

## AIFB Themenheft 2014 Semantik in Big Data



Institut für  
Angewandte Informatik  
und Formale  
Beschreibungsverfahren



**Einladung**

**30. AIK-Symposium  
„Semantik in Big Data“  
Karlsruhe  
31. Oktober 2014**

## Einladung

# 30. AIK-Symposium „Semantik in Big Data“ 31. Oktober 2014 Renaissance Karlsruhe Hotel

## Programm

10:00 – 12:00	Treffen der Lehrbeauftragten und Honorarprofessoren des Instituts AIFB
13:00 – 14:00	Mitgliederversammlung AIK e.V.
14:15 – 14:30	Eröffnung und Begrüßung
14:30 – 15:00	<b>Semantik in Big Data.</b> <b>Cross-linguale und cross-modale Medienanalyse von Livestreams</b> Dr. Achim Rettinger, <a href="#">Institut AIFB</a> , <a href="#">Karlsruher Institut für Technologie</a>
15:00 – 15:30	<b>Semantische Anreicherung und Linking von Presstexten.</b> <b>Eine Luxid Fallstudie</b> Stefan Geißler, <a href="#">TEMIS Deutschland GmbH</a>
15:30 – 16:15	Kaffeepause
16:15 – 16:45	<b>Welchen Teil von Klick hast Du nicht verstanden?</b> <b>Personalisierung, Big Data und die Bedeutung von Mausclicks</b> Dr. Sebastian Blohm, <a href="#">Microsoft</a> , <a href="#">Cambridge</a> , GB
16:45 – 17:15	<b>Maschinelles Lernen mit Semantischen Graphen:</b> <b>eine neue Perspektive für Big Data</b> Prof. Dr. Volker Tresp, <a href="#">Siemens</a> , <a href="#">Corporate Technology</a> und <a href="#">Ludwig-Maximilians-Universität München</a> , <a href="#">Institut für Informatik</a>
17:15 – 17:45	<b>Mit semantischen Technologien von Big Data zu Smart Data</b> Dr. Peter Haase, <a href="#">fluid Operations AG (fluidOps)</a>
17:45 – 18:15	Preisverleihung
ab 18:30	Abendessen

### Anmeldung

Teilnahmebeitrag für AIK-Mitglieder	€ 30
Teilnahmebeitrag für Nichtmitglieder*	€ 70
Beitrag zum Abendessen	€ 30

Wir bitten um Anmeldung zum Symposium und Überweisung des Beitrags bis zum 17. Oktober 2014. Bei Rücktritt bis zum 24. Oktober 2014 werden die entrichteten Gebühren erstattet.

**Den Link zur Anmeldung und weitere organisatorische Informationen finden Sie unter:**  
[www.aik-ev.de](http://www.aik-ev.de)

Übernachtungsmöglichkeiten bestehen im Renaissance Karlsruhe Hotel (Sonderrate unter Stichwort „AIK 30“).

\* Bei gleichzeitigem Vereinsbeitritt sind nur der Mitgliedsbeitrag von € 25 für das Jahr 2014 sowie der Teilnahmebeitrag für Mitglieder in Höhe von € 30 zu entrichten. Unternehmen, die eine Firmenmitgliedschaft im AIK e.V. haben, können bis zu drei Firmenangehörige zu je € 30 entsenden.

*Sehr geehrte Damen und Herren,  
liebe Freunde und Förderer des  
Instituts AIFB,*



wenn sich abzeichnet, dass ein am Institut bearbeitetes Forschungsgebiet hohe Relevanz für die Wirtschaft gewinnt, machen wir es mit dem Verein Angewandte Informatik Karlsruhe (AIK) e.V. zum Thema eines AIK-Symposiums. Das erste fand 1998 zu Business Intelligence statt. 16 Jahre später geht es beim 30. AIK-Symposium „Semantik in Big Data“ wieder um grundlegende Fragen der Nutzung semantischer Technologien zur Unterstützung der Wertschöpfung aus Daten; im Besonderen um den Reifegrad von Big Data Analytics und die Chancen für Unternehmen, die sich aus der Verfügbarkeit mächtiger Datenvolumina und schier unendlicher Informationsquellen ergeben.

Big-Data-Analysen sind im Wesentlichen Auswertungen von Daten zur Entscheidungsfindung. Aber von welchen Daten eigentlich? Das ist die Kernfrage, mit der sich die Forschungsgruppe Wissensmanagement am Institut beschäftigt, die für die inhaltliche Ausrichtung des Symposiums verantwortlich zeichnet. Die Gruppe um Rudi Studer, Semantik-Forscher der ersten Stunde, vertritt die Ansicht, dass der aktuelle Umgang mit Big Data wesentliche Dimensionen zu wenig berücksichtigt. Aus ihrer Sicht ist es wichtig, neben dem Einfluss der Faktoren Verarbeitungszeit und Dynamik vor allem die Heterogenität der Daten, den Grad an Strukturiertheit sowie die Komplexität der Abhängigkeiten zwischen den Daten zu adressieren.

Big-Data-Analysen werden schon heute wichtigen geschäftlichen Entscheidungen zugrunde gelegt. Sie sind Arbeitswerkzeug für Forschung, Marketing und Patentbeobachtung in Industrieunternehmen und dienen als Produktionsrohstoff für Informationsdienste aller Art. Im eCommerce liefern sie den Anbietern tiefe Einblicke in das Verhalten von Kunden. Im Zusammenhang mit Industrie 4.0, Smart Energy, Smart Grids und Smart Cities, ja sogar einer zukünftigen personalisierten Gesundheitsversorgung, werden Big Data analysiert. Anfang des Jahres hat das KIT ein „Smart Data Innovation Lab“ eröffnet, das sich mit ebendiesen Themen befasst.

