

Netzwerke sehen

Matej Ďurčo, ACDH-ÖAW

In diesem Beitrag stellen wir eine Webapplikation zur Visualisierung und interaktiven Erkundung von Graphen und Netzwerken vor. Die Applikation ist ursprünglich im Kontext der Forschungsinfrastruktur CLARIN entstanden, mit dem Ziel die komplexe Datendomäne der Metadaten-Profile der Component Metadata Infrastructure¹ (CMDI) (Broeder et al., 2010) besser fassbar zu machen. (Ďurčo, 2013). Im Laufe der Entwicklung hat sich diese Applikation zu einem generischen Viewer für jede Art von graph-basierten Daten weiterentwickelt.

Die Applikation kann auch im Vergleich mit alternativen weit verbreiteten Tools bestehen. Gephi² bietet zwar wesentlich mehr Funktionalität zum automatischen Analysieren von Graphen, ist aber eine Client-Applikation, die lokal installiert wird, und die Möglichkeiten der dynamischen Navigation im Graphen sind auch nicht so reichhaltig, wie in der vorgestellten Applikation. Die traditionelle command-line Applikation GraphViz³ ist zwar sehr stark im eleganten Layoutieren der Graphen, ist aber eine rein statische Anwendung ohne graphisches User Interface, bietet also keine Möglichkeit interaktiv zu arbeiten.

Die Applikation basiert auf der open-source javascript Bibliothek d3⁴ und läuft nach dem anfänglichen Laden vollständig client-seitig. Es bietet mehrere miteinander verknüpfte Ansichten und eine Reihe von Optionen zum Manipulieren der dargestellten Graphen. So ist es möglich mehrere Knoten auszuwählen und sich aus dem zugrundeliegenden geordneten Graphen beliebig viele Ebenen von Vorgänger- bzw. Nachfolgerknoten anzeigen zu lassen. Ebenfalls werden mehrere vordefinierte Layout-Algorithmen angeboten. Das Layout kann die Stärke der Verbindungen reflektieren, ebenso kann die Größe und Farbe der Knoten verwendet werden, um weitere Dimensionen visuell zu kodieren. Der aktuell angezeigte Graph, kann entweder als Link verschickt oder als SVG-Grafik exportiert und weiter verarbeitet werden.

Daten

Neben den CMDI Metadaten, für welche die Applikation ursprünglich vorgesehen war, wurde die Applikation bereits erfolgreich mit ganz anderen Datensätzen erprobt. So wurde zum Beispiel das „Philosophen-Influenz-Netzwerk“ visualisiert. Dafür wurden die Daten aus der dbpedia über den SPARQL-Endpoint abgefragt (Philosophen und ihre *influenced/influencedBy* Beziehungen), diese wurden durch eine einfache XSLT-Transformation in das Eingabeformat umgewandelt und in die Applikation importiert. Mit minimalem Aufwand konnte dergestalt eine große Datenmenge wesentlich besser erfasst und erkundet werden als dies mit konventionellen Instrumenten möglich wäre. (siehe Abb. 1)

Es ist vorgesehen, weitere Datensätze aufzubereiten und sie über die Applikation verfügbar zu machen. Dies wären zum einen Taxonomien, die in unterschiedlichen DH-Disziplinen Verwendung findet. Hier bietet sich SKOS als primäres Input-Format an, da es weit verbreitet ist und entsprechende Transformationen von SKOS in das interne Graph-Format eine ganze Klasse von Datensätzen für die visuelle Erkundung in der Applikation zugänglich machen würde. Eine weitere Klasse potentieller Daten sind prosopographische Annotationen, durch deren Auswertung sogenannte Kookurenznetzwerke visualisiert werden können. Hierfür wurden schon Experimente mit Daten aus dem Schnitzler Tagebuch (8.500 Personen mit ca. 77.000 Nennungen) und aus der Zeitschrift ‚Die Fackel‘ (15.000 Personen mit über 123.000 Nennungen) durchgeführt.

