht näher auf die AccessTokens eingegangen. Eine Übersicht gültiger AccessTokens kann über den Graph API Explorer eingesehen werden (<https://developers.facebook.com/tools/explorer/>, Stand 30.08.2014).

Während der Einarbeitung in die API wurden verschiedene Verfahren zum importieren und Auswertung evaluiert. Diese wurden dann für die Entwicklung verwendet.

## Entscheidung Setup / Programmiersprache / Framework

Für die Entscheidung welches Betriebssystem, Programmiersprache und Framework verwendet werden soll wurde eine Nutzwertanalyse durchgeführt:

### Betriebssystem:

#### Debian:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
Erfahrung:	9	0,6	5,4
Komplexität der Installation:	7	0,2	1,4
Kosten:	9	0,2	1,8
		Summe:	8,6

#### Windows Server:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
Erfahrung:	5	0,6	2,5

Komplexität der Installation:	3	0,2	0,6
Kosten:	1	0,2	0,2
		Summe:	3,3

## Programmiersprache:

### PHP 5.5:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
Erfahrung:	9	0,6	5,4
SDK verfügbar:	9	0,3	2,7
Dokumentation:	9	0,1	0,9
		Summe:	9

### JavaScript:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
Erfahrung:	5	0,6	3
SDK verfügbar:	9	0,3	2,7
Dokumentation:	9	0,1	0,9
		Summe:	6,6

### Java:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
Erfahrung:	3	0,6	1,8
SDK verfügbar:	9	0,3	2,7
Dokumentation:	9	0,1	0,9
		Summe:	5,4

## Framework:

### Eigenes Framework:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
ORM:	6	0,2	1,2
Lerneffekt:	9	0,4	3,6
MVC:	9	0,2	1,8
Einfachheit:	9	0,2	1,8
		Summe:	8,4

### Zend FW2:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
ORM:	7	0,2	1,4
Lerneffekt:	3	0,4	1,2
MVC:	9	0,2	1,8
Einfachheit:	1	0,2	0,2
		Summe:	4,6

### Laravel:

Kriterium:	Erfüllung:	Gewichtung	Ergebnis
ORM:	6	0,2	1,2
Lerneffekt:	9	0,4	3,6
MVC:	9	0,2	1,8
Einfachheit:	9	0,2	1,8
		Summe:	8,4

### Entscheidung:

Durch die Nutzwertanalyse konnte ermittelt werden, dass Debian mit PHP 5.5 und eigenem oder Laravel Framework die Beste Wahl ist. Durch den höheren Einsatz an Programmieraufwand für das eigenentwickelte Framework wurde dieses zurückgestellt.

Benutzerdaten einlesen

Um Facebook Benutzerdaten einlesen zu können wird eine Facebook App benötigt.

Diese kann unter <https://developers.facebook.com/apps> (Stand: 31.08.2014) angelegt werden. Zusätzlich benötigt Facebook ein verifiziertes Benutzerprofil. Nach Erstellung der App können unter dem Punkt „Dashboard“ die App Zugriffsdaten ausgelesen werden.

Facebook stellt mit seinem PHP SDK 4.0 eine Reihe von Hilfsklassen bereit, die die Kommunikation mit der API übernehmen. Darunter gehören FacebookSession, welches die Zugriffsdaten für die APP und Session hält, FacebookRedirectLoginHelper zur Generierung des Links zur Autorisierung der AccessTokens und FacebookRequest zur Abfrage der API.

### Bestimmung AccessTokens

Facebook arbeitet auf AccessToken Technologie bei der auf einzelne Parts des Profils bestimmte Berechtigungsstufen (Tokens) liegen, um Berechtigung auf alle Daten des Profils zu erhalten und somit keine Probleme bei der Auslesung zu erhalten werden alle Tokens (Stand 30.08.2014) die über den Graph API Explorer hinterlegt sind abgefragt.

### Profildaten auslesen

Das Profil des aktuellen Nutzers kann durch die Angabe des Pfads „/me“ angefragt werden, dabei werden alle Daten die direkt mit einem Profil verknüpft sind ausgelesen.

