

Das IUWIS-Datenmodell

Konstantin Baierer; Ben Kaden

Ziel des Reports

Interne und externe Inhalte werden bei IUWIS in einem so genannten *Informationspool* gesammelt, nach einem Datenmodell strukturiert, nach verschiedenen Kriterien inhaltlich erschlossen und relationiert sowie für den Diskurs, vor allem auf IUWIS, nutzbar gemacht. Dieser Report beschreibt die Grundeigenschaften des Datenmodells für IUWIS und die entsprechende Einbindung in die Plattform.

Verwandte Reports: Report „01/2010 Semantische Kontrolle“, Report „03/2010 Informationspool“

Anforderungen

Im IUWIS-Portal erhalten die Nutzer des Angebots die Möglichkeit, kollaborativ in geschlossenen oder offenen Gruppen einzelne Themenbereiche zu bearbeiten und zu diskutieren.

Zusätzlich sollen neben den so entstehenden internen Diskursrepräsentationen auch themenrelevante, z.B. a) externe Informationsobjekte wie Publikationen, oder b) Ereignisse wie Konferenzen, Gerichtsentscheidungen u.ä., in einer Datenbank verwaltet werden. Diese in einem (intern) so genannten Informations-Pool bzw. Info-Pool verwalteten Objekte bilden somit die Basis der Dossiers (sh. Dossier-Report). Alle Informationsobjekte werden zweischichtig – also sowohl über ein kontrolliertes Vokabular wie auch über freie Tags – mit der IUWIS-Tagsonomy erschlossen und vielfältig nutzbar gemacht.

Elemente des Datenmodells

Da sich bei der Konzeption der beiden Funktionsbereiche der Plattform – dem Bereich Information und dem Bereich Kommunikation – erhebliche Redundanzen in den Eigenschaften der jeweiligen Elemente ergeben haben, werden alle Objekte als Konkretisierung eines abstrakten Infopool-Objekts angesehen und

verarbeitet. Dieser Ansatz erlaubt es, durch Vererbungsbeziehungen eine umfängliche Modellierung mit nur wenigen definierten, generellen Attributen umzusetzen. Diese werden von den jeweiligen Objektarten realisiert, d. h. mit Bedeutung versehen.

Infopool-Objekt

Alle Informations-Objekte der IUWIS-Plattform können mit klassischen Metadaten ausgezeichnet werden:

- Titel
- Untertitel
- Titeltzusätze
- Abstract
- Datum

Darüber hinaus können jedem Objekt assoziierte Themen, Akteure sowie Orte zugewiesen werden, die wiederum Elemente des jeweiligen Vokabulars der Tagsonomy sind. Der Assoziierte Akteur wird bei einer Monographie bspw. zum Autor, bei einem Sammelband zum Herausgeber, bei einer Konferenz zum Veranstalter oder bei einem Benutzerprofil zum Namen des Benutzers.

Um zu verhindern, dass Objekte gar nicht oder bis zur Beliebigkeit übermäßig ausgezeichnet werden, muss für die Bereiche Akteur und Thema ein einzelnes Tag, im Idealfall das Hauptthema bzw. der Hauptakteur, als Haupt-Tag dieser Facette für dieses Objekt definiert werden. Mit diesen Daten lassen sich zudem Untersuchungen über die Relevanz bestimmter Tags umsetzen bzw. können Relevanzfilter entwickelt werden (vgl. den Report zur semantischen Kontrolle).

Die Diskursrelationen sind ebenfalls auf einer abstrakten Ebene definiert, wodurch grundsätzlich jedes Objekt des Infopools mit jedem anderen in Relation gesetzt werden kann. Bisher ist diese Art der Verknüpfung zwar nur für die Publikationen und Diskussionsbeiträge vorgesehen. Sie könnten aber in Zukunft auch zwischen Akteuren, Themen, Ereignissen verwendet werden.

Jedes Objekt des IUWIS-Infopool erhält eine interne ID, die der Node-ID in Drupal entspricht. Auch externe IDs, wie URL, ISBN, DOI, ISSN, PND-Nummer

oder Kurztitel ähnlich dem BibTeX-Formatsollten können für alle Objekte vergeben werden.

Ereignis- und Publikationsdatenbank

Publikationsdatenbank

In der Publikationsdatenbank sollen externe Quellen zum Themenbereich Urheberrecht in Wissenschaft und Bildung gesammelt, erschlossen und auffindbar gemacht werden. Neben klassischen Literaturquellen (Zeitschriften, Sammelbände, Monographien) sind dies insbesondere juristische Werke (Gesetze, Gesetzeskommentare, Urteile, Debatten), aber auch mehr oder weniger informelle Quellen, wie Blog-Postings, Pressemitteilungen oder Arbeitspapiere. Die Relevanz für die Aufnahme einer Quelle ergibt sich dabei aus der Nützlichkeit für den Diskurs. Eine Wertung bei der Auswahl findet nicht statt.

