



Datenanalyse in den Digital Humanities – Eine Annäherung an Kostümmuster mittels OLAP Cubes

Michael Falkenthal¹, Johanna Barzen¹, Simon Dörner², Vadym Elkind²,
Jan Fauser², Frank Leymann¹, Tino Strehl²

¹Institute of Architecture of Application Systems,
University of Stuttgart, Germany
{falkenthal, barzen, leymann}@iaas.uni-stuttgart.de

²Herrmann Hollerith Zentrum
University of Applied Sciences Reutlingen, Germany
firstname.lastname@student.reutlingen-university.de

BIB_TE_X:

```
@inproceedings{Falkenthal2015,  
  author    = {Michael Falkenthal and Johanna Barzen and Simon Dörner and  
              Vadym Elkind and Jan Fauser and Frank Leymann and Tino Strehl},  
  title     = {Datenanalyse in den Digital Humanities – Eine Annäherung an  
              Kostümmuster mittels OLAP Cubes  
              },  
  booktitle = {Datenbanksysteme für Business, Technologie und Web (BTW),  
              16. Fachtagung des GI-Fachbereichs Datenbanken und  
              Informationssysteme (DBIS), 02.-06.3.2015 in Hamburg, Germany.  
              Proceedings.},  
  year      = {2015},  
  pages     = {1--4},  
  series    = {Lecture Notes in Informatics (LNI)},  
  publisher = {Gesellschaft für Informatik e.V. (GI)}  
}
```

© 2015 Gesellschaft für Informatik, Bonn

See also LNI-Homepage: <http://www.gi-ev.de/service/publikationen/lni>



Datenanalyse in den Digital Humanities – Eine Annäherung an Kostümmuster mittels OLAP Cubes

Michael Falkenthal¹, Johanna Barzen¹, Simon Dörner², Vadym Elkind², Jan Fauser²,
Frank Leymann¹, Tino Strehl²

Institut für Architektur von
Anwendungssystemen
Universität Stuttgart
Universitätsstraße 38
70569 Stuttgart
Nachname@iaas.uni-stuttgart.de¹

Hochschule Reutlingen
Herrmann Hollerith Zentrum
Alteburgstr. 150
72762 Reutlingen
Vorname.Nachname@student.reutlingen-
university.de²

Abstract: Im Film ist das Kostüm eines der prominentesten Gestaltungselemente, um Aussagen über eine Rolle, deren Charakter, Stimmung und Transformation, wie auch über Ort- und Zeitgegebenheiten zu kommunizieren. Durch Kostümmuster sollen Kostümbildner befähigt werden, effizient auf bewährte Designlösungen zurückgreifen zu können. Diese Demo zeigt, wie generelle Designprinzipien aus einer großen Anzahl an Kostümen aus Filmen für die Entwicklung dieser Kostümmuster mittels OLAP Cubes abstrahiert werden können. Um generelle Designprinzipien feststellen zu können, werden Kostüme über kategoriale Merkmalstaxonomien beschrieben und in verschiedenen Aggregationsstufen ausgewertet. Die Abstraktion von generellen Lösungen für Kostümmuster wird durch Drill-Down und Roll-Up Mechanismen unterstützt.

1 Einleitung und Motivation

Muster und Mustersprachen sind in vielen Domänen ein weit verbreitetes Mittel, um erprobtes Lösungswissen zu wiederkehrenden Problemen zu erfassen und dieses Wissen effizient für andere nutzbar zu machen. Aus der Architektur kommend [AIS77], hat sich das Konzept des Musters in der Informatik weit verbreitet. Dies lässt sich durch die Vielzahl an Publikationen aus den verschiedensten Bereichen der Informatik belegen, wie z.B. Cloud-Applikationsarchitektur, Unternehmens-IT-Architektur, Green-IT, Applikationssicherheit oder Benutzerschnittstellendesign, um nur einige zu nennen.

Aber auch in anderen Domänen, wie der Pädagogik oder der Produktentwicklung, hat das Konzept des Musters erfolgreiche Anwendung gefunden. Auch in der Domäne von Kostümen in Filmen erscheint der Einsatz von Mustern vielversprechend, indem adäquate textile Ausdrücke für Charaktere, die einfach vom Rezipienten verstanden werden (wie kommuniziert man beispielsweise einen guten Sheriff gegenüber einem bösen Sheriff mittels eines Kostüms?) als Lösungen wiederkehrender Designprobleme eines Kostümbildners in Mustern abgefasst werden [SBL12]. Als ein Projekt der Digital Humanities, bei dem sich geisteswissenschaftlichen Fragestellungen mittels neuer Methoden aus der Informatik genähert werden soll, wird durch interdisziplinäre