Das Response Objekt des Requests stellt Methoden bereit die Daten auszulesen.

Das Standardprofil des Facebook Benutzers enthält die User ID, Geburtsdatum, Werdegang (schulisch und beruflich), Schulfreunde, aktueller Wohnort, Geburtsort, Sprachen, Politische und

religiöse Richtung, sportliche Interessen + Freunde mit denen gemeinsam Sport gemacht wird, letzte Profiländerung und Internetseiten.

Sportliche Aktivitäten, Werdegang (Schule und Firma), politische und religiöse Richtung, ebenso wie Freunde aus Schule und Sport können bei anderen Profilen ebenso vorkommen und werden daher in Property Tabellen ausgelagert welche über eine Table Relation wieder mit dem Profil verknüpft werden.

Diese Property Tabellen enthalten durchgehend eine ID und mind. ein Bezeichner Name. Dabei wird der Name jeder Property Tabelle als eindeutig erachtet, so dass möglichst viele Beziehungen unter den Profilen existieren und diese ausgelesen werden können. Ein weiterer Vorteil ist es, dass mit dieser Technik andere Soziale Netzwerke eingebunden werden können und diese Daten automatisch verknüpft sind.

Neben dem Standardprofil werden Freunde, Nachrichten, Likes und Gruppen ausgelesen.

### *Freunde*

Facebook erlaubt durch den Pfad „/me/friends“ eine eingeschränkte Auslesung der Freunde.

Es werden neben der gesamt Anzahl nur alle Freunde die diese App ebenfalls benutzen aufgelistet.

### *Nachrichten*

Nachrichten werden in Threads zusammengefasst und können über den Pfad „/me/inbox“ geladen werden. Jeder Thread kann beliebig viele Gesprächsteilnehmer haben. Standardmäßig werden pro Thread die letzten 25 Nachrichten geladen, für alle weiteren muss durch das Facebook Paging ein neuer Request gestartet werden.

### *Likes*

Über den Pfad „/me/likes“ werden 25 Facebooks Likes mit Kategorie aufgelistet. Alle weiteren Likes können durch ein größeres Limit oder durch Nutzung des Pagings abgerufen werden.

Während der Entwicklung und der Testes ist aufgefallen, dass der Name innerhalb von Facebook nicht eindeutig ist und somit nicht der 3. Normalform entspricht.

### *Gruppen*

Gruppen können über den Pfad „/me/groups“ ausgelesen werden dabei ist die Paging Technologie ebenso vorhanden. Neben dem Namen und ID der Gruppe sind Informationen wie Administrator, ungelesene Posts und Position vorhanden.

## Überführung in MySQL Tabellenschema

MySQL ist ein in OpenSource entwickeltes Datenbank Management System, welches durch die Datenbank Engine InnoDB Fremdschlüsselbeziehungen abbilden kann. Zur Abbildung der beschriebenen Facebook Daten wurden 33 Tabellen angelegt und miteinander verknüpft, dabei wurde die Zeichenkodierung UTF-8 als Standard genommen, Primärschlüssel heißen ID und haben die Eigenschaften INT, 10 Stellen + nur positiver Zahlenbereich. Fremdschlüssel Spalten werden mit Fremdtabelle + \_id bezeichnet. Zur Erkennung von Erstelldatum und Aktualisierungsdatum wurden die Spalten created\_at und update\_at angelegt.

Ein EER Schema ist im Source Code unter „sql/eer ocm.mwb“ zu finden.

## Testverfahren

Um den Datenimport mit verschiedenen Profilen zu realisieren müssen Entwickler oder Testbenutzer in Facebook hinterlegt werden, diese können auf nicht publizierte Applikationen zugreifen. Die Kommilitonen dieses Projekts haben sich bereit erklärt Ihre Profile zu Verfügung stellen, so dass jeder Import Schritt mit ihren Daten geprüft werden konnte.

