

VIII Experimentelle Entwicklung eines Prototyps für das Monitoring von Entwicklungs- verläufen in einem Forschungsfeld

Karima Haddou ou Moussa, Peter Mutschke und Philipp Mayr
GESIS – Leibniz-Institut für Sozialwissenschaften, Köln

Vorbemerkung	261
1 Datenbasis und technische Basis	262
2 Vorselektion von Projekten und Indikatoren	266
3 Visualisierungsansätze der Indikatoren	268
3.1 Forschungsaktivität	268
3.2 Disziplinbereich	269
3.3 Förderart	270
3.4 Qualifizierungsarbeiten	272
3.5 Methodisches Paradigma	273
3.6 Erhebungs- und Analyseverfahren	274
3.7 Forschungsziel	275
3.8 Forschungslandkarte	275
4 Visualisierung von Indikatorenrelationen	277
5 Schlussfolgerung und Diskussion	280

Vorbemerkung

Die rasante Entwicklung der Informationstechnologien ermöglicht heutzutage in fast allen Wissenschaften die dauerhafte Speicherung großer Datenmengen. Diese Daten werden gesammelt, weil sie wertvolle, relevante Informationen enthalten, die zu bestimmten Zwecken analysiert und ausgewertet werden können. Allerdings ist das Auffinden dieser Informationen in unüberschaubaren Datenstapeln keine leichte Aufgabe. „Ein Bild sagt mehr als tausend Worte“, dieses Sprichwort beschreibt die Vorteile von Bildern gegenüber Sprache und damit die Methodik der wissenschaftlichen Visualisierung. Heutzutage spielt die Visualisierung von Daten eine immer größer werdende Rolle. Dabei werden große Datenmengen in visueller Form dargestellt, weil Bilder mehrere Informationen gleichzeitig liefern und dem Benutzer einen Einblick in die Struktur der Daten geben.

In diesem Kapitel wird ein Prototyp vorgestellt, dessen Ziel es ist, exemplarische Visualisierungen der in den vorangegangenen Kapiteln beschriebenen Indikatoren innerhalb eines experimentellen Prototyps zu demonstrieren, mit der Perspektive, Ideen für einen digitalen Monitoring-Dienst zu generieren, der Veränderungen eines Forschungsfeldes über die Zeit sichtbar macht.¹

¹ Bei diesem Beitrag handelt es sich um eine deutlich erweiterte, deutsche Fassung des Konferenzbeitrages von K. Haddou ou Moussa, U. Sondergeld, P. Mayr, P. Mutschke, und M. Rittberger (2014).

1 Datenbasis und technische Basis

Technologische Basis des Monitoring-Prototyps ist das Online-Portal SOFISwiki von GESIS. SOFISwiki (Sofis.gesis.org) ist eine Community-Plattform, die der Erfassung von und der Suche nach sozialwissenschaftlichen Forschungsprojekten von verschiedenen Fachdisziplinen (primär Soziologie, aber auch angrenzende Fächer wie Erziehungswissenschaft, Psychologie und Politikwissenschaft) aus dem deutschsprachigen Gebiet dient. Derzeit enthält SOFISwiki 53.980 Projektnachweise. Der Monitoring-Prototyp bedient sich eines Subsets von 9.122 abgeschlossenen Forschungsprojekten aus dem Bereich der Bildungsforschung der Jahre 1995 bis 2009. Es handelt sich also im Wesentlichen um dasselbe Korpus,² das in Kapitel II untersucht wurde.

SOFISwiki basiert auf der Semantic MediaWiki-Technologie. Semantic MediaWiki (Krotzsch, Vrandecic, Volkel, Haller & Studer, 2007) (SMW) ist eine erweiterte Version von MediaWiki (Mediawiki.org), welches die Grundlage vieler Wiki-Anwendungen (zum Beispiel Wikipedia) ist. Der Zweck der SMW-Erweiterung besteht darin, ein schnelles semantisches Suchen und Finden von Daten in einem Wiki-System zu ermöglichen. Dazu werden nicht nur reine Text-Seiten auf einer Wiki-Seite hinterlegt, sondern diese werden auch mit Zusatzinformationen angereichert. Durch diese zusätzlichen Informationen, die sogenannten „Wiki-Attribute“, werden Beziehungen zwischen Wiki-Seiten beschrieben. Zur Herstellung von direkten Verbindungen zwischen diesen Seiten werden Hyperlinks gesetzt. Die Seitenrelation wird entweder durch typisierte Verweise und/oder durch Werte der Attribute realisiert. Seitennamen in einem MediaWiki-System bestehen aus einem Namensraum und einem ausgewählten Namen. Namensräume sind Strukturierungskonzepte, die zur Gruppierung von Seiten dienen.

MediaWiki verfügt unter anderem über folgende Namensräume: Kategorie, Attribut und Vorlage. Kategorien (Semantic-mediawiki.org, Editing) ermöglichen die Einordnung von Seiten. Eine Seite kann einer oder mehreren Kategorien zugeordnet werden. Die Zuweisung einer Seite zu einer Kategorie erfolgt durch folgende Syntax: `[[Kategorie:Kategoriename]]`; beispielsweise besagt der Wiki-Text `[[Kategorie:MoBi]]`, dass MoBi aus dem Namensraum „Kategorie“ stammt und „MoBi“ folglich der Name der betreffenden Kategorie ist. Alle MoBi-Projekte sind der MoBi- und (gegebenenfalls sowohl der MoBi – als auch) der Projektkategorie zugeordnet. Daher werden sie am Seitenende wie folgt dargestellt: *Kategorien: Projekte | MoBi*. Attribute (Semantic-mediawiki.org, Properties and types)

2 Da SOFISwiki auf eine „lebende“ Datenbank aufsetzt, gibt es im Monitoring-Prototyp geringfügige Abweichungen zum Untersuchungskorpus von Kapitel II.

sind als Kategorien für Werte in Wiki-Seiten zu betrachten, mit denen Daten semantisch beschrieben werden können. Den Nutzern ist es erlaubt, selbst nach diesem einfachen Schema: `[[Attributname::Attributwert]]` Attribute zu erstellen. Diese Wiki-Schreibweise ordnet dem angegebenen Attribut einen Attributwert zu und zeigt diesen Wert oder die Wertmenge an der entsprechenden Stelle auf der Seite an: Attributwert, zum Beispiel `[[Jahrgang::1997]]`.

