# KONZEPT FÜR EINEN SEMANTIC-WEB-BASIERTEN STUDIENFÜHRER FÜR WIRTSCHAFTSINFORMATIK AN FACHHOCHSCHULEN IM DEUTSCHSPRACHIGEN RAUM

Vera G. Meister
Fachbereich Wirtschaft
Fachhochschule Brandenburg
Magdeburger Str. 50
14770 Brandenburg, Deutschland
E-Mail: vera.meister@fh-brandenburg.de

### **KEYWORDS**

Studiengang Wirtschaftsinformatik, Tätigkeitsprofile für Informatiker, Studiengangschwerpunkt, semantische Webseitenannotation, Domänen-Ontologie, Microdata.

#### **ABSTRACT**

Der Beitrag beschreibt die ersten Phasen eines Entwicklungs- und Einführungsprojektes für ein Semantic-Webbasiertes Informationssystem. Es soll Stakeholder von Wirtschaftsinformatik-Studiengängen an Fachhochschulen im deutschsprachigen Raum in relevanten Entscheidungsfällen unterstützen. Grundlage dafür soll eine Anreicherung der über Studiengangswebseiten bereitgestellten Informationen sein. Dafür wurde eine Ontologie entwickelt, die zugleich domänenspezifisch wie auch standardbasiert ist. Ausgehend von einer Beschreibung der Wissensdomäne, erörtert der Beitrag verschiedene Lösungsansätze und entwirft für den präferierten Lösungsansatz ein Systemdesign. Die grundsätzliche technische Umsetzbarkeit wurde durch einen Proof of Concept nachgewiesen. Der Beitrag endet mit einem kurzen Ausblick auf die weiterführende Arbeit, die sowohl Forschungs- und Entwicklungs- als auch organisationale und Managementaufgaben umfasst.

### EINFÜHRUNG

Als Dachverband der Studiengänge für Wirtschaftsinformatik an Fachhochschulen im deutschsprachigen Raum versteht sich der Arbeitskreis Wirtschaftsinformatik (AKWI) als fachkompetenter Ansprechpartner für alle Stakeholder der vertretenen Studiengänge. Primärere und niedrigschwellig zugänglicher Informationskanal für diese Interessengruppen ist die Webseite des Dachverbandes (AKWI 2015). Ein Angebot für Studieninteressierte auf der Webseite ist beispielsweise der Hochschulkompass der Hochschulrektorenkonferenz (HRK 2015b), der mit einem speziellen Filter auf Wirtschaftsinformatik versehen ist.

Im Jahr 2013 entstand die Idee, einen eigenen Studienführer für Wirtschaftsinformatik an Fachhochschulen zu entwickeln. Dafür wurde eine Arbeitsgruppe unter Lei-

tung der Autorin gebildet, die zunächst Bedarfe und Konzeptansätze sammelte. Auf Initiative der Autorin wurde ein Semantic-Web-basierter Ansatz präferiert und über mehrere Semester in studentischen Forschungs- und Entwicklungsprojekten an der Fachhochschule Brandenburg ausgearbeitet. Als Benchmark für die Informationstiefe des zu entwickelnden Systems wurden von Beginn an die Schwerpunkten der Studiengänge sowie Aspekte der Qualifizierung für bestimmte Tätigkeitsfelder für IT-Fachleute gewählt.

# **BESCHREIBUNG DER DOMÄNE**

Im Jahr 2003 aktualisierte der Fachbereich Wirtschaftsinformatik der Gesellschaft für Informatik e. V. (GI) seine Rahmenempfehlung für die Universitätsausbildung in Wirtschaftsinformatik (WI), in der er unverändert die drei fachlichen Säulen: (1) Betriebswirtschaftslehre, (2) Informatik und (3) Wirtschaftsinformatik als annähernd gleichgewichtig für WI-Studiengänge erklärt. Zusätzlich empfehlen die Autoren der Rahmenempfehlung eine vierte Säule mit Grundlagen aus Mathematik, Statistik, Recht und Verhaltenswissenschaften, die vom Umfang vergleichbar ist (GI 2003). Auch wenn sich alle akkreditierten WI-Studiengänge an Fachhochschulen im deutschsprachigen Raum an dieser Empfehlung orientieren, gibt es doch nennenswerte Variationen an verschiedenen Hochschulen. Diese lassen sich teilweise durch die organisationale Einbindung der Studiengänge in Fachbereiche oder Fakultäten erklären. Es liegt nahe, dass ein WI-Studiengang, der an einem Fachbereich für Informatik angeboten wird, der zweiten Säule mehr Gewicht beimisst, während eine organisationale Einbindung in einen Fachbereich Wirtschaft eher für eine Stärkung der Säulen (1) oder (3) spricht. Ein klares Bild über die fachlichen Schwerpunkte eines WI-Studiengangs lässt sich durch eine detaillierte Analyse des Curriculums erzielen, wie sie im Rahmen von Akkreditierungsprozessen vorgenommen oder gar gefordert wird.