Forschung von Informatikern und Geisteswissenschaftlern eine Mustersprache für Kostüme im Film entwickelt [SBL12]. Basierend auf einem repräsentativen Fundus an Filmen (60 Filme mit jeweils ca. 200 Kostümen) wird eine Datenbasis – auch Lösungsrepository genannt [FBF14] – angelegt, in der Kostüme durch relevante Parameter, organisiert in Taxonomien, beschrieben werden [Ba13]. Jedes Kostüm in dieser Datenbasis repräsentiert eine konkrete Lösung für ein Designproblem. Im Gegensatz zu konkreten Lösungen im Lösungsrepository beinhalten Muster abstrahierte, generelle Lösungsprinzipien [FBB14]. Für die Formulierung von Mustern muss daher die Quintessenz aus einer Menge von Kostümen mit gleichem Effekt bezüglich eines Designproblems extrahiert werden [BL14].

In der Demo wird gezeigt, wie die Datenbasis mit Hilfe von Online Analytical Processing (OLAP) Cubes ausgewertet werden kann, sodass Muster aus der Menge an konkreten Lösungen abstrahiert werden können. Dazu wird eine Werkzeugumgebung vorgestellt, die eine automatisierte Aufbereitung und Übertragung der Daten aus der Datenbasis in ein Data Warehouse (DWH) ermöglicht. Weiterhin wird gezeigt, wie das Datenbankschema des DWH aufgebaut sein muss, um performante OLAP Cubes für die Auswertung des Datenbestandes zu ermöglichen und wie die Analyse der Daten aus filmwissenschaftlichen Gesichtspunkten erfolgt.

Nach der einleitenden Motivation in diesem Abschnitt, wird das grundlegende methodische Konzept der Auswertung in Zusammenhang mit dem Datenmodell der auszuwertenden Kostüme in Abschnitt 2 beschrieben. In Abschnitt 3 wird die Architektur des zu demonstrierenden Systems erläutert, während diese Arbeit mit einer Beschreibung des Ablaufs der Demo in Abschnitt 4 schließt.

2 Datenmodell und methodisches Konzept

Dem vorgestellten Konzept liegt eine Datenbasis mit konkreten Designlösungen (Kostümen) zugrunde. Ein *Kostüm* besteht aus mindestens einem *Basiselement*, wobei ein Basiselement einem Kleidungsstück wie z.B. einem Hemd entspricht. Ein Basiselement setzt sich aus mehreren *Teilelementen* wie Hemdkragen oder Knopfleiste zusammen. Basis- und Teilelemente werden durch Attribute wie Farbe, Form, Material oder Zustand beschrieben, während zur Beschreibung eines Kostüms Attribute wie Charaktereigenschaft, Funktion, Zustand oder Körpermodifikation herangezogen werden. Damit Kostüme im Kontext von rollen- bzw. filmspezifischen Attributen ausgewertet werden können, werden sie mit genau einer *Rolle* verknüpft, welche wiederum genau einem *Film* zugeordnet ist. Eine Rolle wird durch Attribute wie Beruf, Alter, Familienstand, Geschlecht oder Stereotyp spezifiziert. Attribute wie Erscheinungsjahr, Genre, Produktionsort oder Stil werden zur Beschreibung eines Films verwendet. Eine komplette Auflistung der Attribute aller Entitäten und deren Semantik ist in [Ba13] zu finden.

Um wiederkehrende Kostümmuster abstrahieren zu können, müssen Ausprägungen von Attributen und Attributkombinationen der Kostüme in der Datenbasis einfach vergleichbar gemacht werden. Zu diesem Zweck werden Taxonomien vorgegeben, die

den Wertebereich der Attribute festlegen und eine einheitliche Begrifflichkeit bei der Beschreibung von Kostümen gewährleisten [Ba13]. Die Attribute einer Entität können somit als pivotierbare Dimensionen mit kategorialen Ausprägungen verwendet werden, um neue Sichten zur Analyse des Datenbestandes zu generieren. Zusätzlich definiert jede Taxonomie eine Hierarchie von Super- und Subtypen. Dadurch können Kostüme über den Vergleich von Attributkombinationen über unterschiedliche Abstraktionsstufen hinweg auf gemeinsame Ausprägungen hin untersucht werden.

Abbildung 1 zeigt exemplarisch ein Auswertungsszenario, in dem analysiert wird, wie oft Basiselemente des Typs Kleid eine spezifische Farbausprägung haben. Dafür werden Basiselementtypen vertikal und Farben horizontal angeordnet. Die Zahlen im rechten Teil der Abbildung entsprechen dem Auftreten der entsprechend vertikal bzw. horizontal zugehörigen Merkmalsausprägungen. So existieren im Beispiel insgesamt 4 Basiselemente des Typs Kleid mit der Farbe Rot sowie 3 gelbe und 2 blaue. Durch die hierarchische Struktur der Basiselementtypen kann die Analyse detailliert werden, indem durch einen Drill-Down angezeigt wird, wie viele Basiselemente der Subtypen Abendkleid und Sommerkleid entsprechend rot, gelb und blau sind. Damit wird ersichtlich, dass bei der Aggregation das Auftreten von Merkmalsausprägungen von Subtypen aufsummiert wird. Der höchste auswertbare Detaillevel ist mit den Blattelementen einer Taxonomie festgelegt, im Beispiel ist dies für die Basiselementtypen das schulterfreie Kleid und das One-Shoulder Kleid. Über dieses Prinzip der Aggregation von konkreten Typen hin zu abstrakteren Typen – Roll-Up genannt – können Zusammenhänge auf dem Level von Merkmalskategorien ausgewertet werden.