Auf den semantisch angereicherten Wiki-Seiten des SMW können viele Informationen über die einzelnen Seiten anhand von Attributen explizit dargestellt werden. Sie können für verschiedene Arten von Daten verwendet werden wie Zahlen, Datumsangaben oder geografische Koordinaten, wobei jedem Attribut ein Datentyp zugewiesen wird. Andernfalls werden Annotationen, die keine bzw. unpassende Datentypen haben, unberücksichtigt gelassen. Für Attributwerte existieren viele unterschiedliche Datentypen wie String, Page oder Number. So ist beispielsweise die Eigenschaft „Personen“ eines Forschungsprojektes ein Attribut vom Typ String. Somit besitzt jede einzelne Wiki-Seite eine Liste verschiedener Attribute und deren Werte, die „Metadatenschema“ genannt wird. Für dieses Schema existieren zwei Arten von Darstellungen: Nutzer- und Entwickleransicht. Abbildungen 1 und 2 stellen Screenshots der Attribute-Liste eines sozialwissenschaftlichen Projekts dar. Die Nutzeransicht (Abbildung 1) zeigt nur die wichtigsten Informationen an, die zur Beschreibung eines Projekts dienen und für den Nutzer relevant sind. Diese beinhalten die Metadaten des Projekts (Titel, Jahr, Abstract, Methoden, Schlagwörter sowie die beteiligten Personen und Institutionen). Handelt es sich um ein gefördertes Projekt, wird die Fördereinrichtung genannt.

Die Entwickleransicht (Abbildung 2) hingegen beinhaltet viele unterschiedliche Informationen in Form anwählbarer Suchsymbole, die das schnelle Auffinden von Seiten mit identischen Annotationen in Abfragen ermöglicht. Die Entwickleransicht besteht aus zwei Spalten. Die erste Spalte (links) listet die vorhandenen Attribute auf, während die zweite Spalte (rechts) die zugeordneten Attributwerte abbildet. Die Attribute bei der SOFISwiki-Datenbank sind keine atomaren Felder, sondern multiple Felder, die beliebig viele verschiedene Inhalte gleichzeitig erlauben. Mehrfachbelegung eines Feldes ist also möglich. Beispielsweise beinhaltet das Feld „Methode“ in Abbildung 2 sieben Deskriptoren (Beobachtung, empirisch, Feldforschung, Längsschnitt, Querschnitt, Zeitbudget und empirisch-quantitativ). Vorlagen (Semantic-mediawiki.org) sind normale Wiki-Seiten, die sich nach dem Transklusionsprinzip, das heißt Einbindung von Seiteninhalten in andere Seiten, als Baustein (bei häufig verwendeten Elementen) einschließen lassen. Sie werden im MediaWiki als Hilfsmittel eingesetzt, um beispielsweise Attribute gleicher Annotationen zu erstellen. Die Syntax zur Verwendung von Vorlagen lautet `{{Template: page name}}`.

Untersuchung von Leistungsvoraussetzungen und Dispositionen bei Kindern des frühen Schulkindalters (1.-4. Schuljahr)	
Erfassungsnr.:	19952251
Laufzeit von:	1994/01/15
Laufzeit bis:	1997/12/15
Art der Forschung:	Auftragsforschung, gefördert
Kontakt:	Leiter (Tel. 0561-8044416)
Institutionen	
Forschungseinrichtung:	<u>Institut für Sport und Sportwissenschaft (Kassel)</u>
Auftraggeber:	<u>Land Hessen Kultusministerium (Wiesbaden)</u>
Auftraggeber:	<u>Landessportbund Hessen e.V. (Frankfurt am Main)</u>
Finanzierer:	<u>Deutsche Forschungsgemeinschaft (Bonn)</u>
Finanzierer:	Auftraggeber
Beteiligte Personen	
Leitung:	Prof.Dr. Dietrich Martin
Bearbeitung:	Sabine Karoß
Bearbeitung:	Kerstin König
Inhalt	
<p>Untersuchung derjenigen Leistungsvoraussetzungen und Dispositionen, die bei Kindern des frühen Schulkindalters (1.-4. Schuljahr) die sportliche Talentbestimmung gewährleisten und durch den Ausprägungsgrad der Leistungsentwicklung Prognosen für spätere Leistungen erlauben. Normierung talentbestimmender Merkmale. Ziele und Aufgabenfelder der wissenschaftlichen Begleitung von Talentfördersystemen im Sport. Erarbeitung der Wirksamkeit von Nachwuchstrainingssystemen.</p>	
Geographischer Raum:	Hessen
Grundlegender theoretischer / methodischer Ansatz (Forschungsparadigma):	Theorien zum sportlichen Talent; Theorien zu Erbanlagen und genetischen Dispositionen; Längsschnitterhebung über 3 Jahre.
Zeitraum:	1994-1997
Schlagwörter:	Kind, Prognose, Training, Bundesrepublik Deutschland, Begabung, Hessen, Grundschule, Leistungssport, Schulkind, Leistungsfähigkeit, Normierung
Hauptklassifikation:	<u>Bildungswesen Primarbereich</u>
Nebeklassifikation:	<u>Freizeitforschung</u>
Methode:	<u>Beobachtung, empirisch, Feldforschung, Längsschnitt, Querschnitt, Zeitbudget, empirisch-quantitativ</u>

Abb. 1: Nutzeransicht im SOFISwiki. Anzeige von Metadaten eines Forschungsprojekts

Untersuchung von Leistungsvoraussetzungen und Dispositionen bei Kindern des frühen Schulkindalters (1.-4. Schuljahr)	
<u>Abgabe Erhebungsdaten</u>	eventuell + ↗
<u>Auftraggeber</u>	Land Hessen Kultusministerium (Wiesbaden) + Ⓜ , Landessportbund Hessen e.V. (Frankfurt am Main) + Ⓜ
<u>Finanzierer</u>	Deutsche Forschungsgemeinschaft (Bonn) + Ⓜ
<u>Finanzierer Freitext</u>	Auftraggeber
<u>Forschungsart</u>	Auftragsforschung + ↗ , gefördert + ↗
<u>Forschungseinrichtung</u>	Institut für Sport und Sportwissenschaft (Kassel) + Ⓜ
<u>Forschungseinrichtungs counter</u>	1 + Ⓜ
<u>Förderart</u>	Auftragsforschung + Ⓜ , Gefördert + Ⓜ
<u>Geographischerraum</u>	Hessen + ↗
<u>Hauptklassifikation</u>	Bildungswesen Primarbereich + Ⓜ
<u>Hauptklassifikationsuch</u>	Erziehungswissenschaft + Ⓜ
<u>Id</u>	19952251 + ↗
<u>Informationsquelle</u>	GESIS, Forschungserhebung 1995 + ↗
<u>Inhalt de</u>	Untersuchung derjenigen Leistungsvorausset ... irksamkeit von Nachwuchstrainingsystemen.
<u>Institution</u>	Institut für Sport und Sportwissenschaft (Kassel) + Ⓜ , Land Hessen Kultusministerium (Wiesbaden) + Ⓜ , Landessportbund Hessen e.V. (Frankfurt am Main) + Ⓜ , Deutsche Forschungsgemeinschaft (Bonn) + Ⓜ
<u>Jahrgang</u>	1995 + Ⓜ
<u>Jahrgang ende</u>	1997 + Ⓜ
<u>Jahrgang start</u>	1994 + Ⓜ
<u>Klassifikation</u>	Freizeitforschung + Ⓜ
<u>Kontakt</u>	Leiter (Tel. 0561-8044416)
<u>Laufzeit Bis</u>	15 Dezember 1997 + ↗
<u>Laufzeit Von</u>	15 Januar 1994 + ↗
<u>Methode</u>	Beobachtung + Ⓜ , Empirisch + Ⓜ , Feldforschung + Ⓜ , Längsschnitt + Ⓜ , Querschnitt + Ⓜ , Zeitbudget + Ⓜ , Empirisch-quantitativ + Ⓜ
<u>Methodisches Paradigma</u>	Empirisch + Ⓜ , Empirisch-quantitativ + Ⓜ
<u>Personen</u>	Dietrich Martin + ↗ , Sabine Karoß + ↗ , Kerstin König + ↗
<u>Pjjsn</u>	36709 + ↗
<u>Rnr</u>	19.942.984 + ↗
<u>Schlagworte</u>	Kind + ↗ , Prognose + ↗ , Training + ↗ , Bundesrepublik Deutschland + ↗ , Begabung + ↗ , Hessen + ↗ , Grundschule + ↗ , Leistungssport + ↗ , Schulkind + ↗ , Leistungsfähigkeit + ↗ , Normierung + ↗