Ein anderer Aspekt zur Beurteilung von Studiengängen, insbesondere an Fachhochschulen, ist ihre Ausrichtung im Hinblick auf bestimmte berufliche Qualifikationen. In

(Beckmann e. a. 2015) analysieren Beckmann, Lindner und Werner die aktuelle Situation auf dem deutschen Arbeitsmarkt für IT-Fachleute. Sie nehmen dabei auf eine Klassifikation von Tätigkeitsfeldern Bezug, die im Jahr 2010 von der Agentur für Arbeit aktualisiert wurde. Danach sind für WI-Hochschulabsolventen die in Tabelle 1 dargestellten Tätigkeitsfelder einschlägig.

Tabelle 1: Tätigkeitsfelder für IT-Fachleute auf dem deutschen Arbeitsmarkt (Beckmann e. a. 2015)

Tätigkeitsfeld	Erforderliche Kompetenzen
Informatik	Hard- und Softwarelösungen oder
	komplexe IT-Systeme konzipie-
	ren, erstellen, installieren,
	betreuen oder erforschen
Beratung	IT-Systeme analysieren, Anwend-
	er beraten, IT-Produkte vertreiben
Administration	IT-Netzwerke einrichten und
	betreuen; IT von Unternehmen
	und Organisationen koordinieren
	und organisieren; IT-Systeme und
	Web-Anwendungen administrie-
	ren; Datenbanken einrichten und
	administrieren
Software-	Software konzipieren, entwickeln
Entwicklung	und programmieren
IT Management	IT-Projekte, IT-Abteilungen, IT-
	Arbeitsgruppen führen; IT-Strate-
	gien entwickeln und umsetzen; IT-
	Governance verantworten

Bild 1 zeigt die Passung zwischen den drei fachlichen Säulen der WI und den fünf Tätigkeitsfeldern. Diese Zuordnung folgt der in (Großmann e. a. 2005) beschriebenen Methode der Analyse von Arbeitsaufgaben, angewandt auf die Tätigkeitsfelder für IT-Fachleute. Den Bezugsrahmen bilden die typischen Lernziele der Studiengangsmodule, die den Säulen der WI zuordenbar sind.

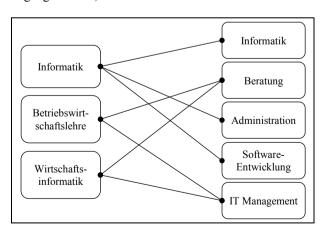


Bild 1: Säulen und Tätigkeitsfelder der WI

Dieses komplexe Bild bedenkend, erscheint die folgende Fragestellung als gültig und untersuchenswert: In welchem Maße qualifiziert ein bestimmter WI-Studiengang für oben genannten Tätigkeitsfelder für IT-Fachleute? Im Folgenden wird diese Fragestellung ins Zentrum des zu entwickelnden Studienführers für WI gestellt.

### **PROBLEMDEFINITION**

Die Stakeholder von Studiengängen sind sowohl innerhalb der Hochschulen (Studiengangsentwickler, Marketingverantwortliche, Hochschulleitung) als auch außerhalb zu finden (Studieninteressierte, Arbeitgeber, Projektpartner). Zunächst ging es darum, für alle Gruppen von Interessenten relevante Entscheidungsfälle mit Bezug zu fachlichen Schwerpunkten und/oder Tätigkeitsfeldern für WI-Absolventen zu sammeln. Ausgehend von eher allgemeinen Entscheidungskriterien, soll im Folgenden eine Auswahl dieser speziellen Entscheidungsfälle dargestellt werden.

