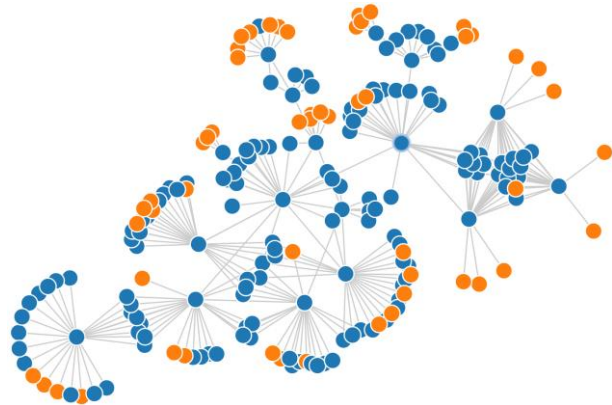




Ingestion: Datenaufnahme und -analyse für semantische Unternehmensnetzwerke

Unternehmensnetzwerke umfassen ein komplexes Geflecht verschiedener Organisationen. Dazu gehören zum einen die Zulieferer und Partner eines Unternehmens, zum anderen lassen sich aber auch die Kundenbeziehungen des Unternehmens als Kanten in einem Unternehmensgraphen verstehen. Neben externen Beziehungen lassen sich auch strukturelle Informationen wie beispielsweise Tochterunternehmen eines Konzerns als Kanten eines Graphs modellieren.



Die Erstellung und Analyse solcher Netzwerke stellt aufgrund der hohen Heterogenität der Eingabedaten eine Herausforderung dar, bietet aber auch eine Reihe interessanter Anwendungen. So kann die Analyse eines solchen Netzwerks dazu dienen, Entscheidungsprozesse signifikant zu unterstützen. Besonders die Bereiche Risiko- sowie Konkurrenz- und Betrugsanalyse können von der Auswertung eines solchen Netzwerks profitieren.

Projektbeschreibung

Das Ziel des Projektes ist die systematische Erweiterung eines Systems zur Erstellung, Analyse und Visualisierung von Unternehmensnetzwerken um neuartige Komponenten und Methoden. Der erste Fokus liegt auf der Automatisierung des komplexen Datenaufnahmeprozesses (*data ingestion*):

- **Named Entity Disambiguation:** Zur Zuordnung von Erwähnungen in Texten zu konkreten Unternehmen im Netzwerk existieren diverse Ansätze. Diese sollen in das System integriert und erweitert werden, so dass es automatisiert und zuverlässig Textströme aufnehmen kann.
- **Zeitbezogene Informationen:** Da sich Unternehmen und deren Beziehungen im Laufe der Zeit ändern, sollen zeitbezogene Daten modelliert und realisiert werden. So soll es möglich sein, z.B. die Veränderung eines Netzwerks über Zeiträume hinweg zu visualisieren und analysieren.
- **Relationship Extraction:** Beziehungen zwischen Unternehmen sind vielfältig und können in verschiedensten Quellen entdeckt werden. Entsprechende Verfahren sollen in das System integriert und erweitert werden.

Der zweite Fokus liegt auf der Analyse des resultierenden, sehr großen Netzwerks:

- **Graphexploration:** Die vorhandene Lösung zur interaktiven Exploration des Graphen sollen erweitert werden, um verschiedene Fragestellungen von Risikoanalysten abzubilden.
- **Graphanalyse:** Zum Ranking von Unternehmen kann z.B. der PageRank bestimmt werden; zur Entdeckung von Lieferketten sollen Pfade zwischen zwei Unternehmen untersucht werden; Teilgraphen von Konkurrenten oder ganzen Branchen sollen verglichen werden.

Technologien, Architektur und Vorgehensmodell



Das aktuelle System ist als Client-Server Architektur implementiert. Dabei besteht der Server aus einem RESTful Webservice der mittels Python und dem Web-Framework Flask implementiert ist. Der Client, der zur Visualisierung und Exploration des Graphen verwendet wird, ist in AngularJS und D3 implementiert. Im Backend kommen im Wesentlichen zwei Datenbanken zum Einsatz. Zur Speicherung der Rohdaten wird MongoDB verwendet, anschließend wird die Graphdatenbank Neo4j verwendet um das effiziente Speichern sehr großer Graphen als auch die schnelle Ausführung verschiedener Analysen zu ermöglichen. Weiterhin wird Apache Solr verwendet einen effizienten Volltextindex auf den zu verarbeitenden Dokumenten zu realisieren.

Um eine gemeinsame und flexible Konkretisierung der zu erreichenden Ziele zu ermöglichen, ist ein agiles Vorgehen wünschenswert; als Vorgehensmodell wird Scrum empfohlen.

Projektpartner

Die **Commerzbank** ist eine führende, international agierende Geschäftsbank mit Standorten in mehr als 50 Ländern. Mit den Geschäftsbereichen Privatkunden, Mittelstandsbank, Corporates & Markets und Central & Eastern Europe bietet sie ihren Privat- und Firmenkunden sowie institutionellen Investoren ein umfassendes Portfolio an Bank- und Kapitalmarktdienstleistungen an. Die Commerzbank finanziert über 30 Prozent des deutschen Außenhandels und ist unangefochtener Marktführer in der Mittelstandsförderung. Mit den Töchtern comdirect und der polnischen mBank verfügt sie über zwei der weltweit innovativsten Online-Banken.

Die Commerzbank betreibt mit 1.100 Filialen sowie rund 90 Geschäftskundenberatungszentren eines der dichtesten Filialnetze der deutschen Privatbanken. Insgesamt betreut sie rund 15 Millionen Privat- sowie 1 Million Geschäfts- und Firmenkunden. Im Jahr 2014 erwirtschaftete sie mit durchschnittlich rund 52.000 Mitarbeitern Bruttoerträge von knapp 9 Milliarden Euro.

Das Projekt für 6-8 Studenten beginnt am 1. Oktober 2016 und wird durch Prof. Dr. Felix Naumann, Michael Loster und Toni Gruetze betreut. Es setzt die erfolgreiche Kooperation in einem Bachelorprojekt des vergangenen Jahrgangs fort. Fragen können gerne an michael.loster@hpi.de oder felix.naumann@hpi.de gerichtet werden.