		Farbe			
		Rot	Gelb	Blau	
Kleidung	Kleid	4	3	2	Aggreg.lvl 2
	Abendkleid	3	2	1	Aggreg.lvl 1
	Schulterfreies Kleid	1	2	0	Detaillevel
	One-Shoulder Kleid	2	0	1	
	Sommerkleid	1	1	1	

Abbildung 1: Aggregation über Kategorieebenen von Merkmalstaxonomien

3 Systemumgebung

Die in den ausgewählten Filmen vorkommenden Kostüme werden durch Domänenexperten betrachtet und über eine Webanwendung in eine MySQL Datenbank eingetragen. Mittels der Microsoft Integration Services werden die Daten aus der MySQL Datenbank in einem ETL-Prozess aufbereitet und in ein Microsoft SQL

Server 2012 DWH geladen. Das erstellte DWH verwendet das Galaxy-Schema. Somit können mehrere Faktentabellen erstellt werden, die gezielt auf bestimmte Analyseszenarien ausgelegt sind und nur die Szenario-relevanten Attribute enthalten. Über Microsoft Analysis Services werden OLAP Cubes bereitgestellt, die zur Auswertung des Datenbestands in Microsoft Excel eingebunden und visualisiert werden können.

4. Beschreibung der Demo-Session

In der Demo wird insbesondere darauf eingegangen, wie eine Werkzeugumgebung in einem Digital Humanities Projekt gestaltet werden kann, sodass neue informationstechnische Auswertungsmethoden für Geisteswissenschaftler bereitgestellt werden und gleichzeitig so viel informationstechnische Komplexität wie möglich für die Endanwender verborgen bleibt. Dafür umfasst die Demo alle Schritte vom Vorbereiten der Rohdaten bis hin zur Auswertung in Excel. Die wichtigsten Punkte des ETL-Prozesses werden erläutert, die zur Bereitstellung des DWH führen. Der Aufbau des DWH wird gezeigt. Anschließend wird die Einbindung und die Vorbereitung des DWH mittels Microsoft SQL Data Tools vorgestellt, in denen die Cube Einstellungen vorgenommen werden. Weiterhin zeigt die Demo wie der Datenbestand in Excel ausgewertet werden kann. Der Endanwender wählt die benötigten Attribute aus, um den Datenbestand mit bestimmten Fragestellungen zu untersuchen. Mittels Pivot-Tabellen können die Daten in jeder Dimension analysiert und als Pivot-Charts visualisiert werden. Durch das Auftreten von Spitzen in der Häufigkeitsverteilung der Pivot-Charts können erste Rückschlüsse auf mögliche Muster gezogen werden.

Literaturverzeichnis

- [AIS77] Alexander, C.; Ishikawa, S.; Silverstein, M.: A Pattern Language: Towns, Buildings, Construction. Oxford University Press, New York, 1977.
- [Ba13] Barzen, J.: Taxonomien kostümrelevanter Parameter: Annäherung an eine Ontologisierung der Domäne des Filmkostüms. Universität Stuttgart, Stuttgart, 2013.
- [BL14] Barzen, J.; Leymann, F.: Costume Languages as Pattern Languages. In: Proceedings of Pursuit of Pattern Languages for Societal Change - Preparatory Workshop 2014, Krems, 2014 (akzeptiert).
- [FBB14] Falkenthal, M.; Barzen, J.; Breitenbücher, B.; Fehling, C.; Leymann, F.: From Pattern Languages to Solution Implementations. In: Proceedings of the 6th International Conference on Pervasive Patterns and Applications, Venedig, 2014; S. 12–21.
- [FBF14] Fehling, C.; Barzen, J.; Falkenthal, M.; Leymann, F.: PatternPedia – Collaborative Pattern Identification and Authoring. In: Proceedings of Pursuit of Pattern Languages for Societal Change - Preparatory Workshop 2014, Krems, 2014 (akzeptiert).
- [SBL12] Schumm, D.; Barzen, J.; Leymann, F.; Ellrich, L.: A Pattern Language for Costumes in Films. In: Proceedings of the 17th European Conference on Pattern Languages of Programs (EuroPLOP), Irsee, 2012. ACM Press, New York, 2012.