Zunächst soll die Entscheidung von Studieninteressierten für einen Studiengang in den Blick genommen werden. Ihre Entscheidungskriterien können tatsächlich sehr stark voneinander abweichen. Vielfach stehen geographische und organisationale Kriterien, wie Ort, Rechtsform und Image der Einrichtung an erster Stelle. Diese Aspekte sollen als allgemeine Kriterien bezeichnet werden. Insbesondere Studieninteressierte mit erstem berufsqualifizierendem, akademischem Abschluss oder mit beruflicher Vorbildung und/oder Erfahrung können bereits klare Karriereziele bzw. konkretere Schwerpunktpräferenzen aufweisen. Für diese Gruppe gewinnen fachspezifische Entscheidungskriterien an Relevanz.

Ein starkes Interesse an Informationen über spezifische Qualifikationen im Hinblick auf Studiengänge kann bei Unternehmen und Organisationen beobachtet werden, die auf der Suche nach qualifizierten Fachkräften sind. Darüber hinaus können diese Stakeholder an einer Projektpartnerschaft mit der Hochschule interessiert sein und deshalb einen tieferen Einblick in fachliche Schwerpunkte bestimmter Studiengänge wünschen. Nicht zuletzt sind auch die Hochschulen selbst daran interessiert, ihre Angebote erkennbar zu differenzieren, um von Studieninteressierten und anderen Interessengruppen gezielt gefunden zu werden. Eine formale Spezifikation der Studienschwerpunkte und/oder der Qualifikationsoptionen für bestimmte Tätigkeitsfelder durch einen Studiengang kann dieses Bemühen wirksam unterstützen.

Bezüglich der allgemeinen Entscheidungskriterien für Studiengänge ist die Informationsbereitstellung in Deutschland auf einem vergleichsweise hohen Niveau. Hier zeichnet die Hochschulrektorenkonferenz verantwortlich. Ihr Webangebot "Hochschulkompass" richtet sich primär an Studieninteressierte (HRK 2015a). Es erlaubt eine erweiterte Suche nach Studiengängen entlang der folgenden Filterkriterien:

- Studientyp (grundständig oder weiterführend);
- formale Aspekte eines Studiengangs, wie Hauptunterrichtssprache, Studienbeginn, Akkreditierung;

- Studienformen, wie Vollzeit- oder Teilzeitstudium, ausbildungsintegrierend, berufsbegleitend etc.;
- geografische Merkmale, wie Studienort, Postleitzahl und Bundesland:
- Zugangsvoraussetzungen für Personen ohne Abitur, wie z. B. Meisterprüfung, mehrjährige Berufsausbildung oder Berufserfahrung;
- institutionelle Merkmale der Hochschule, wie Hochschultyp und Trägerschaft;
- fachliche Suche durch Eingabe eines Suchbegriffs oder durch Spezifikation vorgegebener Kategorien von Fächergruppen, Studienbereichen und Studienfeldern.

Auch wenn, wie im letzten Punkt benannt, eine (grobe) fachliche Suche durch den Hochschulkompass unterstützt wird, so ist doch eine detaillierte fachliche Suche mit diesem System nicht möglich. Wirtschaftsinformatik kann zwar als Studienfeld ausgewählt und auch mit Erfolg als Suchterm eingesetzt werden, spezielle Entscheidungskriterien, wie oben beschrieben, werden jedoch verständlicherweise nicht unterstützt. Ein Interessent könnte diese Informationen ggf. den Studiengangswebseiten entnehmen, allerdings hat er mit einer Reihe von Schwierigkeiten zu kämpfen. So werden Informationen dieser Art oftmals nicht explizit dargestellt oder es werden heterogene Begriffe und Ausdrücke verwendet. In den seltensten Fällen werden diese Informationen in maschinenlesbarer Form vorgehalten. Das ist auch der Grund dafür, dass sich eine automatische Analyse dieser Kriterien als schwierig bis unmöglich erweist. Das zu lösende Problem besteht also darin, ein Informationssystem zu entwerfen, zu entwickeln und zu implementieren, das die beschriebenen Entscheidungsfälle unterstützt. Gleichzeitig muss die Wartung eines solchen Systems so gestaltet sein, dass keine oder nahezu keine zusätzliche redaktionelle Arbeit anfällt und auch die Pflege der Daten ohne Wartungsaufwand an der Peripherie auskommt.