## Statistische Daten erzeugen

Auf Basis der importierten Profildaten aus Facebook können statistisch vergleichbare Werte generiert werden. Dabei ist es abhängig wie gut die einzelnen Profile gepflegt sind und wie viele Daten zur Analyse gezogen werden. Bereits während der Evaluierung der Software mussten Änderungen zur Optimierung durchgeführt werden. Eine weitere Arbeit könnte sich intensiver mit der Auswertung der Daten beschäftigen und Verfahren wie Text-Mining implementieren. Durch den Import der Daten ergeben sich hier besonders Vorteile, da komplexere Abfragen in SQL geschrieben werden können und auch BI Techniken zum Einsatz kommen können.

Die einfache Analyse durch diese Arbeit beschränkt sich der Zeit auf 14 Faktoren.

- Dauer bei einem Arbeitgeber in Jahren
- Anteil Admin Gruppen zu Gruppen
- Anzahl Freunde / Alter
- Anzahl Gruppen / Anzahl Freunde
- Anzahl Nachrichten der 7 Tage / Anzahl Freunde
- Anzahl Schulfreunde / Anzahl Freunde
- Anzahl Likes / Kategorie
- Politische Orientierung
- Anzahl Sprachen
- Anzahl Sportarten
- Anzahl Freunde
- Bildungslevel
- Arbeitsqualifikation (einfaches Text-Mining)
- Anzahl Bildungswegdegang / Anzahl Arbeitsplatzwechsel

## Statistische Daten exportieren

Die generierten statistischen Daten werden in diesem Projekt nicht gewichtet und nicht ausgewertet, daher muss ein externes Tool diese aggregieren und auswerten.

Diese Funktionalität ist durch die Generierung einer JSON Datei pro Profil implementiert, welche innerhalb des Projekts unter „app/storage/out/“ abgelegt wird.

Zur späteren Import Verifizierung wird ein generierter Hash aus Person ID, Import Gültigkeit bis, einer Zufallszahl und einem geheimen Schlüssel erstellt.

Der Dateiaufbau hat einen Head Bereich mit Steuerdaten, welche beim Import zurückgespielt werden und einen Daten Bereich der die statistischen Daten enthält.

### JSON Datei Aufbau:

#### Head Bereich

Feld	Datentyp	Beschreibung
id	INT	Export ID
version	FLOAT	Export Format Version
created_at	ISO Date	Erstelldatum
expires_at	ISO Date	Gültigkeit bis
person_id	INT	Person Ident.nummer
hash	CHAR(32)	Schlüssel zur Verifizierung der Berechtigung beim Import

#### Daten Bereich

Feld	Datentyp	Beschreibung
frequenz_job_change	FLOAT	Dauer bei einem Arbeitgeber in Jahren
percent_group_administrator	FLOAT	Anteil Gruppen Admin zu Anzahl Gruppen
percent_friends_age	FLOAT	Anzahl Freunde / Alter
percent_group_friends	FLOAT	Anzahl Gruppen / Anzahl Freunde
percent_threads_friends	FLOAT	Anzahl Nachrichten letzter Woche / Anzahl Freunde
percent_friends_schoolfriends	FLOAT	Anzahl Schulfreunde / Anzahl Freunde
percent_likes_category	FLOAT	Anzahl Likes / Anzahl Kategorien
orientation_political	CHAR(45)	Politische Orientierung
languages	INT	Anzahl Sprachen
sports	INT	Anzahl Sportarten
friends	INT	Anzahl Freunde

education_level	INT	Bildungsabschlusslevel: Grundschule => 5 Realschule / Gymnasium => 10 Hochschule => 20 Master => 30
work_level	INT	Joblevel: Geschäftsführer => 100 Teamleiter => 50 Gruppenleiter => 60 Head Of => 70 TCO => 80 Leiter => 35
balance_work_education	INT	Anzahl Bildungswerdegang / Anzahl Arbeitsplätze

## Five Factor Daten importieren

Durch ein externes Tool, welches nicht Teil dieser Arbeit ist, werden die exportierten statistischen Daten ausgelesen, geprüft, aggregiert und in das Five Factor Modell überführt. Dabei legt das Tool eine JSON Dateien in den „app/storage/in/“ Ordner an.