Da bei IUWIS eine möglichst granulare Erschließung der Werke angestrebt wird, gleichzeitig aber eine Vielzahl von Werkarten zugelassen werden sollen, haben wir uns auch hier für ein abstraktes System der Vererbung entschieden: Dies erlaubt ein weitgehend reduziertes Modell für die Publikationen mit einem simplen Basisobjekt, von dem die konkreten Ausprägungen bestimmte Eigenschaften erben. Das Modell enthält Felder für allgemeine bibliographische Angaben, die wiederum die Attribute des Basisobjekts des Infopools realisieren. So wird beispielsweise der Autor automatisch zum Assoziierten Akteur. Oder es können extern einem Objekt zugeordnete Tags als *Freie Tags* in die Tagsonomy übernommen werden.

Entitäten der Publikationsdatenbank (=Informationsobjekte), die in anderen Entitäten enthalten sind, können mit diesen optional über eine *isPartOf-Relation* verbunden werden. Dadurch werden Redundanzen weitgehend vermieden. Es ist nach diesem Ansatz möglich, ganze Ketten von Zusammenhängen im Publikationsprozess und alle Teile einer Publikation, die für den Benutzer relevant sein könnten einzeln zu diskutieren, zu *taggen* und als Einstiegspunkt für weitere Recherchen oder Grundlage für neue Dossiers zu nutzen.

In erster Linie mit Blick auf das deutsche Urheberrecht und verwandte Rechtsgebiete sollen Gesetze möglichst präzise beschrieben werden. Gesetzesänderungen, wie sie u.a. beim Zweiten Korb anfielen, verändern ja nicht das Gesetz als Ganzes, sondern greifen auf der Ebene bestimmter Paragraphen. Oft werden auch nur die Formulierungen einzelner Absätze oder Sätze modifiziert. Um dem gerecht zu werden, müssen Sätze nicht nur mit Paragraphen und Paragraphen mit Gesetzen verbunden sein, sondern auch mit den jeweiligen Gesetzesänderungen. Auch dies wird über *isPartOf*-Beziehungen realisiert.

Die Kapitel einer Monographie, eines Sammel- oder Konferenzbande sowie Zeitschriften- oder Zeitungsartikel werden als ein gesonderter Objekttyp angesehen. Fasst man diesen als Kapitel auf, so sind sie den Werken direkt untergeordnet, versteht man sie aber als Artikel, so können sie optional noch einer einzelnen Ausgabe der jeweiligen Zeitschrift oder Zeitung zugewiesen werden.

Ereignisdatenbank

In der Ereignisdatenbank werden Konferenzen, Gerichtsprozesse, Veranstaltungen aller Art, sowie Aktionstage wie z.B. die Open Access Week verzeichnet. Über eine nicht-hierarchische *hasRelated*-Verknüpfung können darüber hinaus Ereignisse mit Publikationen verbunden werden. Beispiele dafür sind Konferenzen mit einem Konferenzband sowie akademische Ereignisse mit zugehörigen Hochschul- oder Festschriften.

Zusätzlich zu diesen expliziten Ereignistypen ist auch die Möglichkeit *impliziter Ereignisse* vorgesehen. Diese werden automatisch aus den Informationen der Publikationsdatenbank generiert, beispielsweise aus dem Erscheinungsdatum einer Publikation oder dem Gültigkeitszeitraum einer Gesetzesfassung. Über die impliziten Ereignisse können diskurstheoretische Auswertungen vorgenommen werden. So lässt sich auf dieser Basis die Entwicklung des Diskursgeschehens zu einem konkreten Thema sehr präzise modellieren. Als Mehrwert dieser erhöhten Transparenz wird u.a. eine argumentative Straffung des Diskussionsgeschehens in IUWIS mit exakten Referenzierungsmöglichkeiten gesehen.

Interaktiver Diskursbereich

Der interaktive Diskursbereich deckt alle den Benutzern möglichen Aktivitäten auf der Plattform ab.

Dazu zählen u.a.:

- die Nutzerprofile
- die nutzergenerierten Inhalte
- Dossierdiskussionen
- Informationsobjekte
- Annotationen der Informationsobjekte
- Freie *Tags* (Personomien)
- Verknüpfungen mit anderen Nutzern
- Individuelle Kalender-/Ereignisangaben

Benutzer

Jeder Benutzer hat eine Profilseite, das er mit wenigen obligatorischen und weiteren optionalen Angaben zur Person füllen kann. Obligatorisch ist eine zeitstabile individuelle Identifizierung. Idealerweise ist dies der Klarname, einerseits, da es sich um einen Fachdiskurs handelt und andererseits um negativen Effekten der Online-Kommunikation (*Trollen, Flaming*) vorzubeugen

Weiterhin können erfasst werden:

- institutionelle Zugehörigkeit
- Kontaktdaten
- Interessen

Auch die Personomie des Nutzers ist Bestandteil seines Profils („meine Tags“). Sie ergibt sich aus den den eigenen Beiträgen (Diskussionsbeiträgen oder Informationsobjekten) zugewiesenen freien Tags.