# LÖSUNGSANSÄTZE

Mittel der Wahl zur Gewährleistung der Maschinenlesbarkeit von im Web bereitgestellten Informationen sind semantischen Technologien (Allemang und Hendler 2011). Voraussetzung dafür wiederum ist die Verfügbarkeit einer problemspezifischen Ontologie, die zugleich weitestgehend standardkonform sein sollte (Suarez-Figueroa 2013). Im Wahlpflichtmodul "Grundlagen semantischer Technologien" im Sommersemester 2014 wurden erste Architekturansätze für ein solches System gesammelt und im September 2014 auf der 27. Jahrestagung der AKWI an der OTH Regensburg vorgestellt.

# Beschreibung alternativer Ansätze

Erster, naheliegender Gedanke war, auch in Zukunft auf den Daten des zentralen HRK-Systems aufzusetzen. Die Daten müssten dann in RDF transformiert werden (Resource Description Framework ist eine zentrale Basistechnologie im Technologie-Stack des Semantic Web). Die transformierten Daten wären dann um semantisch ausgezeichnete Datenelemente anzureichern, welche Informationen zu fachlichen Schwerpunkten und Tätigkeitsfeldern im Rahmen der Domänenontologie bereitstellen. Um aktuell zu bleiben, müssten diese ergänzenden Daten regelmäßig neu gesammelt werden. Die aggregierten Daten könnten dann über eine Suchanwendung auf Basis von SPARQL - der Analysesprache des Semantic Web – abgefragt werden. Bei diesem Ansatz wäre der Entwicklungsaufwand für die Datenbank vergleichsweise gering, allerdings wären dauerhaft große organisatorische und redaktionelle Anstrengungen nötig, um das System aktuell zu halten. Einzig die Transformation der existierenden Datenbank könnte automatisiert werden, solange das Datenbankschema unverändert bleibt. Darauf hat der AKWI jedoch keinen Einfluss.

Die zweite Idee kann als Graswurzel- oder inkrementeller Ansatz bezeichnet werden. Interessierte Fachhochschulen mit WI-Studiengängen annotieren ihre Studiengangswebseiten gemäß der Domänenontologie. Ein automatisierter Webdienst parst diese Seiten und speichert die extrahierten Daten in Form von RDF-Triples in einer semantischen Datenbank. Die Informationsabfrage kann analog zum ersten Ansatz erfolgen. Ein solches System könnte weitestgehend automatisiert betrieben werden, sofern die Annotationsaufgabe in einer Weise unterstützt wird, dass im Vergleich zur allfälligen redaktionellen Arbeit kein Zusatzaufwand anfällt. Es entstünde also ein initialer Entwicklungsaufwand, aber keine zusätzlichen Pflegeaufwände in der Zukunft. Der Fakt, dass zumindest zu Beginn nicht alle Studiengänge für WI von diesem System erfasst wären, kann als ein Nachteil dieses Ansatzes angesehen werden. Andererseits kann so nach der Methode des Rapid Prototyping verfahren werden, was schnelle Entwicklungs- und Feedbackzyklen erlaubt (Balzert 2008). Studiengänge, die sich erst später dem System anschließen, könnten von den bislang gesammelten Erfahrungen profitieren und damit den internen Einführungsaufwand minimieren.

Die dritte und letzte Idee folgt einem Big-Data-Ansatz. Er setzt auf einen Web-Crawler, der sämtliche einschlägige Webseiten und Dokumente durchforstet. Unter Einsatz einer Komponente für Natural Language Processing könnten die großen Mengen der gecrawlten Daten analysiert und den Konzepten der Domänenontologie zugeordnet werden. Speicherung und Abfrage folgen den bisher beschriebenen Modellen. Dieser Ansatz kann als nahezu vollständig automatisiert gekennzeichnet werden. Es wären keinerlei Abstimmungen mit Verantwortlichen der Hochschulen notwendig. Allerdings gibt es hier zwei bemerkenswerte Nachteile: zum einen kann die Richtigkeit der Informationen nur mit einer gewissen Wahrscheinlichkeit gesichert werden und zum anderen würde ein solches System wesentlich mehr technische Ressourcen benötigen, als die anderen beiden Varianten.