¹ <http://clarin.eu/cmd/>

² <https://gephi.github.io/>

³ <http://www.graphviz.org/>

⁴ <https://github.com/mbostock/d3/>

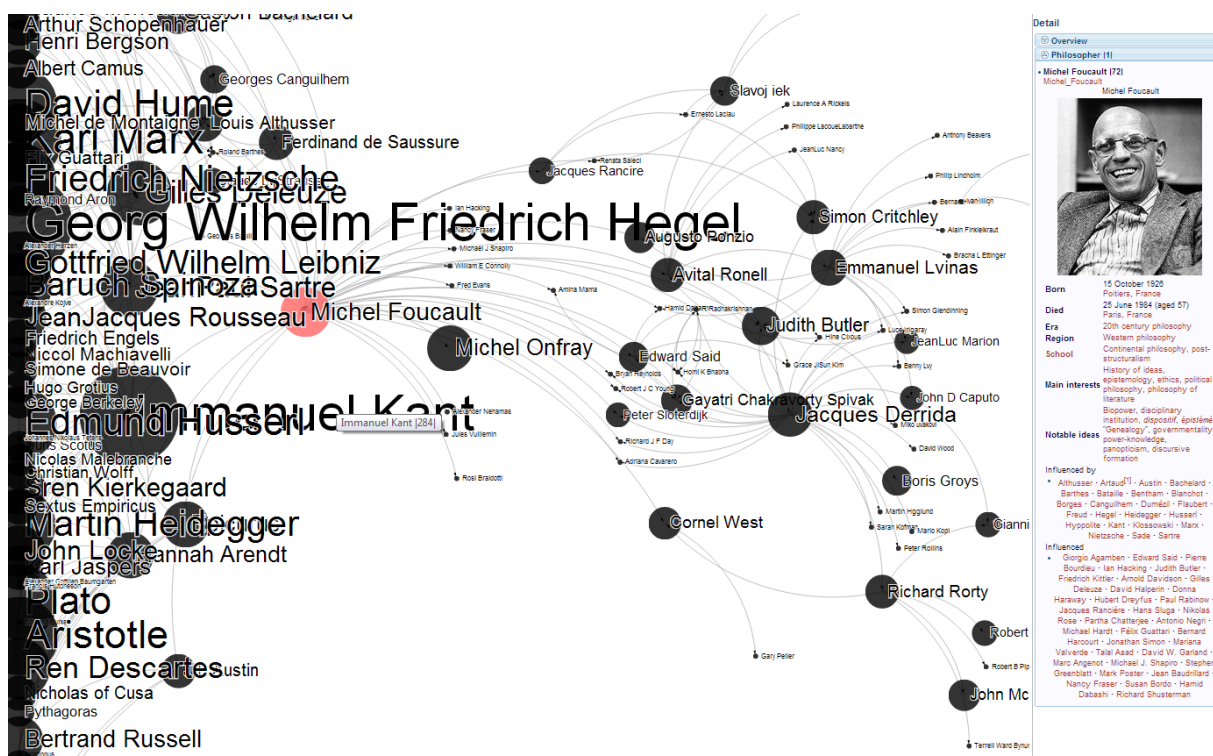


Abb 1 Screenshot der Visualisierung des "Philosophen-Influenz-Netzwerks". Es werden die VorgängerInnen und NachfolgerInnen von Michel Foucault dargestellt.

Nächste Schritte

Die Applikation ist bereits lauffähig und wird in ihrem ursprünglichen Kontext produktiv eingesetzt⁵, sie wird aber auch laufend weiterentwickelt. So ist eine Erweiterung geplant, um verbreitete standardisierte Graph-Formate (wie GraphML, GDF, GML) als Input und Output Formate zu unterstützen. Ebenso ist ein Refactoring des Codes notwendig, um die domänenspezifischen Aspekte von der generischen Applikation zu trennen und diese als ein sauberes wiederverwendbares konfigurierbares javascript-Modul anzubieten, das leicht in komplexere Applikationen eingebaut werden kann. Im Hinblick auf die eigentliche Visualisierung ist es wünschenswert, andere Darstellungsformen für die Knoten (momentan nur Kreise) anzubieten. Für tiefgreifende Analysen sind Graph-Operationen erforderlich, z.B. Vergleich von zwei Graphen und das Berechnen des gemeinsamen Subgraphen, Clustering u.ä.

Die Applikation wird als eigener Visualisierungsservice, der auch externe Daten annehmen, verarbeiten und visualisieren kann, als einer der Dienste des ACDH-ÖAW angeboten werden. Der Code wird open-source frei zur Verfügung gestellt.

Referenzen

- Broeder, D.; Kamps-Snijders, M.; Uytvanck, D. V.; Windhouwer, M.; Withers, P.; Wittenburg, P. & Zinn, C. (2010). A Data Category Registry- and Component-based Metadata Framework. In Calzolari, N.; Choukri, K. & others (Eds.). Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA).
- Đurčo, M. (2013). SMC4LRT - Semantic Mapping Component for Language Resources and Technology. Technical University, Vienna.

⁵ <http://clarin.oeaw.ac.at/smc-browser>