Transferbericht-1: Jacqueline Franßen (496804)

Im folgenden werde ich den Praxisbezug des Faches „Analyse semi- und unstrukturierter Daten“ näher beschreiben.

**Transfer: Inhalte aus dem Modul in meinem beruflichen Alltag**

Ich konnte einige Inhalte des oben genannten Faches in meinem beruflichen Alltag beobachten. Dazu gehören unter anderem die verschiedenen Techniken, die natürliche Sprache zu analysieren (auch bekannt unter dem Begriff des Natural Language Processing).

Seit Beginn 2019 darf ich in meinem Beruf in einer Fachgruppe teilnehmen. Ziel der Fachgruppe ist es, eine intelligente Suche zu entwickeln, die zwei Suchmodi erlauben soll: Q&A und Document Search. Bei ersterem Typ handelt es sich um einen einfachen „Frage“-„Antwort“-Mechanismus, bei dem die User gezielte Fragen an die Plattform stellen und anschließend in einem „Fragenkatalog“ nach einer passenden Antwort gesucht wird. Diese Form der Sprach-Verarbeitung ist eine sehr einfache Form, welche mittels einer Map mit Key-Value-Paaren implementiert werden kann. Dem gegenüber steht die Form „Document Search“, welche mittels NLP-Techniken die eingegebene Frage (es muss hier keine grammatikalisch-korrekte Frage sein) verarbeitet, anschließend alle Einträge aus der Datenbank auswertet und das Ergebnis mit den meisten Matches auf die eingegebene Frage ausgibt.

**Bezüge zwischen dem beruflichen Alltag und dem gewählten Modul**

Die im ersten Abschnitt genannte Fachgruppe entwickelt unter Verwendung von Techniken des Natural Language Processing (Stemming, Lemmatisierung etc.). Was mir in der beruflichen Praxis vor allem geholfen hat, mein Verständnis für diese Prozesse zu stärken, war folgender Inhalt des o.g. Moduls: „STIRL-Mechanismus“ sowie die Einführung in die NLP mittels Python. Der „STIRL-Mechanismus“ beschreibt die Verarbeitung unterschiedlich strukturierter Daten zu einem „Corpus“, also einem einheitlichen „Datenpool“, aus dem anschließend „Antworten“ oder „Topics“ erstellt werden können. Dieser Schritt der Datenaufbereitung war unter anderem notwendig zu implementieren, da bei unserem Projekt oben ebenfalls Quellen unterschiedlicher Herkunft verwendet wurden.

**Transfer: Kenntnisse aus meinem beruflichen Alltag, die mir bei dem Modul geholfen haben**

In meinem beruflichen Alltag programmiere ich vor allem das Frontend (Webanwendungen)

 für verschiedene Kunden. Daher ist mir der Umgang mit Gitlab- und Github-Anwendungen bekannt. Während meines Vorstudiums (Bachelor of Science) nutzte ich diese Versionsverwaltung vor allem für Code aber auch für das Erstellen meiner Bachelorarbeit. Es ermöglicht ein einfaches, betriebssystem-unabhängiges Arbeiten, welches gleichzeitig die Historie des Repositories/der hinterlegten Daten anzeigt. So lassen sich ältere Stände auf einfache Weise wiederherstellen.

Im Falle eines Ausfalls von Hardware (z.B. Rechner stürzt ab oder kann nur durch ein Backup wiederhergestellt werden) kann es schnell passieren, dass relevante Daten verloren gehen würden. Insbesondere in diesem Fall ist es vorteilhaft, eine Sicherung zu haben, die gegebenenfalls auf mehrere „Einheiten“ verteilt ist. D.h. nicht nur die Daten auf einer externen Festplatte zu speichern, sondern ebenfalls in der Cloud oder innerhalb einer Git-Versionsverwaltung.

Ein weiterer Punkt, der mir aus meinem beruflichen Alltag weitergeholfen hat, war das Zeitmanagement und eine frühzeitige Planung bzw. Eingrenzung der Problemstellung.

Da in o.g. Modul die Prüfungsleistung eine Projektarbeit ist und die Auswahl des Themenbereiches uns überlassen war, habe ich schon früh mit der Recherche über mögliche

Themen begonnen. Dabei hat mir unter anderem die Aufzeichnung verschiedener möglicher Themenpfade und Strategien der Implementierung geholfen, welche ich in einem zweiten Schritt gegenübergestellt habe und gegeneinander abgewogen habe. Als Ergebnis dieses Prozesses kristallisierte sich dann ein „mögliches“ Thema heraus, welches ich anschließend mit dem Dozenten besprach.

