

HOUSTON, WIR HABEN EIN PROBLEM

SIEBEN AUF EINEN STREICH

SCHAU MIR IN DIE AUGEN, KLEINES!

GESCHÜTTELT

HASTA LA VISTA, BABY

NICHT GERÜHRT!

Online-Textsuche in Videomaterial

Ein Filmzitat schwirrt im Kopf, aber man kommt partout nicht darauf, welcher Szene aus welchem Film man es zuordnen soll. Wir stellen ein System vor, mit dem die gesuchte Filmszene innerhalb weniger Sekunden online angezeigt werden kann. Dazu genügt es in der Suchmaske des Browsers das gesuchte Wort einzugeben und alle Szenen, in denen das Wort gesagt wird, können mit einem Klick online abgespielt werden. Zusätzlich kann das gefundene Filmzitat auch in anderen Sprachen ausgegeben werden.

Angenommen, man möchte sich die Szene anschauen, in der der berühmte Satz aus dem Film Casablanca »Schau' mir in die Augen, Kleines.« fällt. Die einfache Suche nach dem Wort »Augen« liefert mit unserem System alle Textstellen im gewählten Kontext, in welchem das Wort »Augen« vorkommt. Bei der Auswertung der angezeigten Treffer erkennt man, dass das Zitat in dieser Form gar nicht existiert, sondern richtig »Ich seh' Dir in die Augen, Kleines.« heißt. Das System eignet sich aber für weit mehr als dafür, falsche Zitate aufzuspüren - es kann als eine universelle Filmszenensuchmaschine verwendet werden - mit dem Plus, dass die entsprechende Filmszene zum gesuchten Text im Browser unmittelbar angezeigt werden kann. Darüber hinaus kann das System die Filmdialoge auch in ihrer Übersetzung anzeigen und unterstützt so multimedial den Fremdsprachenunterricht. Auch Sprachwissenschaftler kommen auf ihre Kosten, denn es handelt sich bei den Filmzitaten um ein begehrtes Textkorpus der gesprochenen Sprache, dessen Wörter zudem mit der jeweiligen Wortart und Grundform ausgezeichnet sind. Eine solche Korpus-Suchmaschine ermöglicht die Beantwortung spannender sprachwissenschaftlicher Fragen.

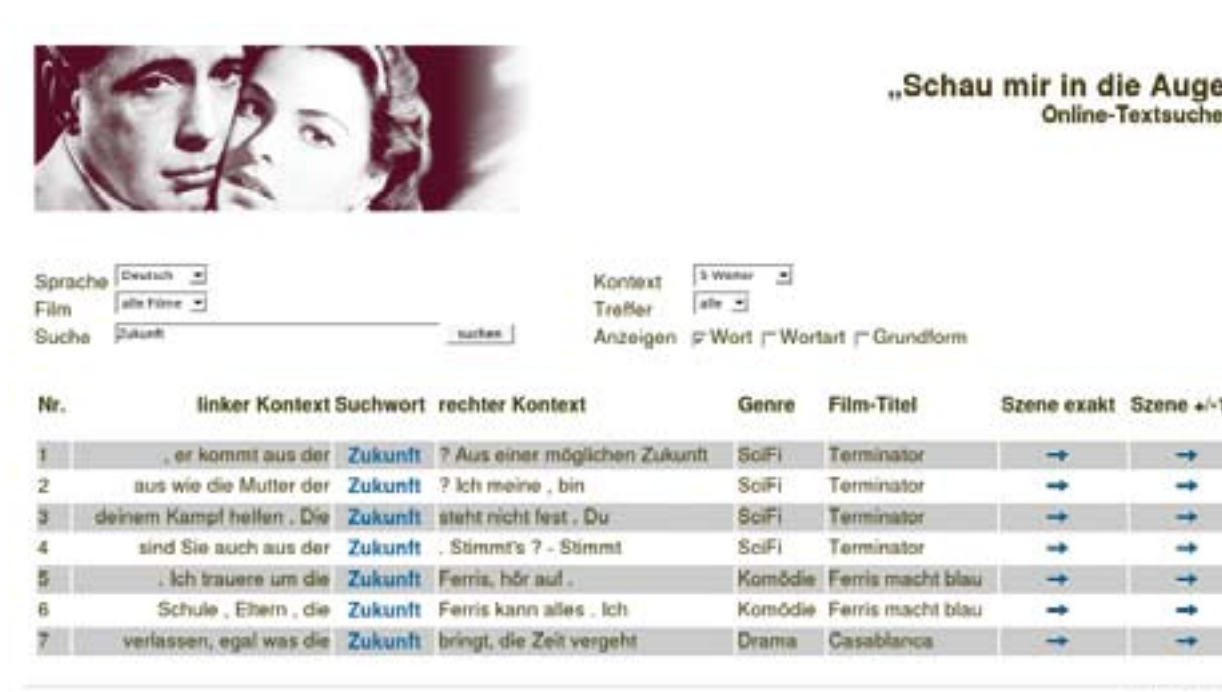
Das Schaubild zeigt die Online-Schnittstelle mit der Ausgabe aller Treffer zum gesuchten Begriff »Zukunft«, wobei alle gespeicherten deutschen Filme durchsucht werden. Die anklickbaren Pfeile führen zum konkreten Filmausschnitt, alternativ auch zu einer Fassung mit mehr Kontext.

Technische Realisierung des Projektes: Die Textsuche erfolgt auf vorher extrahierten und

indexierten Untertiteln der Film-DVDs. Bei jedem Untertitel ist der exakte Zeitcode der entsprechenden Filmszene angegeben, er wird im Index mitgespeichert. Das Anklicken des Pfeilsymbols in der Ausgabe führt im Hintergrund zum sekundenschnellen Schneiden der jeweiligen Filmszene aus dem gesamten Film, so dass dem User unmittelbar die gesuchte Szene angezeigt werden kann. Das System läuft auf einem Linux-Server; clientseitig wurde die Online-Schnittstelle in der Programmiersprache PHP umgesetzt. Da die Online-Textsuche in einem Webbrowser ermöglicht und zum Abspielen der Filmausschnitte das Plugin des VLC media players eingesetzt wird, ist ein betriebssystemübergreifender Zugriff auf unsere Filmdatenbank gewährleistet.

Einsatz im Projekt »LatinUS«: Das Projekt LatinUS am Institut für Romanistik hat zum Ziel, den Gebrauch des Spanischen in den US-amerikanischen Medien längerfristig zu dokumentieren und mit verschiedenen Methoden zu analysieren. Als Basis dafür dient ein multimodales Korpus, das kontinuierlich ausgebaut wird, bestehend aus Text, Video- und Tonmaterialien in spanischer Sprache aus den öffentlichen Kommunikationsbereichen in den USA. Das oben beschriebene System wird in diesem Projekt bereits prototypisch eingesetzt, mit dem Unterschied, dass die Suche nicht auf extrahierten Untertiteln erfolgt, sondern auf dem zuvor manuell transkribierten Text der Ton- bzw.

Videomaterialien. Die Anzeige der passenden Videoausschnitte erfolgt dann analog zur obigen Online-Filmdatenbank.



HU | Institut für Romanistik

Petra Prochazkova | petra.prochazkova@gmail.com

<http://euroling.de>

 multimedia projekt
HUMBOLDT-UNIVERSITÄT ZU BERLIN