Diese muss einen Head und Daten Bereich haben. Der Head Bereich muss die Daten aus der exportierten Datei enthalten, so dass die Daten einer entsprechenden Person zugewiesen werden kann und der Zugriff auf diese Person verifiziert werden kann.

### JSON Datei Aufbau:

#### Head Bereich

Feld	Datentyp	Beschreibung
id	INT	Export ID
version	FLOAT	Export Format Version
created_at	ISO Date	Erstelldatum
expires_at	ISO Date	Gültigkeit bis
person_id	INT	Person Ident.nummer
hash	CHAR(32)	Schlüssel zur Verifizierung der Berechtigung beim Import

#### Daten Bereich

Feld	Datentyp	Beschreibung
open	INT 0 - 10	Offenheit für Erfahrungen
extraversion	INT 0 - 10	Extraversion
conscientiousness	INT 0 - 10	Gewissenhaftigkeit
compatibility	INT 0 - 10	Verträglichkeit
neuroticism	INT 0 - 10	Neurotizismus

Nach erfolgreicher Verifizierung und Validierung werden die Daten in das System importiert.

## Grafische Vergleichbarkeit von Personen

Zur grafischen Auswertung von Personen wird eine Internetseite verwendet die mit einem Login versehen ist. Nach der Authentifizierung steht ein Balkendiagramm mit 2 Mehrfachauswahlboxen zu Verfügung.

Das Balkendiagramm beinhaltet auf der X-Achse 11 Werte (0-10), welches die Ausprägung darstellt. Auf der Y-Achse beliebig viele, max. die Anzahl der Personen, die genau diese Ausprägung haben.

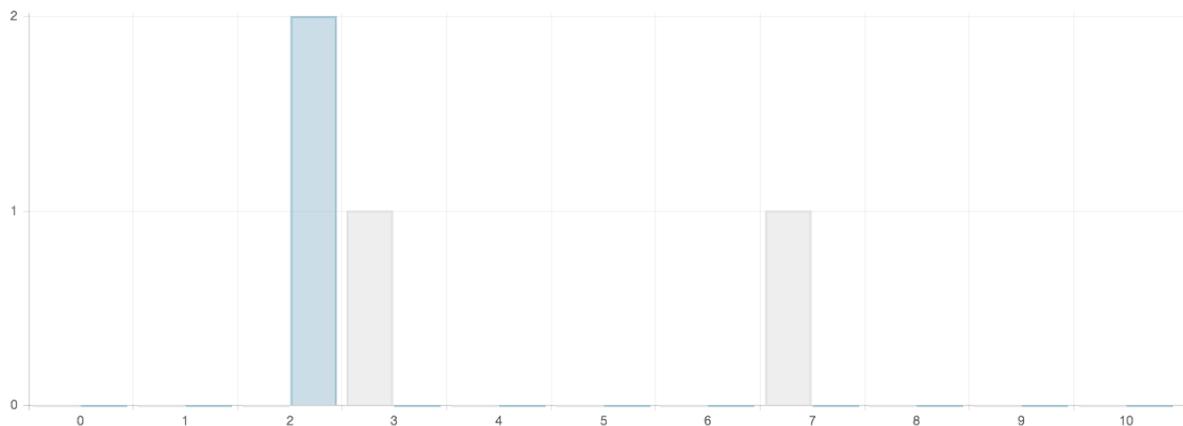
Die zwei Auswahlboxen enthalten alle importieren Personen mit Vor- und Nachname.

Diese können an- und abgewählt werden und somit gegeneinander verglichen werden.

Tobias Albrecht	Tobias Albrecht
Daniel Urstöger	Daniel Urstöger
Florian Hofsäss	Florian Hofsäss
Grundl Flo	Grundl Flo

Filtern

Gewissenhaftigkeit



## Installationsanleitung

Zur Installation der Software wird ein funktionierendes Debian Betriebssystem mit Lighttpd oder nginx als Webserver mit PHP5.5 und MySQL benötigt.

Nginx und Lighttpd benötigen Rewrite Rules, die alles bis auf JS und IMG Ordner auf die index.php Datei weiterleiten.