Zusätzlich können Benutzer Publikationen, die von ihnen (extern) verfasst sind, in den IUWIS-Informationspool einbringen. Sie können dort bereits vorhandene Titelangaben mit weiteren Metadaten versehen und explizit mit ihrem Profil verknüpfen, um sie dort als eine Art Bibliografie auszugeben. Im Idealfall führt so das Streben der Benutzer nach Sichtbarkeit ihrer Veröffentlichungen zu einer

kontinuierlich gepflegten, gut erschlossenen und umfangreichen Publikationsdatenbank zum erweiterten Bereich der Urheberrechtsproblematik.

Schließlich soll es für die Benutzer möglich sein, sich mit anderen Nutzern direkt über entsprechende elektronische Kommunikationsformen zu vernetzen.

IUWIS-Textinhalte

Alle IUWIS-internen Beiträge sind werden mit Zeitstempel und der IUWIS-Benutzer-ID des Verfassers gespeichert. Darüber hinaus können optional alle allgemeinen Attribute eines Infopool-Objekts verwendet werden. Insbesondere sollen die Nutzer bei der Verwendung von Themen- und Akteurtags animiert werden, den Datenkorpus für die semantische Kontrolle zu erweitern.

Innerhalb der Text-Objekte wird unterschieden zwischen Featuretexten, Blogbeiträgen und Diskussionsbeiträgen. Jeder Text kann sich auf ein anderes Infopool-Objekt in einer diskursiven oder formalen Relation beziehen.

Feature-Text

Hierbei handelt es sich um Texte, die ein bestimmtes Thema näher erläutern und bevorzugt von ausgewiesenen Experten des Fachgebiets, in jedem Fall aber in einem redaktionellen Kontext verfasst werden. Sie können nicht direkt kommentiert, aber dadurch diskutiert werden, dass Diskussionsbeiträge auf sie Bezug nehmen. Feature-Texte werden in Dossiers als einleitende Problematisierungen, historische Überblicke, juristische Analysen o. Ä. auf der Hauptseite des jeweiligen Dossiers als besonders lesenswert hervorgehoben.

Diskussionsbeitrag

Diskussionen beginnen in IUWIS immer mit einem Initialbeitrag mit einem aussagekräftigen, das Thema bestimmenden Titel. Der Beitrag kann sich optional auf ein Objekt des Infopools – etwa einen Feature-Text oder ein bestimmtes Thementag – beziehen. Rekursiv können sich nun auf diesen Initialbeitrag weitere Diskussionsbeiträge beziehen. Sie bilden dadurch einen Baum von Beiträgen (*Diskussionsthread*). Der Bezug der Diskussionsbeiträge untereinander kann durch Angabe diskursiver Qualitäten semantisch ausgezeichnet werden.

Alle *Diskussionsthreads*, deren Initialtext sich auf ein bestimmtes Objekt des Infopools bezieht, sollten mit diesem verknüpft und in der Ansicht als Diskussionen zu diesem *Text/dieser Publikation/diesem Thema/diesem Dossier* gelistet werden.

Blögeintrag

Die IUWIS-Redaktion führt ein Blog, in dem organisatorische und administrative Neuigkeiten verbreitet werden. Strukturell sind Blögeinträge identisch mit Initialtexten für Diskussionen, erscheinen aber in einem dossierunabhängigen Kontext.

Dossier

(siehe auch Report Dossier)

Im Dossier werden die interaktiven Objekte zu einem abgegrenzten Themenbereich zusammengefasst. Jedes Dossier hat neben den allgemeinen Infopool-Objekt-Attributen immer mindestens einen Moderator. Falls *Öffentlich-Attribute* gesetzt sind, können sich alle IUWIS-Benutzer als Mitglied des Dossiers eintragen, andernfalls kann der Moderator über die Mitgliedschaft anderer Benutzer entscheiden.

Mit dem Dossier können Diskussionen, Featuretexte, Ereignisse, Publikationen oder Kalendereinträge verbunden werden. Die semantische Verknüpfung der Elemente erfolgt über die Tagsonomy. Indem bspw. das Tag *Dossier Creative Commons* einer Publikation vergeben wird, wird die Publikation in diesem Dossier aufgeführt.