Wegen der oben beschriebenen Gründe und nicht zuletzt weil der Fokus in den genannten Projekten an der FH Brandenburg auf den Semantic-Web-Technologien liegt, die eher dem Wissensmanagement zuzuordnen sind, wurde der zweite Ansatz für einen Proof of Concept ausgewählt. Der verbleibende Teil des Beitrags ist diesem Ansatz gewidmet, der als Föderation dezentralisierter, annotierter Daten unter Einhaltung von Semantic-Web-Standards bezeichnet werden soll.

# Entwurf der Systemarchitektur

Eine entscheidende Rolle im Systemdesign spielt die zugrundeliegende Domänen- bzw. Problem-Ontologie, welche im folgenden Abschnitt beschrieben wird. Sie begründet ein einheitliches Schema, das bei der Annotation von Webseiten durch die teilnehmenden Hochschulen bzw. Studiengänge einzuhalten ist. Zugleich stellt sie die Basis für die SPARQL-Abfragen zur Analyse der föderierten Daten dar. Für die Annotation der Webseiten soll zunächst nur die Microdata-Technologie (Hickson 2013) berücksichtigt werden, da sie sich perfekt in HTML5 einpasst – dem bevorzugten Standard für moderne Content Management Systeme (CMS).

Bild 2 zeigt eine Skizze des Gesamtdesigns. Die HTML-Quelltexte der registrierten Webseiten werden regelmäßig gecrawlt. Sinnvoll erscheint ein Turnus von einmal täglich bis zweimal wöchentlich, da den betrachteten Informationen keine hochvolatilen Daten zugrunde liegen. Ein Datenbankdienst prüft, ob es Differenzen zwischen den zuletzt gecrawlten und den bislang gespeicherten Daten gibt. Ist das der Fall, wird der neue Quelltext zu einer Komponente geroutet, die aus dem Code die semantischen Daten destilliert. Mit den extrahierten RDF-Daten wird dann die semantische Datenbank (der Triple Store) aktualisiert.

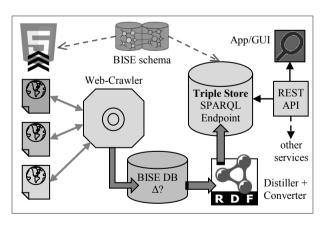


Bild 2: Systementwurf auf Basis der Föderation dezentralisierter Daten

Benutzer interagieren mit dem System über eine Oberfläche, die flexible und performante Suchtechnologien unterstützt. Die semantische Datenbank kann von der Suchanwendung über den SPARQL-Endpoint aufgerufen werden. Ferner erlaubt eine REST-Schnittstelle die Anbindung zusätzlicher interner und externer Dienste zur weiteren Analyse der bereitgestellten Daten, wie auch in (Fagnoni 2014) empfohlen.

Da organisationale Webseiten in fast allen Fällen über CMS ausgeliefert werden, müssen die Annotationstechnologien mit diesen Systemen kompatibel sein. Eine vorläufige Untersuchung ergab, dass die überwiegende Mehrzahl der Hochschulen in Deutschland auf TYPO3 setzt. Dennoch müssen die redaktionellen und steuernden Prozesse für den Betrieb der CMS bzw. einzelner Webseiten im Rahmen eines CMS untersucht werden und geeignete Annotationsansätze dargestellt werden. Moderne CMS sind mit Modulen oder Features für die semantische Annotation ausgestattet, vorwiegend mit Fokus auf schema.org (Brickley e. a. 2015), dem von führenden Suchmaschinen gehosteten Auszeichnungsvokabular. Daneben gibt es das von der Apache Software Foundation betriebene Stanbol-Projekt, das Komponenten der semantischen Erweiterung auch für andere CMS bereitstellt (Apache 2010, Behrend 2013).