Abb. 2: Entwickleransicht im SOFISwiki. Anzeige der internen Darstellung von Attributen und Attributwerten

2 Vorselektion von Projekten und Indikatoren

Der Prototyp ermöglicht eine Vorselektion der Projekte, die anhand des Projektstatus (abgeschlossen, beginnend, laufend) und des geografischen Raums (Deutschland oder Gesamtbestand, das heißt Deutschland, Österreich und die Schweiz) visualisiert werden können. Diese Möglichkeit wird durch eine eigens dafür entwickelte PHP-Extension unterstützt, die als Spezialseite angelegt wird. Da die MoBi-Datenbasis nur abgeschlossene Projekte enthält, wurden bei der bisherigen Implementierung auch nur diese berücksichtigt (siehe Abbildung 3).

Projekt-Auswahlkriterien

Projektstatus:

angefangen
 laufend
 abgeschlossen

Geographischer Raum:

Gesamtbestand
 Deutschland

SUBMIT

Abb. 3: Projektauswahlkriterien nach Projektstatus und geografischem Raum

Nach der Selektion des Projektstatus und des geografischen Raums kann der Anwender den zu visualisierenden Indikator (absteigende Reihenfolge: Forschungsaktivität, Disziplinbereich, Förderart, Qualifizierungsarbeit, Methodisches Paradigma, Erhebungs- und Analyseverfahren sowie Forschungsziel) und den Zeitraum auswählen. Die Nutzer haben die Möglichkeit, entweder den Zeitraum selbst zu bestimmen (selbst definierter Zeitraum) oder den Standardzeitraum (Gesamtzeitraum) (Abbildung 4) zu betrachten. Für diesen Zweck wurde eine eigene PHP-Extension entwickelt (Abbildung 5), die eine Jahresselektion ermöglicht und für alle Indikatoren einsetzbar ist.

Selektion der Indikatoren und Zeitraum

Indikatoren:

- Forschungsaktivität
- Disziplinbereich
- Förderart
- Qualifizierungsarbeit
- Methodisches Paradigma
- Erhebungs- und Analyseverfahren
- Forschungsziel

Zeitraum:

- Gesamtzeitraum
- selbstdefinierter Zeitraum**

Abb. 4: Selektierbare Indikatoren und Zeitraum

Forschungsaktivität

Startjahr: Endjahr:

keine Auswahl

keine Auswahl

1995

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009




Abb. 5: Selektion der Jahre – Beispiel für Forschungsaktivität

3 Visualisierungsansätze der Indikatoren

Um auf der Basis der SOFISwiki-Struktur Indikatoren visualisieren zu können, wurden bestimmte Attribute in der Entwickleransicht selektiert, deren Werte als Basis für die Visualisierung dieser Indikatoren dienen. Weil die Attribute in der SOFISwiki-Datenbank mehrfach belegt werden können, kann diese Eigenschaft bei der Datenanalyse zu einem inkorrekten Ergebnis führen. Daher beschränkt sich dieses Kapitel auf die Vorstellung eines exemplarischen Monitoring-Dienstes mithilfe der Technologie Semantic MediaWiki und ihrer Visualisierungsmöglichkeiten. Für den Indikator „Disziplinbereich“ konnte ein vorhandenes Feld abgefragt und ausgewertet werden. Für die anderen Merkmale – Forschungsaktivität, Förderart, Qualifizierungsarbeit, methodisches Paradigma, Erhebungs- und Analyseverfahren – sowie Forschungsziel wurden neue Attribute aus den bereits bestehenden SOFISwiki-Attributen anhand von Templates und zusätzlichen PHP-Prozeduren generiert. Die Mehrfachbelegung der Felder wurde dabei nicht behandelt. Diese Problematik lässt sich rechnerisch bzw. technisch leider nicht beheben, weil unklar ist, welche Methoden, Verfahren und Ziele bzw. in welchem Umfang diese eingesetzt wurden.

Für die visuelle Darstellung der Indikatoren wurden die von Semantic MediaWiki zur Verfügung gestellten Visualisierungserweiterungen wie Sparkline, D3, jqPlot (Semantic-mediawiki.org) eingesetzt und entsprechend angepasst. Anhand dieser Erweiterungen können beispielsweise Linien- und Balkendiagramme sowie Ringdiagramme und so genannte Treemaps dargestellt werden. Damit die Indikatoren mithilfe dieser Verfahren visualisiert werden können, wurden programmiertechnische Anpassungen auf der Basis von Templates durchgeführt. Der Monitoring-Prototyp generiert alle Visualisierungen on-the-fly aus der Datenbasis.

3.1 Forschungsaktivität

Der Indikator Forschungsaktivität gibt Auskunft über die Anzahl der Projekte pro Jahr und wurde im MoBi-Projekt als Anzahl abgeschlossener Projekte definiert. Für diesen Indikator wurde das Attribut „Jahrgang Ende“ für die Entwickleransicht mit einem Wiki-Template generiert. Der Wert dieses Attributs wurde aus dem SOFISWiki-Feld „Laufzeit bis“ durch eine entwickelte PHP-Extension extrahiert und im Zahlenformat in das neue Feld eingefügt. Für die Darstellung dieses Indikators wurden diverse Ergebnisformate des Semantic MediaWiki getestet (wie Sparkline und jqPlot). Abbildung 6 zeigt das Ergebnis einer jqPlot-Darstellung (Semantic-mediawiki.org) in Form eines Balkendiagramms.

