

Ähnlichkeitssuche in den Digital Humanities: Semi-automatische Identifikation von Kostümmustern

,
johanna.barzen@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

,
Falkenthal@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

,
Frank.Hentschel@uni-koeln.de
Universität zu Köln, Deutschland

,
Leymann@iaas.uni-stuttgart.de
Universität Stuttgart, Deutschland

,
Tino.Strehl@student.reutlingen-university.de
Hochschule Reutlingen, Deutschland

Ausgangslage

Kostüme in Filmen sind ein wichtiges Gestaltungselement der diegetischen Welt. Mit MUSE (MUster Suchen und Erkennen) verfolgen wir das Ziel, Konventionen zu identifizieren und darstellbar zu machen, die sich entwickelt haben, um Kostüme als kommunikatives, bedeutungstragendes Element zu nutzen. Um diese Konventionen zu identifizieren, verwenden wir das Konzept des Musters nach Christopher Alexander et al. (1977). In dieser Tradition kann ein Kostümmuster als abstrakte und bewährte Lösung eines wiederkehrenden Designproblems, wie beispielsweise der adäquate textile Ausdruck eines bestimmten Charakters, verstanden werden.

Um die Identifikation und das Verfassen von Mustern zu unterstützen, haben wir ein Lösungs- und ein Musterrepository konzipiert und implementiert. Während das Lösungsrepository ein detailliertes Erfassen der Kostüme aus Filmen ermöglicht (konkrete Lösungen für Designprobleme), können im Musterrepository abstrakte Designlösungen (Kostümmuster) abgelegt werden (Fehling et al. 2014). Wie aber identifiziert man diese Kostümmuster aus der Menge der multidimensional beschriebenen Kostümdateien?

Einen ersten Ansatz haben wir mittels der Analyse aufbauend auf OLAP Cubes vorgestellt (Barzen 2015). Dieser Ansatz erlaubt multidimensionale Abfragen auf den Kostümdateienbestand, beschränkt sich allerdings auf die Analyse der Kostümdateien durch konkrete Abfragen. Bei konkreten Abfragen nicht vermutete Zusammenhänge im Datenbestand können dabei nicht identifiziert werden. Um solche Zusammenhänge der Daten sichtbar zu machen, gewinnen besonders in Industrie und Naturwissenschaften Techniken aus dem Bereich des Data Minings an Gewicht. Diese erlauben mögliche „Auffälligkeiten“ oder Cluster in Datensätzen zu finden. Was wir in diesem Poster vorstellen möchten, ist eine Werkzeugumgebung, die verschiedene Algorithmen und entsprechende Visualisierungen der Analyseergebnisse zur Identifikation von „Kostümmusterkandidaten“ unterstützt. Dem vorgegebenen Umfang geschuldet, beschränken wir uns in diesem Abstrakt auf das Vorstellen einer der angewandten Methoden: Wie kann man die Ähnlichkeit der Daten selektiv auswerten um durch die Visualisierung ähnlicher Ausprägungen von Kostümen aus dem Lösungsrepository Hinweise auf Kostümmuster zu erhalten?

Methodischer Ansatz (exemplarisch)

Um ähnlich wirkende Artefakte (hier die konkreten Kostüme und deren Basiselemente wie Hosen, Pullover, etc.) zu identifizieren und zu visualisieren, machen wir uns die taxonomische Strukturierung (Barzen 2013) des Datenbestandes als Hintergrundwissen zunutze. Um eine detaillierte und strukturierte Erfassung der Kostüme zu gewährleisten, werden sie durch die Eingabe der kostümrelevanten Parameter (Attributbeschreibungen wie Farbe, Material, Zustand etc.), deren Wertebereich durch zugrundeliegende Taxonomien vorstrukturiert ist, beschrieben und im Lösungsrepository gespeichert. In der Literatur gibt es bewährte Verfahren, um aus einer Taxonomie, die Ähnlichkeiten von Objekten berechnen zu können. Insbesondere in der Biologie (Lord 2003), aber auch in der Linguistik (Jiang 1997) haben sich beispielsweise Verfahren zur Ähnlichkeitsbestimmung von Genotypen oder Sprachbausteinen bewährt. Dieser Ansatz soll auf die Kostümdateien übertragen werden.