Kalender

Kalendereinträge sind Textinhalte mit einem Bezugsdatum. Diese werden in einem Kalender visualisiert.

Tagsonomy

Die Tagsonomy (von Tags und Taxonomy) besteht aus fünf Facetten, also voneinander abgegrenzten Wortschätzen, davon vier kontrollierten (Thema,

Akteur, Ort, Diskursrelation) und einer freien Facette, die von allen Benutzern nach eigener Maßgabe ergänzt werden kann.

Aus technischer Sicht handelt es sich bei diesen Wortschätzen um im Content-Management-System *Drupal* definierte Vokabularien. Die kontrollierten Vokabularien werden redaktionell erarbeitet, das freie Vokabular wird regelmäßig daraufhin überprüft, ob häufig verwendete Idiome in die kontrollierten Wortschätze übernommen werden sollen. (vgl. Report zur Semantischen Kontrolle)

Eine weitere Besonderheit des IUWIS-Ansatzes zur Verschlagwortung mit *Tags* besteht darin, dass Tags zum einen als Metadaten fungieren, wie dies bei Anwendungen wie *Delicious* oder *flickr* geschieht, zum andern aber auch selbst wieder Objekte des Info-Pools sind. Einem Tag können somit – analog zu den Informationsobjekten des Info-Pools – weitgehend dieselben Metadaten zugewiesen werden. Das Tag fungiert damit als ein Einstiegspunkt, um im Info-Pool zu recherchieren und ergänzt die unkomplizierten Möglichkeiten des Taggings um eine zusätzliche semantische Ebene, die konventionelle Anwendungen nicht bieten können. Über ein Definitionsfeld dient die *Tagsonomy* in gewisser Weise als Glossar: Der Aufruf der Metadatenseite zu einem Thementag zeigt gleichzeitig eine dort hinterlegte Definition sowie mögliche hinterlegte Relationen innerhalb der Tagsonomy an.

Tags als Metadaten

Das umfangreichste Vokabular ist der Bereich *Thema*, der sich aus einer Ontologie zum Themenbereich *Urheberrecht in Wissenschaft und Bildung* speist. Die Vokabularien zu Ort und Akteur werden kontrolliert, um eindeutige Ansetzungsformen sicherzustellen und Querverbindungen zu IUWIS-internen Ressourcen wie Benutzern oder Dossiers herzustellen. Zudem sollen sie in einem späteren Entwicklungsschritt als Basis für Visualisierungen und verschiedene *Browsing-Zugänge* (Zeitleisten, Themenmaps) zu den erfassten Objekten verwendet werden. Diese beiden Vokabularien sind weitgehend normiert.

Das Vokabular *Diskursrelation* bietet die Möglichkeit, einzelne Objekte des Infopools qualitativ-semantic zueinander in Verbindung zu setzen. Es sollte nur wenige Beziehungen wie *zitiert, kritisiert, widerlegt* o. Ä. enthalten.

Die freien Tags schließlich entsprechen dem herkömmlichen Ansatz von *Folksonomies*, bei dem Benutzer Objekten Zeichenfolgen explizit zuweisen. Dabei bestehen die gleichen Probleme wie bei anderen Taggingansätzen auch (beispielsweise Unklarheit, auf welcher semantischen Ebene getaggt wird; verschiedene Ansetzungsformen; uneinheitliche Interpunktion usw.). Das nehmen wir in Kauf, um dem Benutzer die Möglichkeit zu geben, eine Verschlagwortung nach eigenen Maßstäben vorzunehmen (*Personomy*), andererseits auch die Synergieeffekte der Menge verschlagwortender Benutzer zu nutzen (*Folksonomy*). So kann neu aufkommende Themen oder Akteure Rechnung getragen werden oder Lücken im ontologischen Konzept aufgefüllt werden. Der Unschärfe in den Ansetzungsformen soll über die Elaboration von (syntaktisch basierten) Äquivalenzklassen begegnet werden.

Tags als Infopool-Objekte

Tags als Objekte des *Infopools* verhalten sich ähnlich wie *Dossiers*. Man kann sie auch als Mini-Dossiers auffassen. Anders als in den *Dossiers* erfolgt die Zuweisung von Infopool-Objekten automatisch anhand des entsprechenden Tags.

Die Tag-Objekte können untereinander über ein Äquivalenzattribut als gleichbedeutend gekennzeichnet werden. Aus einer Klasse von äquivalenten Tag-Objekten wird für alle Funktionalitäten, bei denen eine einheitliche Ansetzungsform gewünscht ist (z.B. *Tag Clouds, Autocompletion*), nur das Tag angezeigt, bei dem das Hauptansetzung-Attribut gesetzt ist. (siehe auch Report „Semantische Kontrolle“, letzter Abschnitt).