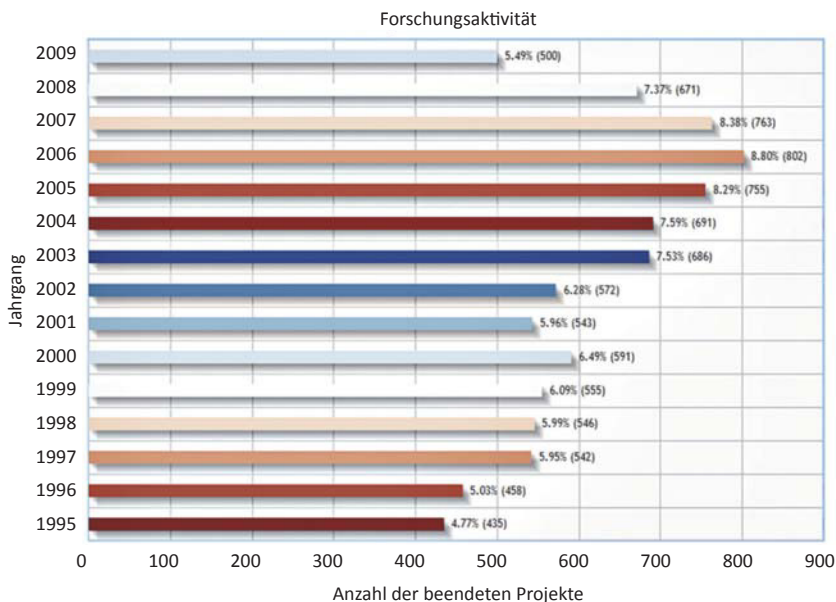


Abb. 6: Darstellung der Forschungsaktivität (Anzahl der abgeschlossenen Projekte pro Jahr) als Balkendiagramm

3.2 Disziplinbereich

Um einen Überblick über das Forschungsfeld zu geben, werden die Projekte zu Disziplinbereichen zusammengefasst dargestellt. Der Indikator „Disziplinbereich“ ordnet alle Forschungsprojekte den zwölf Disziplinbereichen zu, die auf Klassen der Klassifikation Sozialwissenschaften von GESIS (Gesis.org) beruhen. Diese lauten wie folgt:

- Gesellschafts- und Geisteswissenschaften
- Soziologie
- Bevölkerungswissenschaft
- Politikwissenschaft
- Erziehungswissenschaft
- Psychologie
- Kommunikationswissenschaften
- Wirtschaftswissenschaften
- Sozialpolitik
- Arbeitsmarkt- und Berufsforschung
- Interdisziplinäre Fachgebiete
- Geschichtswissenschaft

Für die Visualisierung des Indikators „Disziplinbereich“ wurden folgende, von SMW bereitgestellte Darstellungsmöglichkeiten verwendet: Sparkline, jqPlot, D3 und Tag Cloud. Beispielsweise zeigt Abbildung 7 die Ergebnisformate als Tag Clouds (Semantic-mediawiki.org).



Abb. 7: Tag Cloud-Visualisierung der Disziplinverteilung

3.3 Förderart

Um Arten der Förderung bzw. Finanzierung von Projekten darzustellen, wird zwischen Eigenprojekten, geförderter Forschung und Auftragsforschung unterschieden. Die Förderart wird aus dem Attribut „Forschungsart“ ermittelt. Im SOFISwiki kann das Attribut die folgenden neun Merkmale oder Kombinationen daraus enthalten:

- Auftragsforschung
- gefördert
- Eigenprojekt
- Gutachten
- Dissertation
- Habilitation
- Abschlussarbeit
- Sonstige
- Keine Angabe

Durch den Einsatz der eigens entwickelten PHP-Extension und eines Wiki-Templates wird das neue Feld „Förderart“ erzeugt, das nur die drei Merkmale Eigenprojekte, geförderte Forschung und Auftragsforschung identifiziert. Mithilfe der Erweiterungen D3 (Semantic-mediawiki.org) und jqPlot wurden die Ergebnisse in Abbildung 8 (D3: Bubble und Treemap, jqPlot: Pie und Donut) erzielt.

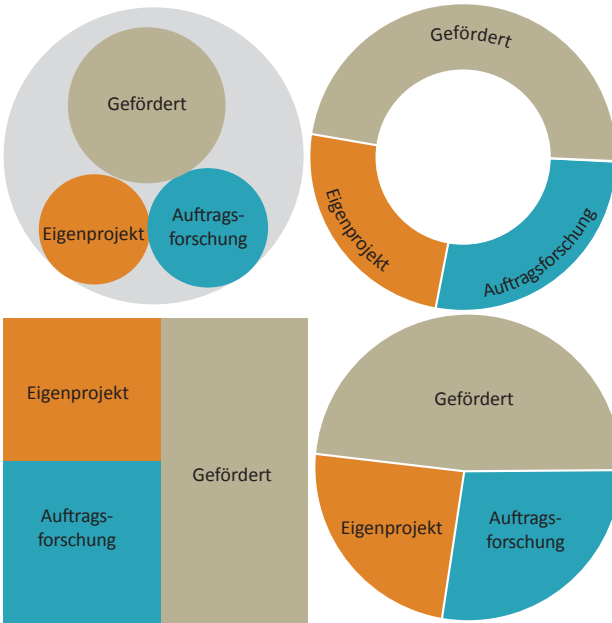


Abb. 8: Verschiedene Darstellungen des Indikators Förderart

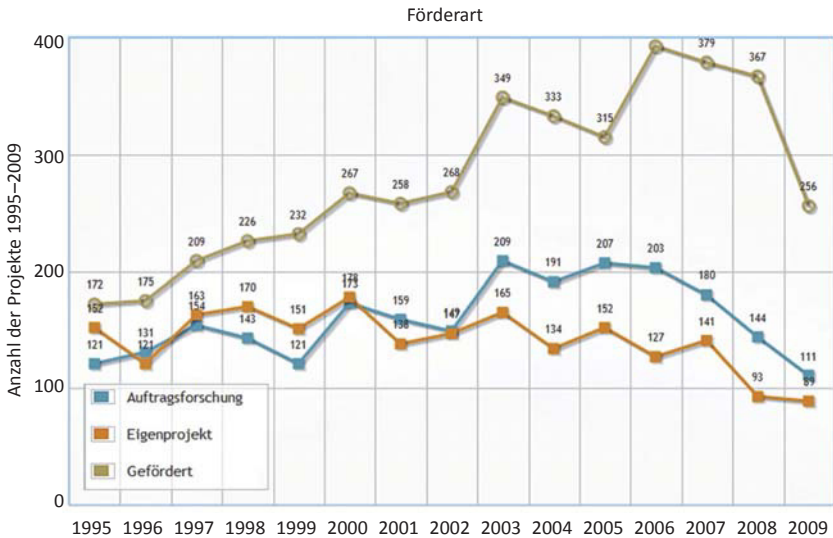


Abb. 9: Anzahl der Projekte mit unterschiedlichen Finanzierungsarten pro Jahr

Damit der Indikator „Förderart“ im zeitlichen Verlauf visualisiert werden kann, wurde die Jqplotseries-Erweiterung (Semantic-mediawiki.org) auf die aufbereiteten Förderarten angewandt (siehe Abbildung 9).