Um die Ähnlichkeit von Artefakten zu bestimmen, wird die Struktur der Taxonomie als Hintergrundwissen einer Distanz-Funktion als Graph bereitgestellt. Aufbauend auf der Distanzmetrik, die Wu und Palmer für die Bestimmung konzeptueller Entfernung zwischen Begriffen (Palmer 1994) entwickelt haben, soll die Ähnlichkeit von Artefakten über die jeweiligen Distanzen ihrer Attributsausprägungen bestimmt werden. Eine Anwendung dieser Metrik auf die Attributsausprägungen „Farbe“ soll in Abbildung 1 demonstriert werden. Hier wird den Farbklassen „Hellblau“ und „Gelbtöne“ über

Bestimmung des gemeinsamen Elternknotens (C3) und der Kantenanzahlen von jeder Klasse (C1 und C2) zu dem Elternknoten (N1 und N2), sowie von Elternknoten zu Wurzelknoten (N3) durch die Anwendung der Distanzmetrik ein Ähnlichkeitswert von 0,4 zugeordnet (wobei 1 mit Identität und 0 mit völliger Verschiedenheit korrespondiert).

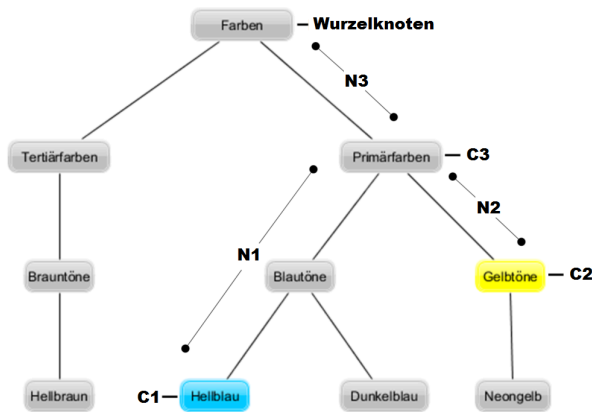


Abb. 1: Distanzbestimmung der Attribute

Visualisierung: Hinweise auf Musterkandidaten

Die Ergebnisse der Ähnlichkeitsanalyse können dann als Graph visualisiert werden. Abbildung 2 zeigt eine Beispielauswertung. Der Übersichtlichkeit halber haben wir die Anfrage auf Basiselemente, welche mit „negativ belegten“ Charaktereigenschaften assoziiert und von „weiblichen“ Rollen getragen werden, sowie auf die Kostumeigenschaften „Design“, „Farbe“ und „Zustand“ in der Ähnlichkeitsanalyse beschränkt. Die größte Ähnlichkeit bei den abgefragten Kostümen liegt bei „Unifarben“, „Gold/Silber“ und „Saubere“.

Diese so identifizierten Häufungen bzw. Cluster ähnlicher Attributsausprägungen können als Hinweise auf mögliche Kostümmuster gewertet werden. Wie die Ergebnisse bewertet werden und ob ein gehäuftes Auftreten ähnlicher Ausprägungen als Kostümmuster bewertet werden kann, bedarf einer weiterführenden Interpretation der Ergebnisse durch einen Domänenexperten.

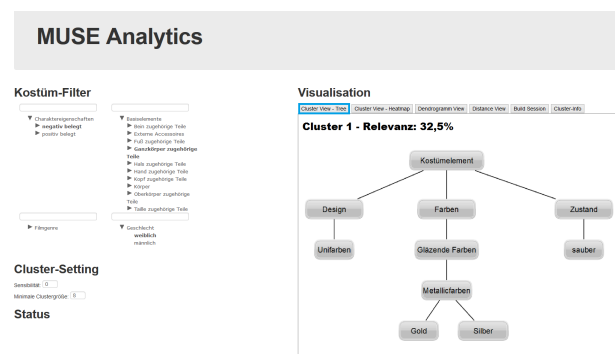


Abb. 2: Visualisierung der gemeinsamen Merkmale

Um die Analyse und Visualisierung einfach zugänglich zu machen, ist das Tool über ein Web Frontend erreichbar und erlaubt über Filtermöglichkeiten und unterschiedliche Visualisierung ein differenziertes Auswerten der Daten. Einen kleinen Ausblick auf die unterschiedlichen Ansätze und Diagrammarten, die das Tool unterstützt, soll durch die folgenden Screenshots (Abbildungen 3 und 4) gegeben werden.

