

Masterprojekt 2013/2014:

Semantische Erschließung des tele-TASK-Archivs

Das Webportal von www.tele-TASK.de ist ein gewaltiges Archiv von Aufzeichnungen von Vorlesungen, Fachvorträgen, Kolloquien etc. Etwa 4.500 Aufnahmen mit über 15.000 Kapiteln und 1.700 Rednern sind bislang über die Suchfunktion zu finden. Des Weiteren können Aufnahmen mit Tags versehen, bewertet und annotiert werden. Folienextraktion aus den aufgezeichneten Videos und Texterkennung (OCR), sowie die Generierung von Vortragsstruktur/Inhaltsverzeichnis laufen automatisiert ab. Ein weiteres Forschungsthema ist die automatische Spracherkennung (ASR). Bei all diesen Schritten entstehen große Mengen an Metadaten, die nicht nur durchsuchbar sein sollten, sondern auch miteinander verknüpft werden sollen. Dazu muss der stetig wachsende Suchraum des tele-TASK-Vorlesungsarchivs semantisch erschlossen werden. Durch die Verwendung von Hintergrundwissen wird bei einer semantischen Suchmaschine die inhaltliche Bedeutung von Texten und Suchanfragen berücksichtigt. Es wird nicht nur nach Wörtern im Text, wie bei Keyword-basierten Suchmaschinen, gesucht. Dadurch kann eine Suchanfrage präziser erfasst und mit den inhaltlich relevanten Texten in Verbindung gebracht werden. Somit werden inhaltlich korrekte Suchergebnissen bereitgestellt. Semantische Suche imitiert gewissermaßen das menschliche Gehirn, indem Wissen und Assoziationen zur Suche genutzt werden.

Open Data bedeutet die freie Verfügbar- und Nutzbarkeit von, meist öffentlichen, Daten. Sie beruht auf der Annahme, dass vorteilhafte Entwicklungen eingeleitet werden, wenn Daten für jedermann frei zugänglich gemacht werden. Davon ausgehend besteht der Wunsch, das Archiv von tele-TASK zugänglicher zu machen und der Welt zu öffnen. So soll eine Suchanfrage, sowohl vom Portal aus gestartet, aber auch über externe Suchanbieter, nicht nur zum betreffenden Vorlesungsvideo, sondern auch gleich an die richtige Stelle im Video springen. Darüberhinaus soll von da aus eine intelligente Benutzerführung durch sinnvolle, thematisch passende Vorschläge, erfolgen und so den Lernprozess mit tele-TASK intensivieren. Außerdem sollen Suchergebnisse durch Hintergrundwissen und Assoziationen genauer werden.

Die Aufgabe dieses Masterprojekts ist es, das tele-TASK-Webportal zu untersuchen und zu analysieren, welche Metadaten kreiert werden und wie diese akquiriert werden können. Ein semantisches Modell ist zu entwerfen und anschließend ist die Architektur des Portals dahingehend zu erweitern. Die Genauigkeit der Ergebnisse sowie die Nutzerakzeptanz sind zu untersuchen und auf weitere Verbesserungsmöglichkeiten zu prüfen.

Ansprechpartner: Prof. Dr. Christoph Meinel
Matthias Bauer, tele-TASK