Aufgrund der möglichen Mehrfachbelegung kann ein Projekt gleichzeitig sowohl als „Eigenprojekt“ als auch als „gefördert“ oder als „Auftragsforschung“ indiziert werden. Somit ist eine eindeutige Zuordnung eines Projekts zu einer einzigen Förderart nicht möglich, mit dem Ergebnis, dass nur absolute Zahlen visualisiert werden und auf die Berechnung relativer Zahlen verzichtet wird.

3.4 Qualifizierungsarbeiten

Die Förderung des wissenschaftlichen Nachwuchses ist bei der Evaluierung von Forschungsorganisationen ein wichtiger Indikator, der über die Anzahl abgeschlossener Qualifizierungsarbeiten erhoben wird. Für die Darstellung von Projekten mit dem Ziel einer Dissertation oder Habilitation wird der Indikator „Qualifizierungsarbeit“ gebildet. Diese Werte werden aus dem Feld „Forschungsart“ programmiertechnisch ausgelesen und in das neu generierte Feld „Qualifizierungsarbeit“ eingefügt. Bei der visuellen Präsentation der Qualifizierungsarbeiten werden dieselben Visualisierungserweiterungen wie beim Indikator „Förderart“ verwendet. Abbildung 10 zeigt in Form eines Säulendiagramms, wie sich die Anteile von Dissertationen und Habilitationen über den Zeitraum von 1995 bis 2009 verändern.

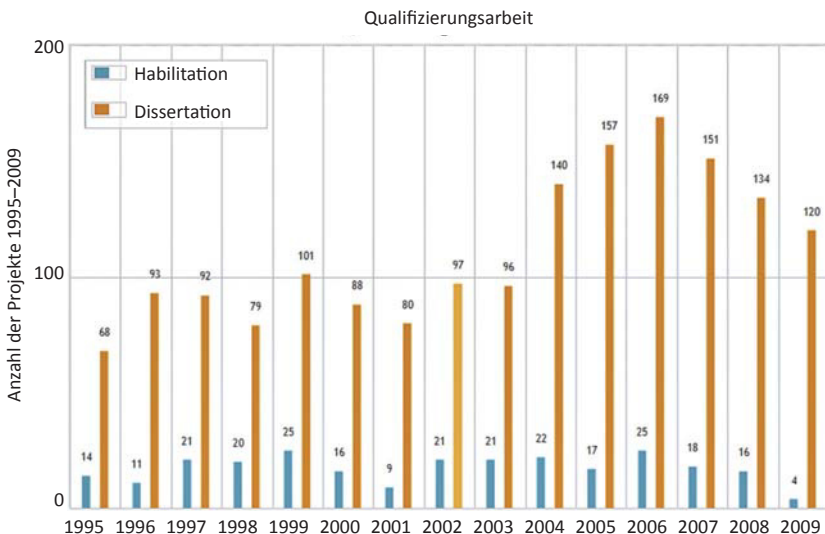


Abb. 10: Balkendiagramm-Darstellung der Qualifizierungsarbeiten pro Jahr

3.5 Methodisches Paradigma

Um zu prüfen, welche empirischen Methoden in den Forschungsprojekten eingesetzt werden, wurde das Feld „Methodisches Paradigma“ aus dem SOFISwiki-Attribut „Methode“ generiert, das nur die Deskriptoren „empirisch-quantitativ“ und „empirisch-qualitativ“ beinhaltet. Bei der Indexierung der empirischen Methoden stehen den SOFIS-Dokumentaren folgende Begriffe zur Verfügung: „empirisch-quantitativ“, „empirisch-qualitativ“ und „empirisch“. Während der Begriff „empirisch“ auch allein vergeben werden kann, findet bei den Begriffen „empirisch-quantitativ“ und „empirisch-qualitativ“ immer ein Upposting statt, das heißt, der Attributwert „empirisch“ wird zusätzlich automatisch vergeben. Das bedeutet, dass der Begriff empirisch allein vergeben werden kann, die Begriffe empirisch-quantitativ oder empirisch-qualitativ aber immer mit dem Begriff empirisch einhergehen. Das Attribut enthält neben den empirischen Methoden verschiedene Merkmale, die in SOFISwiki nach einer einheitlichen Methodenklassifikation vergeben werden. Diese Klassifikation beinhaltet zwei Facetten (Gesis.org):

1. Primäre Forschungskennzeichnung: anwendungsorientiert, Evaluation, deskriptive Studie, Dokumentation, empirisch, historisch, empirisch-qualitativ, empirisch-quantitativ, Grundlagenforschung, Methodenentwicklung, Theorieanwendung.
2. Methodisches Design und Erhebungsverfahren: Aggregatdatenanalyse, Aktenanalyse, Befragung, Beobachtung, Simulation, biografische Methode, Exploration, Fallstudie, Feldforschung, Gruppendiskussion, Inhaltsanalyse, Längsschnitt, Netzwerkanalyse, Sekundäranalyse.

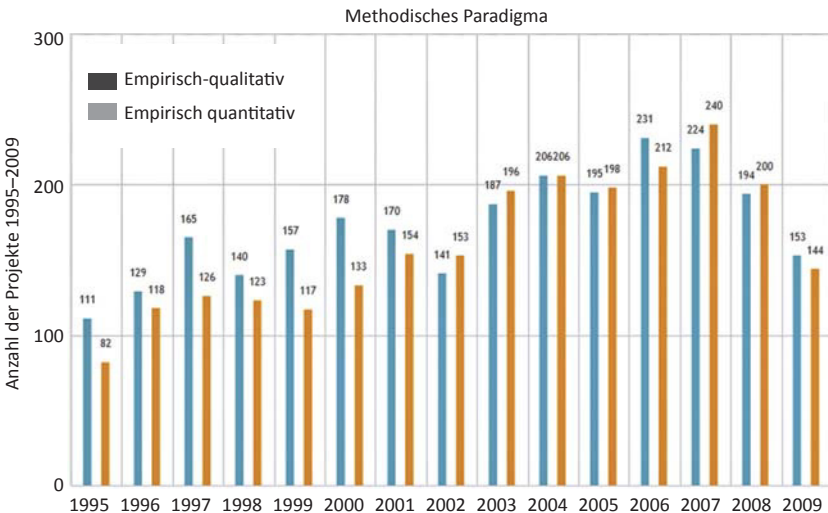


Abb. 11: Verteilung der empirisch-qualitativen und empirisch-quantitativen Methoden über den Zeitverlauf 1995 bis 2009

Methoddeskriptoren können mehrfach vergeben werden, wobei die primäre Forschungskennzeichnung immer angegeben sein muss. Wegen der möglichen Mehrfachbelegung des Attributs „Methode“ ist eine eindeutige Zuordnung eines Projekts zu einer einzigen Methode nicht möglich. Daher ist es sehr wahrscheinlich, dass Projekte teilweise in beide Kategorien fallen. Abbildung 11 zeigt die Anzahl derjenigen Projekte, die empirisch-qualitative und/oder empirisch-quantitative Methoden verwenden.