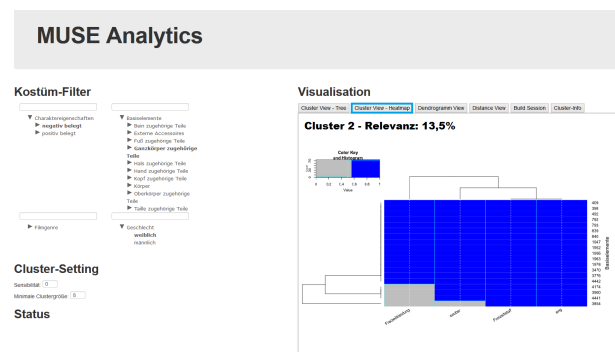


Abb. 3: Web Frontend: Heatmap

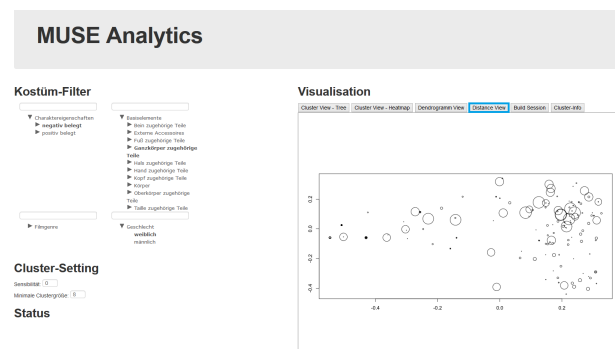


Abb. 4: Web Frontend: Distanzdiagramm

Über das Kostüm hinaus kann dieser Ansatz auch für andere Domänen der Digital Humanities fruchtbar gemacht werden. So zum Beispiel ist der Einsatz bei der Identifikation musikalischer Muster angedacht. Hier wurde

bereits mit der Erstellung musikalischer Taxonomien als Grundlage begonnen.

Fußnoten

1. Zur Projektbeschreibung s. auch <http://www.iaas.uni-stuttgart.de/forschung/projects/MUSE/>.

Bibliographie

Alexander, Christopher / Ishikawa, Sara / Silverstein, Murray / Jacobson, Max / Fiksdahl-King, Igrid / Angel, Shlomo (1977): *A Pattern Language*. Towns, Buildings, Constructions. Oxford: Oxford University Press.

Barzen, Johanna (2013): *Taxonomien kostümrelevanter Parameter*. Annäherung an eine Ontologisierung der Domäne des Filmkostüms. Technischer Bericht Nr. 2013/04, Universität Stuttgart.

Barzen, Johanna / Falkenthal, Michael / Hentschel, Frank / Leymann, Frank (2015): „Musterforschung in den Geisteswissenschaften: Werkzeugumgebung zur Musterextraktion aus Filmkostümen“, in: *Book of Abstracts zur Tagung der Digital Humanities im deutschsprachigen Raum 2015*, Graz 59-64 <http://gams.uni-graz.at/o:dhd2015.abstracts-poster> [letzter Zugriff 21. Januar 2016].

Fehling, Christoph / Barzen, Johanna / Falkenthal, Michael / Leymann, Frank (2014): „PatternPedia – Collaborative Pattern Identification and Authoring“, in: *Proceedings of PURPLSOC (Pursuit of Pattern Languages for Societal Change)*. The Workshop 2014. Krems 252-284.

Palmer, Martha / Wu, Zhibiao (1994): „Verb Semantics and Lexical Selection“, in: *ACL '94 Proceedings of the 32nd annual meeting of the Association for Computational Linguistics*, Stroudsburg, USA 133-138.

Schmitt, Ingo (2005): *Ähnlichkeitssuche in Multimedia-Datenbanken*. Retrieval, Suchalgorithmen und Anfragebehandlung. München: Oldenbourg Wissenschaftsverlag.

Jiang, Jay J. / Conrath, David W. (1997): „Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy“, in: *Proceedings of International Conference Research on Computational Linguistics (ROCLING X)*, Taiwan.

Lord, Phillip W. / Stevens, Robert D. / Brass, Andrew / Goble, Carole A. (2003): “Investigating semantic similarity measures across the gene ontology: The relationship between sequence and annotation”, in: *Bioinformatics* 19, 10: 1275-1283.