### Kurzbeschreibung der Ontologie

Die Kernkonzepte der Ontologie wurden bereits im Kapitel "Beschreibung der Domäne" dargestellt. Im Mittelpunkt steht eine Klasse für die WI-Studiengänge. Auch wenn sich das Anwendungsfeld der Ontologie im deutschsprachigen Raum befindet, gehört es zur gängigen Praxis, semantische Modellierungen in englischer Sprache auszuführen. Deshalb heißt die zentrale Klasse: BISEDegreeProgram (BISE steht für Business & Information Systems Enginieering – eine übliche englischsprachige Bezeichnung für Wirtschaftsinformatik). Eine technische Dokumentation der Ontologie findet sich in (Meister 2015).

Die Wiederverwendung von Standardvokabular-Elementen gilt in der Ontologie-Entwicklung als weithin akzeptierte gute Praxis (Suarez-Figueroa 2013). Auf dem Gebiet der Annotation von Webseiten ist zuvorderst schema.org zu berücksichtigen. Das zentrale Konzept "WI-Studiengang" jedoch ist zu spezifisch und - zumindest zum jetzigen Zeitpunkt - nicht Bestandteil dieses Vokabulars, das sich tatsächlich noch stark im Aufbau befindet. (Am umfassendsten sind aktuell die Domänen Medizin/Pharmazie sowie E-Commerce ausspezifiziert.) In diesem Fall ist es empfehlenswert, ein allgemeineres Konzept zu suchen, das das spezifische Konzept als Unterklasse beherbergt. Leider kennt schema.org aktuell auch kein Konzept "Studiengang". Deshalb wählten wir zur semantischen Einordnung der zentralen Klasse das generische Konzept Service. In der Tat kann ein Studiengang als ein Dienst angesehen werden, den eine Hochschule anbietet. Das lässt sich in schema.org gut abbilden. Es gibt die Klasse CollegeOrUniversity welche als Range der Object Property provider im Hinblick auf die Klasse BISEDegreeProgram auftritt. Letzere erbt die genannte Property von der übergeordneten Klasse Service.

Wie bereits beschrieben, sind wir daran interessiert, bestimmte qualitative Eigenschaften von Studiengängen zu explizieren und zu analysieren. Im Einzelnen sind das die Ausprägung der WI-Säulen und die Qualifikation für bestimmte Tätigkeitsfelder, die einen WI-Studiengang charakterisieren. Folglich brauchen wir zwei weitere spezifische (nicht-Standard) Klassen BISEPillar und BISEJobProfile, beide als Unterklassen von schema:QualitativeValue und beide mit Referenzen vom Studiengang über die Standard-Property schema: valueReference. Die Relation zwischen diesen beiden spezifischen Klassen kann als educationRequirements modelliert werden, einer weiteren Standard-Property aus schema.org. Die Klassifikation als schema:QualitativeValue impliziert einen Enumerationstyp für beide Klassen, d. h. sie verfügen über eine vordefinierte Anzahl von Instanzen (auch Named Individuals genannt), die alle möglichen Ausprägungen des zentralen Dienstes im Hinblick auf eine qualitative Eigenschaft abdecken. Diese vordefinierten Instanzlisten wurden bereits im Kapitel "Beschreibung der Ontologie" eingeführt und begründet. Weitere Elemente der Ontologie, wie in Bild 3 dargestellt, sind leicht nachzuvollziehen. Nahezu alle Elemente sind schema.org entnommen.

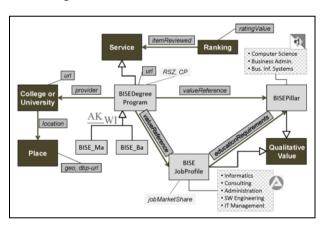


Bild 3: Die BISE-Ontologie

### PROOF OF CONCEPT

Die folgenden Komponenten des entworfenen Systems wurden einem Proof of Concept unterworfen:

- Ontologie-basierte Annotation von exemplarischen, lokal gespeicherten WI-Studiengangsseiten durch unmittelbare Implementierung im HTML-Code;
- Destillation von RDF-Daten aus dem HTML-Code und Konvertierung dieser Daten in Turtle-Code einschließlich einer Qualitätskontrolle (was schrittweise zu Verbesserungen bei der Annotation führte);
- Föderation der destillierten Daten auf einem semantischen Server und Verschmelzung mit der BISE-Ontologie;
- Entwicklung und Test grundlegender SPARQL-Abfragen.