3.6 Erhebungs- und Analyseverfahren

Der Indikator „Erhebungs- und Analyseverfahren“ liefert Informationen über konkrete Verfahren der Datenerhebung und Datenanalyse. Diese Informationen stammen aus der zweiten Facette der Methodenklassifikation. Daher wurde das Feld „Methoden“ erneut abgefragt und ausgewertet, um das neue Attribut zu erzeugen. Abbildung 12 stellt unter Berücksichtigung der Mehrfachbelegung die Verteilung einiger ausgewählter Erhebungs- und Analyseverfahren über die Zeit dar.

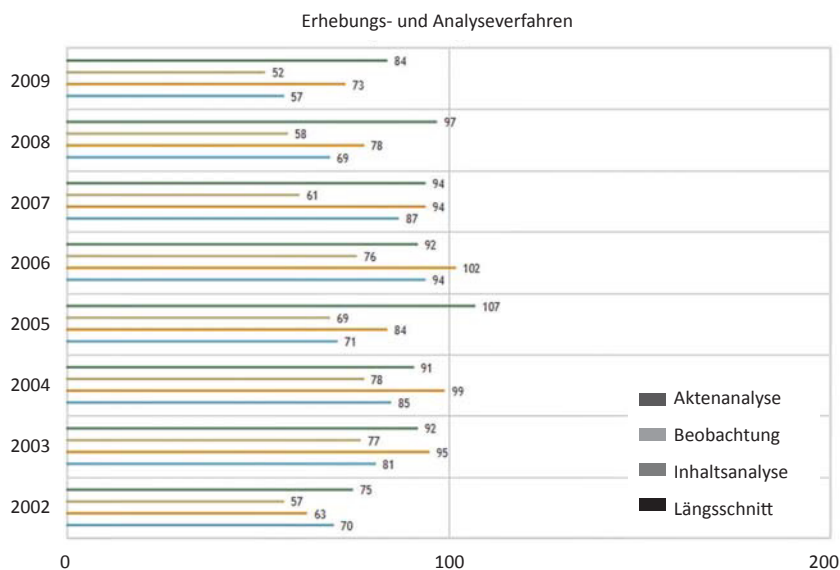


Abb. 12: Verteilung der Erhebungs- und Analyseverfahren über den Zeitraum 2002 bis 2009

3.7 Forschungsziel

Der Indikator „Forschungsziel“ beschreibt das inhaltliche Ziel eines Forschungsprojekts. Der Ausgangspunkt dieses Indikators ist das SOFISwiki-Feld „Methode“, aus dem das neue Attribut „Forschungsziel“ generiert wurde. Im Feld „Forschungsziel“ sind ausschließlich Werte aus der ersten Facette der Methodenklassifikation enthalten. Eine Besonderheit beim Indexierungsvorgang ist ein automatisches Upposting, das bei bestimmten Begriffen stattfindet: Bei der Methode *Evaluation* beispielsweise wird auch automatisch der Begriff *anwendungsorientiert* als Oberbegriff vergeben. Ebenso wird für die Begriffe *Methodenentwicklung* oder *Theoriebildung* immer der Oberbegriff *Grundlagenforschung* vergeben. Die Begriffe *Grundlagenforschung* und *anwendungsorientiert* als Oberbegriffe können hingegen allein vergeben werden. Dies führt zu Mehrfachbelegungen, die eine eindeutige Darstellung erschweren. Abbildung 13 visualisiert die Anzahl der Projekte mit den verschiedenen Forschungszielen über den Zeitraum 1997 bis 2003.

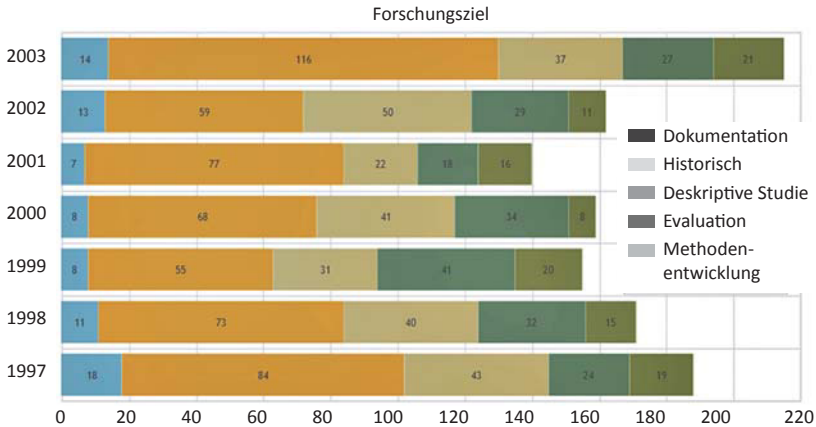


Abb. 13: Verteilung einiger Forschungsziele über den Zeitraum 1997 bis 2003

3.8 Forschungslandkarte

Die „Forschungslandkarte“ wurde über das SOFISwiki-Attribut „Forschungseinrichtung“ generiert und ermöglicht einen Überblick über Standorte einzelner Institutionen im deutschsprachigen Raum. Bei einem Klick auf den Standort (Abbildung 14) werden detaillierte Informationen über die Einrichtungen und ihre Projekte angezeigt (Abbildung 15).



Abb. 14: Landkarte der Forschungseinrichtungen

Abteilung Didaktik der Biologie (Göttingen)

Institutionsname: [Abteilung Didaktik der Biologie](#)

Fachbereich / Fakultät: [Biologische Fakultät \(Göttingen\)](#)

Universität: [Universität Göttingen \(Göttingen\)](#)

Institutions-Nr.: 029510

Inhaltliche Ausrichtung: [Erziehungswissenschaft](#)

Organisationstyp: [Universität](#)

Homepage: <http://www.zeus-bio.de/>

Straße: Waldweg 26

Ort: [Göttingen](#)

PLZ: 37073

Land: [Bundesrepublik Deutschland](#)

Kontaktperson: Sekretariat

Email: ubrand1@uni-goettingen.de

Telefon: 0511 39-3495

Projekte

Zur Zeit sind 9 Projekt(e) für diese Institution vorhanden.

0-9	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
...																											