In meinem beruflichen Alltag spielt neben dem Zeitmanagement auch die Agilität und die Reaktion auf Kunden-Anforderungen bzw. Änderungen der Anforderungen eine zentrale Rolle. Im vorherigen Absatz hatte ich bereits mein Thema eingegrenzt und konnte auf dieser Basis nun meine Literaturrecherche beginnen. Ziel dieser Phase war es, bereits implementierte Lösungen zu erkennen und diese mit in den „Pool“ an verwandten Arbeiten aufzunehmen. Diese Phase diente lediglich der Spezialisierung auf meine eigens zu entwickelnde Lösung. Durch die Erfahrungen aus meinem beruflichen Alltag, insbesondere das Reagieren auf sich ändernde Kundenanforderungen, konnte ich meine implementierte Lösung in Iterationen entwickeln und so mit jeder weiteren Iteration die Implementierung verbessern. Während der Implementierung meiner Projektarbeit legte ich viel Wert auf die Versionierung und das regelmäßige Testen der Anwendung, allerdings legte ich keine festen "Meilensteine" fest. Dies erschwerte mir manchmal die Weiterentwicklung. Auch hätte ich mir gewünscht regelmäßig eine Art "Peer-Reviewing" durchzuführen, da die Meinung einer anderen Person häufig andere Fehlerquellen aufzeigen kann und somit eventuell die Projekt-Entwicklung verbessern und in einen anderen effizienteren Weg einleiten kann.

**Zusätzlichen Inhalte, die ich meiner beruflichen Erfahrung für wichtig erachte, in das Modul zu integrieren bzw. im Curriculum zu vertiefen:**

Da ich bei der Implementierung meiner Projektarbeit wiederholte Male auf das Problem der begrenzten Commodity-Hardware gestoßen bin, finde ich es sinnvoll, bereits zu Beginn der Vorlesung einige Alternativen für das Nutzen solcher Hardware aufgezeigt zu bekommen.

Wir hatten bereits den Zugang für Amazon Web Services Accounts, allerdings wäre eine rechtzeitige Anleitung oder Einführung der Anwendung sinnvoll gewesen.

Auch eine Gegenüberstellung und Empfehlung von Entwicklungsumgebungen wie Spyder für Python und dem Juypter Notebook wäre sinnvoll gewesen. Für meine Projektarbeit habe ich ausschließlich in meinem Juypter Notebook implementiert, allerdings zeigt dies häufig keine konkrete Fehlerbeschreibung an, was den Debug-Prozess etwas verlängert und erschwert.

Außerdem fände ich es sinnvoll, wenn der Dozent ein Beispiel-Projekt aufsetzen könnte, welches er dann in der Vorlesung Schritt für Schritt durchführt/implementiert, sodass sich die Studierenden daran orientieren können. Darüber hinaus könnte durch „festgelegte“, regelmäßige Meilensteine (innerhalb des Semesters, vergleichbar mit Sprints aus Scrum) ein stetiger Austausch der Projektstände gewehrleistet werden und somit evt. Probleme bei der Projektdefinition frühzeitig verhindert oder behoben werden. Es könnte noch einen Schritt weitergegangen werden, indem man Scrum-ähnliche Rollen (z.B. Scrum-Master, Product-Owner) innerhalb des Semesters definiert, welche jeweils über einen Sprint von Studierenden übernommen werden. Somit könnten bei jedem Treffen, der aktuelle „Scrum-Master“ oder „Product-Owner“ mit einer (symbolischen) „Brille“ einen Blick auf die Zwischenstände der Kommilitonen werfen und gezielt Feedback darauf geben. Anschließend können die Ergebnisse mittels Projekt-Management-Tools (z.B. Jira oder alternativ Trello) festgehalten werden. Dieses Verfahren ist meiner Meinung nach sinnvoll in o.g. Modul, da eine Projektarbeit entstehen soll, welche in der Praxis in einem „iterativen“, agilen Vorgehen stattfindet. Gleichzeitig begründet die Form des regelmäßigen Austausches bzw. der Dokumentation der Zwischenergebnisse eine effiziente Arbeitsweise für alle "Projektmitglieder" und es kann zeitnah auf fehlerhafte oder verlangsamende Vorgehensweisen reagiert werden.