Wie bereits erwähnt, erfolgte die Annotation unter Verwendung der Microdata-Technologie. Sie erlaubt verschiedene Vorgehensweisen für die Einbindung strukturierter, semantischer Daten in den HTML-Code einer Webseite. Da das Projekt darauf abzielt, die Aktualität des Systems über die allfälligen redaktionellen Prozesse auf der Studiengangswebseiten abzusichern, erscheint eine direkte Platzierung von kompakten semantischen Daten im Head-Element einer HTML-Seite, d.h. ohne Verbindung zu einem sichtbaren oder zumindest editierbaren Element, von geringem Interesse. Umgekehrt verspricht die Annotation editierbarer Elemente, wie Überschriften, Absätze oder Text-Abschnitte, eine wirksame Unterstützung der gewöhnlichen redaktionellen Aktivitäten auf den dezentralen Studiengangswebseiten.

Microdata basiert auf der Verwendung von fünf Attributen, die für die Mehrzahl aller HTML-Elemente zulässig sind: itemtype, itemscope, itemprop, itemid und itemref. Die ersten drei gelten als besonders wichtig für die Optimierung von Webseiten im Hinblick auf Suchmaschinen (Ludwig 2014). In unserem Fall kann die Suchmaschinen-Optimierung als ein erfreulicher Nebeneffekt angesehen werden. Die Vernachlässigung itemid-Attributen in Microdata-Annotationen erzeugt eine Menge leerer Knoten in den destillierten RDF-Daten. Das führt dazu, dass die Datenföderation und damit eine effektive Datenanalyse mindestens behindert, möglicherweise sogar verhindert werden. Bild 4 zeigt einen Ausschnitt eines beispielhaften Webseiten-Codes, welcher sowohl direkt ausgezeichnete Daten (mit Hilfe von span- und data-Elementen), als auch die Annotation von redaktionell bearbeitbaren HTML-Elementen, wie h1, enthält.

Bild 4: Ausschnitt eines annotierten HTML-Codes

Für den nächsten Schritt wurde das Tool pyMicrodata (Herman 2012) eingesetzt. Es destilliert und konvertiert die RDF-Daten der Webseite. Das Tool wird vom W3C bereitgestellt und kann sowohl über eine Weboberfläche als auch über eine Programmierschnittstelle angesprochen werden. Die destillierten Daten wurden zusammen mit der BISE-Ontlogie in den Default Graph eines

Apache Jena Fuseki Servers verschmolzen (Apache 2015). Dieser als Open Source bereitgestellte Server ist leicht zu installieren. Er bietet eine vollumfängliche Unterstützung sämtlicher SPARQL-Funktionalitäten und zeigt bei nicht zu großen Datenbanken eine zufriedenstellende Performance. Im letzten Schritt erfolgten Tests auf Basis von SPARQL-Abfragen zur Beantwortung von Fragen, wie: "Welche Studiengänge qualifizieren für eine Tätigkeit als IT-Berater?"

Zusammenfassend kann der Proof of Concept als erfolgreich angesehen werden. Die größten Herausforderungen sind in der Bereitstellung von Annotations-Hilfsmitteln zu sehen, nicht zuletzt im Hinblick auf die Vielfalt von CMS-Implementierungen. Auch wenn die meisten CMS heute Annotationen grundsätzlich erlauben, kann ein Hindernis darin bestehen, dass der Zweck für semantische Annotation primär in der Suchmaschinenoptimierung gesehen wird und deshalb die Microdata-Technologie nicht in vollem Umfang unterstützt wird.

### FAZIT UND AUSBLICK

Auch wenn der Proof of Concept die technische Machbarkeit des Zielsystems in wesentlichen Aspekten belegt hat, verbleibt ein großer Teil der Arbeit um das beschriebene System in voller Funktionalität zu entwickeln und zu implementieren. Zunächst ist geplant, einen Prototypen umzusetzen, der zwar die volle Funktionalität implementiert, aber nur eine kleine Anzahl von Hochschulpartnern im Sinne von Early Adopters berücksichtigt. Auf der Grundlage der in diesem Partnernetzwerk gesammelten Erfahrungen können dann Handreichungen für weitere Systempartner entwickelt werden, die neuen Teilnehmern den Anschluss erleichtern.