- [Entwicklung eines Konzepts für naturbezogene Kindergärten, implementation und Evaluation](#)

Abb. 15: Metadaten-Ansicht der ausgewählten Forschungseinrichtung

4 Visualisierung von Indikatorenrelationen

Die bisherigen Visualisierungen stellen einzelne Indikatoren grafisch dar. Diese Indikatoren stehen jedoch in Wechselbeziehung zueinander. Eine grafische Darstellung dieser Beziehungen hilft dem Nutzer, Zusammenhänge zwischen einzelnen Indikatoren zu erschließen. Daher wurde im Projekt eine Anwendung entwickelt, welche Suchanfragen nach Indikatorenrelationen zu einer Datenbank unterstützt und die Ergebnismenge mit den entsprechenden Zusammenhängen visualisiert. Durch die visuelle Darstellung wird explizit dargestellt, welche Indikatoren wie oft gemeinsam auftreten.

Um Relationen zwischen den Ergebnismengen der Indikatoren grafisch darzustellen, gibt es mehrere Techniken. Eine Möglichkeit besteht darin, Indikatorenrelationen als Matrix abzubilden und die Intensität der Relationen farblich zu kennzeichnen. Diese Art von Visualisierung nennt man „Heatmap“. Eine Heatmap (Caraux & Pinloche, 2005) ist ein gutes Gestaltungsmittel, das zur Visualisierung zweidimensionaler Daten verwendet wird und sich vor allem dazu eignet, die Stärke und Häufigkeit des gemeinsamen Auftretens von Indikatoren auszudrücken. Die Farben geben dabei den Grad der Intensität der untersuchten Indikatoren an. Die Farbtöne ergeben sich anhand einer Skala, die eine numerische Variable (z. B. in Form einer Hexadezimalzahl) darstellt. Diese Art der Darstellung hat auch die Besonderheit, dass sie jeden Farbton zwischen zwei anderen annehmen kann. Somit können Wertunterschiede schnell erkannt und Informationen in einer Heatmap in jedem Punkt mit anderen Positionen verglichen werden, so dass ein guter Überblick über die Verteilung möglich ist.

Um die Anwendung zu realisieren, ist es zunächst notwendig zu wissen, wie die Daten, in diesem Fall Indikatoren und deren Inhalte, in der Datenbank abgespeichert werden. Da das Datenmodell der Datenbank SOFISwiki sich nicht strukturiert abfragen lässt, wird die Abfrage von Indikator-Beziehungen auf der Oracle-Datenbank durchgeführt, von welcher der Datenimport für SOFISwiki erfolgt. Die Visualisierung der Beziehungen wiederum wird innerhalb des oben beschriebenen Monitoring-Prototyps angezeigt. Dazu wurde eine PHP-Extension für den Prototyp entwickelt, welche die relevanten Daten anhand von SQL-Statements aus der Oracle-Datenbank abfragt, auswertet und als Heatmap darstellt.

Für die Gewichtung der Indikatoren-Relationen wurde der Jaccard-Index (Ni-wattanaku, Singthongchai, Naenudorn, & Wanapu, 2013) herangezogen, der die Ähnlich- bzw. Unterschiedlichkeit zweier Merkmale (z. B. A und B) anhand des Verhältnisses ihres gemeinsamen Auftretens zur Häufigkeit der einzelnen Merkmale berechnet:

Der Index wurde berechnet, um zu erfahren, wie oft die Merkmale zweier verschiedener Indikatorenmengen gemeinsam auftreten. Das Ergebnis liegt immer zwischen null und eins. Um Kommazahlen zu vermeiden, wurden die errechneten Zahlen aufgerundet und als Prozentzahl dargestellt. Auf der Basis dieser Ergebnisse wurden in der Folge die für die Farbcodierung als Parameter benötigten Minima und Maxima berechnet. Danach wurden die Werte mithilfe eines eigens entwickelten Algorithmus in das entsprechende Format konvertiert, um Farben zu vergeben. Da für die Auswahl effektiver Farbschemata für thematische Karten kein Standardverfahren existiert (Harrower & Brewer, 2003), wurde hier die übliche Farbgebung von rot bis blau für Heatmaps verwendet, das heißt, das Farbspektrum erstreckt sich von blau über grün und gelb bis rot, wobei die blaue Farbe auf eine schwache Beziehung hinweist, die rote hingegen auf eine starke. Je stärker der Zusammenhang zwischen zwei Indikatoren ist, desto farblich „wärmer“ wird der Bereich markiert. Durch die Farbtöne der Heatmap-Zellen, die jedes Indikatoren-Paar verbinden, wird ein Überblick über die Häufigkeit des gemeinsamen Auftretens der vorliegenden Indikatoren gegeben. Abbildung 18 zeigt exemplarisch den Zusammenhang zwischen Disziplinbereich und Erhebungs- und Analyseverfahren in Form einer Heatmap, Abbildung 19 eine Heatmap für die Relation zwischen Disziplinbereich und Forschungszielen.

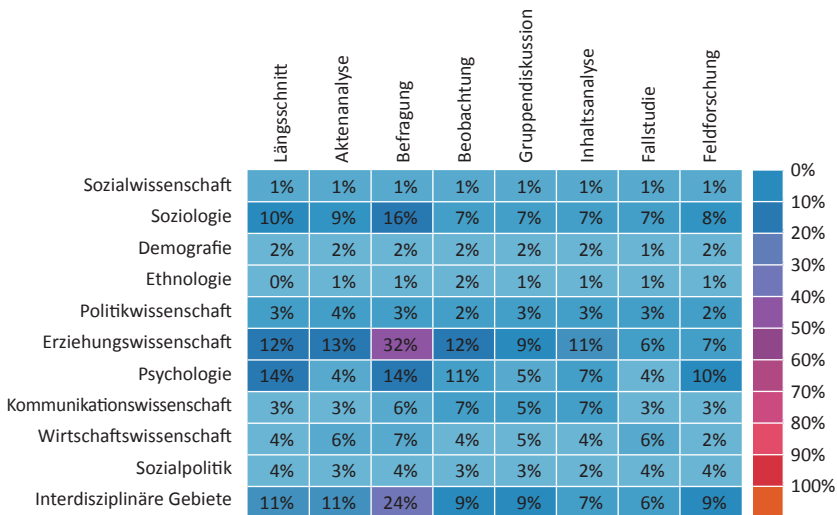


Abb. 16: Heatmap-Visualisierung der Relation zwischen Disziplinbereich und Erhebungsverfahren