PHP5.5 benötigt FPM, MCRYPT, MYSQLND und CURL als Pakete

MySQL benötigt einen Benutzer (siehe Projekt Konfiguration).

Laravel Framework wird installiert durch Composer.

Zusätzlich muss das Crontab erweitert werden um die automatisiert ausgeführten Shell Befehle

- artisan user:FBImport
- artisan user:FACTgen
- artisan user:ProfileExport
- artisan user:AnalyseImport

## Literaturverzeichnis

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York:Holt.

Angleitner, A., Ostendorf, F., & John, O. (1990). Towards a taxonomy of personality descriptors in German: A psycho-lexical study [Special Issue: Personality language]. *European Journal of Personality*.

Baumgarten, F. (1933). *Die Charaktereigenschaften / The character traits*. Bern, Switzerland: Beitrage zur Charakter- und Persoenlichkeitsforschung.

Büttner, R. (2014). A Framework for Recommender Systems in Online Social Network Recruiting: An Interdisciplinary Call to Arms. 47th Hawaii International Conference on System Sciences.

Cattell, R. B. (1946). *Description and measurement of personality*. New York: World Book.

Dang-Xuan, L., & Stieglitz, S. (2014). *Social Media Analytics*. Wirtschaftsinformatik.

Digman, J. (1990). *Personality Structure: Emergence of the Five-Factor Modell*. Honolulu, Hawaii : Department of Psychology, University of Hawaii at Manoa.

Digman, J. (2004). *Personality structure: emergence of the five-factor model*. London Sage. 2: 71–93: The psychology of individual differences.

Faullant, R. (2006). *Psychologische Determinanten der Kundenzufriedenheit*.

Fritz, O., Angleitner, A., & John, O. P. (2004). The lexical approach to personality: a historical review of trait taxonomic research. *The psychology of individual differences*.

Furnham, A., Dissou, G., Sloan, P., & Chamorro-Premuzic, T. (2007). *Personality and Intelligence in Business People: A Study of Two Personality and Two Intelligence Measures*.

Ing Weller, W. M. (2009). *Persönlichkeit und Personalforschung. (Jahrg. 23, H.3)*. German Journal of Research in Human ResourceManagement.

Manovich, L. (2012). *Trending: The Promises and the Challenges of Big Social Data*. U of Minnesota P: Digital Humanities. Ed. Matthew K. Gold. Minneapolis.

- McCrae, R., & John, O. (1992). An introduction to the five-factor model and its applications. *Journal of Personality* 60 (2).
- Ostendorf, P. B. (1993). NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae Handanweisung. Göttingen.
- Saucier, G. (1994). Mini-Markers: A Brief Version of Goldberg's Unipolar Big-Five Markers. *Journal of Personality Assessment*.
- Schulten, M., Mertens, A., & Horx, A. (2012). *Social Branding*. Springer Verlag.
- Scott, J., & Carrington, P. (2011). *Social network analysis: developments, advances, and prospects*. London, UK: SAGE Publications Ltd.
- Sponheuer, B. (2010). *Employer Branding als Bestandteil einer ganzheitlichen Markenführung*. Wiesbaden: GWV Fachverlage GmbH , 1. Auflage 2010.
- Suda, W., & Fouts, G. (1980). Effects of Peer Presence on Helping in Introverted and Extroverted Children.
- Winter, D., & Barenbaum, N. (2003). History of modern personality theory and research. *Handbook of personality: theory and research*.
- Witte, T. S. (2007). Teamfähigkeit und Performance. *Gruppendynamik und Organisationsberatung*, 38. Jahrg. Heft I 2007.
- Zeng, D., Chen, H., Li, S.-H., & Lusch, R. (Nov/Dez 2010). *Social Media Analytics and Intelligence*. Published by the IEEE Computer Society.
- Zhang, R.-P., & Tsingan, L. (2013). *Extraversion and Neuroticism Mediate Associations Between Openness, Conscientiousness, and Agreeableness and Affective Well-Being*. Springer Science+Business Media Dordrecht 2013.

## ANLAGEN

Quellcode: organizationculturematching.zip