Der Erfolg des Projektes ist eng verknüpft mit der Bereitschaft von WI-Studiengängen an Fachhochschulen im deutschsprachigen Raum, sich dem beschriebenen System anzuschließen. Folglich sind auch Maßnahmen zur stärkeren Wahrnehmung dieses Projektes zu planen und umzusetzen. Alternativ oder ergänzend kann parallel ein Projekt zur Umsetzung des als Big-Data-Ansatz beschriebenen Konzeptes angedacht werden. Erste Gespräche mit einschlägigen Experten haben bereits stattgefunden. Als Fernziel kann auch der Transfer eines solchen Informationssystems auf andere Fachrichtungen betrachtet werden.

### **OUELLEN**

- AKWI. 2015. Wirtschaftsinformatik an Fachhochschulen im deutschsprachigen Raum. Zugang http://akwi.de, (abgerufen am 18.10.2015).
- Allemang, D. und J. Hendler. 2011. *Semantic Web for the Working Ontologist*. 2<sup>nd</sup> Ed., Morgan Kaufmann, Amsterdam e.a.
- Apache Software Foundation. 2015. *Apache Jena Fuseki*. Zugang https://jena.apache.org/documentation/fuseki2/, (abgerufen am 18.10.2015).
- Apache Software Foundation. 2010. *Apache Stanbol*. Zugang https://stanbol.apache.org/, (abgerufen am 18.10.2015).
- Balzert, H. 2008. Lehrbuch der Softwaretechnik Softwaremanagement. 2. Aufl., Spektrum Akademischer Verlag, Heidelberg.
- Beckmann, R.; S. Lindner und S. Werner. 2015. *Der Arbeitsmarkt für IT-Fachleute in Deutschland*. Bundesagentur für Arbeit, Nürnberg.
- Behrend, W. und V. Damjanovic (Ed.). 2013. *Developing Semantic CMS Applications The IKS Handbook*. Salzburg Research Forschungsgesellschaft, Salzburg.
- Brickley, D.; S. Corlosquet e. a. 2015. *schema.org*. Zugang http://schema.org/, (abgerufen am 29.09.2015).
- Fagnoni, E. 2014. Business Ontology Toolkit (BOTK). Zugang http://ontology.it/tools/botk4/overview/, (abgerufen am 02.11.2015).
- Gesellschaft für Informatik e. V. 2003. "Rahmenempfehlung für die Universitätsausbildung in Wirtschaftsinformatik". In *Informatik Spektrum*, Bd. 26/2, Springer, Heidelberg, S. 1 20.
- Großmann, N; T. Krogoll und V. Meister. 2005. *Ausbilden mit Lernaufgaben*. Bd. 1, Christiani, Konstanz.
- Herman, I. 2012. *Microdata to RDF Distiller*. Zugang http://www.w3.org/2012/pyMicrodata/, (abgerufen am 18.10.2015).
- Hickson, I. (Ed.). 2013. *HTML Microdata*. Zugang http://www.w3.org/TR/microdata/, (abgerufen am 18.10.2015).
- HRK. 2015a. *Hochschulkompass erweiterte Suche*. Zugang http://hochschulkompass.de, (abgerufen am 18.10.2015).
- HRK. 2015b. *Hochschulrektorenkonferenz*. Zugang http://hrk.de, (abgerufen am 18.10.2015).
- Ludwig, T. 2014. How to use HTML5 and structured Microdata Schema.org for SEO? Zugang https://ludwig.im/, (abgerufen am 18.10.2015).
- Meister, V. 2015. *BISE Ontologie Dokumentation*. Zugang http://www.essepuntato.it/lode/lang=de/http://univera.de/FHB/StudyScout/bise\_schema.rdf, (abgerufen am 02.11.2015).
- Suarez-Figueroa, M.C. 2013. Neon Methodology for Building Ontology Networks: Specification, Scheduling and Reuse. IOS Press, Clifton, VA.