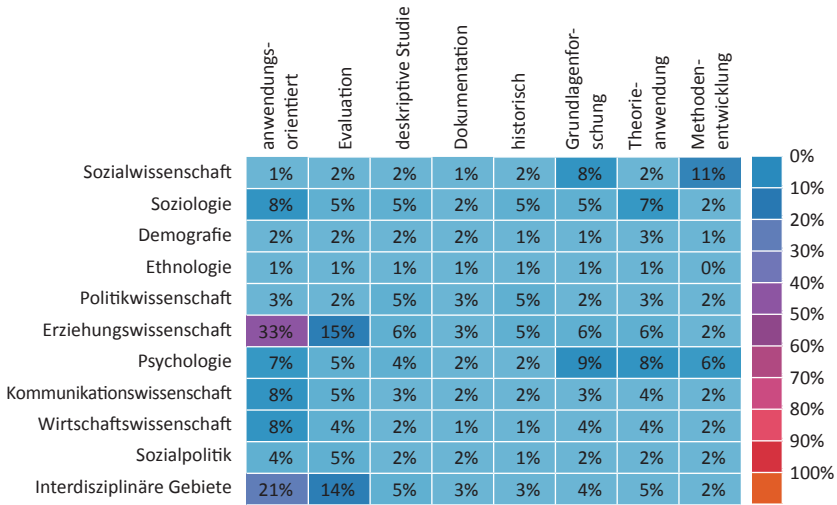


Abb. 17: Heatmap-Visualisierung der Relation zwischen Disziplinbereich und Forschungsziel

Über die reine Visualisierung hinaus bietet der Prototyp dem Nutzer die Möglichkeit, sich alle Forschungsprojekte mit einem bestimmten Merkmal anzeigen zu lassen. Beispielsweise werden beim Anwählen eines Disziplinbereichs die entsprechenden Unterdisziplinen angezeigt, welche ihrerseits ebenfalls adressierbar/zugänglich sind und zu den Forschungsprojekten führen, denen sie zugeordnet sind. Genauso kann man einzelne Methoden, Erhebungsverfahren oder Forschungsziele per Mausklick anwählen und bis zu den Forschungsprojekten gelangen, in denen die angewählte Methode eingesetzt wurde.

5 Schlussfolgerung und Diskussion

Die visuelle Darstellung soll den Informationsgehalt von Daten effektiv präsentieren, damit die Nutzer nicht nur einen schnellen, sondern auch einen möglichst zuverlässigen Zugang zum Informationsraum erhalten. Semantic MediaWiki ist eine leistungsfähige Software, um Datenauswertungen zu visualisieren. Das Projekt hat allerdings gezeigt, dass der Mehrwert von Indikatoren-Visualisierungen in hohem Maße von der Trennschärfe der Kategorienbildung abhängt. Dies gilt sowohl für die Ebene der Indikatoren selbst als auch, und noch viel mehr, für die Ebene der Wertebelegung aller zugrunde liegenden Variablen (wie Feldstruktur und Feldbelegung in den abgefragten Datenbanken). Die Darstellungsform muss für jeden einzelnen Indikator dabei so gewählt werden, dass der spezifische Informationsgehalt adäquat abgebildet wird. Um hier zu einer sicheren Einschätzung zu gelangen, wären die Entwicklung konkreter Anwendungsszenarien und systematische Evaluationen mit Nutzern erforderlich.

Ein grundsätzliches Problem des derzeitigen Monitoring-Prototyps ist der Umstand, dass es keine internationale, mit den Inhaltsbereichen und der Metadatenstruktur von SOFISwiki vergleichbare Datenbasis der Bildungsforschung gibt, sodass die Generierung von externen Vergleichswerten nicht möglich ist. Eine zentrale Zielsetzung für ein Folgeprojekt bestünde daher darin, signifikante Veränderungen, Abweichungen oder Anomalien von Indikatoren innerhalb des gegebenen Korpus im Zeitverlauf darzustellen und Abhängigkeiten aufzuzeigen. Um dies zu realisieren, müssten jedoch ganz neue Metriken entwickelt werden.

Literaturverzeichnis

Literaturquellen

- Caraux, G. & Pinloche, S. (2005). PermutMatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics*, 21 (7), 1280–1281.
- Haddou ou Moussa, K., Sondergeld, U., Mayr, P., Mutschke, P. & Rittberger, M. (2014). Assessing educational research: An information service for monitoring a heterogeneous research field. In *Libraries in the Digital Age (LIDA) Proceedings: Assessing libraries and library users and use* (LIDA Proceedings, Bd. 13, S. 1-8). Zadar: LIDA. Verfügbar unter: <http://ozk.unizd.hr/proceedings/index.php/lida/article/view/166/166>
- Harrower, M. & Brewer, C. (2003). ColorBrewer.org: An online tool for selecting colour schemes for maps. *Cartographic Journal*, 40 (1), 27–37.
- Krotzsch, M., Vrandečić, D., Volkel, M., Haller, H. & Studer, R. (2007). Semantic Wikipedia. *Journal of Web Semantics*, 5 (4), 251–261.
- Niwattanakul, S., Singthongchai, J., Naenudorn, E. & Wanapu, S. (2013). Using of Jaccard coefficient for keywords similarity. *Proceedings of the International MultiConference of Engineers and Computer Scientists, IMECS 2013, March 13–15, 2013, Hong Kong* (Bd. 1, S. 380384). Hong Kong: International Association of Engineers.

Internetquellen

- Gesis.org. *Klassifikation Sozialwissenschaften*. Abgerufen am 13. 11. 2014 von http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/KassifikationSozialwissenschaften_Stand_Juli_2013_dt_en__2_.pdf
- Gesis.org. *Methodenliste SOFIS-SOLIS*. Abgerufen am 13. 11. 2014 von http://www.gesis.org/fileadmin/upload/dienstleistung/tools_standards/Methodenliste_SOFIS-SOLIS.pdf
- Mediawiki.org. *MediaWiki*. Abgerufen am 13. 11. 2014 von <http://www.mediawiki.org/wiki/MediaWiki/de>
- Semantic-mediawiki.org. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:Semantic_templates
- Semantic-mediawiki.org. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:Result_formats
- Semantic-mediawiki.org. *D3 chart format*. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:D3_chart_format
- Semantic-mediawiki.org. *Editing*. Abgerufen am 13. 11. 2014 von <http://semantic-mediawiki.org/wiki/Help:Editing>
- Semantic-mediawiki.org. *Jqplotchart format*. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:Jqplotchart_format
- Semantic-mediawiki.org. *Jqplotseries format*. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:Jqplotseries_format
- Semantic-mediawiki.org. *Properties and types*. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:Properties_and_types
- Semantic-mediawiki.org. *Tagcloud format*. Abgerufen am 13. 11. 2014 von http://semantic-mediawiki.org/wiki/Help:Tagcloud_format
- Sofis.gesis.org. *SOFISwiki*. Abgerufen am 13. 11. 2014 von <http://sofis.gesis.org/sofiswiki/Hauptseite>

Danksagung

Unser Dank gilt den GESIS-Mitarbeitern *Simon Bachenberg* und *Sascha Schüller* für die Beschaffung der Daten sowie die Unterstützung bei der Entwicklung des Prototyps, *Julia Achenbach* für das Korrekturlesen sowie der ehemaligen DIPF-Mitarbeiterin *Ute Sondergeld* für die Projektkoordination.