Sie sind an Big Data und ihrem Mehrwert interessiert? Wir laden Sie herzlich ein, sich von Referenten aus Wissenschaft und Wirtschaft beim 30. AIK-Symposium über die semantischen Aspekte von Big Data informieren zu lassen und diese mit den Vortragenden und mit uns zu diskutieren. Die Referenten stellen erfolgreiche Big-Data-Anwendung in Unternehmen und aktuelle Forschungsprojekte vor, sprechen offene Forschungsfragen an und stellen Lösungsansätze zur Diskussion.

Welche Forschungsarbeiten uns über Big Data hinaus im abgelaufenen Jahr am Institut beschäftigt haben, welche Erfolge und Höhepunkte das Institutsjahr brachte, darüber informiert Sie dieses Themenheft auf den Seiten 8 bis 11. Hervorheben möchten wir daraus die 18 erfolgreichen Dissertationen und drei Rufe auf Professuren, die Institutsmitarbeiter erhalten haben.

Wenn Sie sich für spezielle Themen unserer Forschung und Lehre interessieren, nehmen Sie gerne Kontakt mit uns auf. Aus dem regen Austausch zwischen dem Institut, den Mitgliedern des Vereins AIK e.V. sowie Besucherinnen und Besuchern der AIK-Symposien sind bereits zahlreiche produktive Forschungs- und Innovationsprojekte entstanden.

*Wir danken Ihnen für Ihr Interesse an unserem Institut  
und freuen uns darauf, Sie beim 30. AIK-Symposium  
zu treffen.*

Andreas Oberweis

Hartmut Schmeck

Detlef Seese

Rudi Studer

## Volume, Velocity, Variety: Big Data hat mehr als eine Dimension

Bei der Erschließung des Potentials von Big Data werden aus Sicht der Forschung am Institut AIFB wesentliche Dimensionen noch zu wenig adressiert. Es geht nicht nur um das *Volumen*, die schier unendliche Menge verfügbarer Daten. Dynamik und Verarbeitungsgeschwindigkeit, *Velocity*, haben ebenso Einfluss auf die Auswertungen, besonders allerdings die *Variety*. Dieser Fachbegriff fasst die Herausforderungen zusammen, die sich ergeben aus der Struktur, in der die Daten vorliegen, der Vielfalt der Datenquellen, Datenformate und Landessprachen. Vor allem aber geht es darum, Rohdaten so zu beschreiben, dass deren Bedeutung und Inhalt maschinell interpretierbar wird.

Semantik in Big Data greift diese Herausforderungen auf. In diesem Spezialgebiet des Wissensmanagements werden Methoden und Technologien zur Datenrepräsentation und deren Annotation erforscht und entwickelt.

Ziele der Forschung sind, durch semantische Verfahren a) Rohdaten automatisch zu höherwertigen Datenstämmen zu veredeln, also automatisch Metadatensätze abzuleiten, die den Inhalt beschreiben, b) heterogene Datenquellen zu verknüpfen und auswertbar zu machen sowie c) Big-Data-Analysen durch semantische Datenererschließung zu verbessern, um zuverlässige Ausgangsinformationen für Interpretationen und daraus abgeleitete Vorhersagen zu gewinnen.

### Wertschöpfung aus Big Data

Resultierend aus der immer weiter steigenden Verfügbarkeit großer Datenmengen ist *Big Data* seit 2012 ein zentraler Begriff in der IT-Landschaft geworden. Mit der Datenmenge wächst auch das Interesse. Bereits 2012 wurden über 10 Milliarden US-Dollar Umsatz im Bereich Big Data generiert. Bis 2017 sollen sich die Zahlen verfünffachen.

Meist liegt der Fokus auf dem *Volumen*. Also der Größe und Anzahl der Datenpunkte, die in Form von Texten, Audios, Videos, Grafiken und Bildern, Fakten-, Forschungs- und Messdaten von unzähligen Anwendern auf der ganzen Welt generiert werden. Bei YouTube beispielsweise werden derzeit pro Minute ca. 60 Stunden an Videomaterial hochgeladen. Etwa 250 Millionen Tweets sowie 800 Millionen Facebookbeiträge und Klicks werden pro Tag generiert.

Um mit diesen Datenmengen umzugehen, wurden zur Verteilung und Parallelisierung der Berechnungen in den letzten Jahren sehr erfolgreich neue Modelle und Software Frameworks wie *Apache Hadoop* entwickelt. Dadurch und durch die einfache Verfügbarkeit von Cloud Computing sind große Datenmengen heute selbst für mittelständische Unternehmen mit begrenzten finanziellen und personellen Ressourcen verarbeitbar.

Am Beispiel eines Videokanals dargestellt bezeichnet *Volume* die reine Datenmenge, also hauptsächlich die Auflösung des Videosignals und die daraus resultierende Menge an Pixeln. *Velocity* befasst sich mit der flüssigen Verarbeitung des Videosignals und kurzen Latenzzeiten. *Variety* beschäftigt sich mit dem tatsächlichen Inhalt der Sendungen. Das sind z.B. Informationen, die a) aus der gesprochenen Sprache im Tonsignal extrahiert werden können, b) zeigen, welche Metainformationen zum Video vorliegen, oder c) welche Objekte im Bildsignal zu erkennen sind und deren Abhängigkeiten untereinander.

Viele der mit *Variety* verbundenen Aspekte fallen in das Themengebiet der *Semantik*, also der inhaltlichen Bedeutung von Daten. Dieser Bereich ist Kernkompetenz unserer Gruppe am Institut AIFB. Wir gehen somit davon aus, dass semantischen Technologien eine zentrale Rolle zukommt, um die Herausforderungen von Big Data zu bewältigen.

### Semantische Datenanalyse für Big Data

Die Arbeit unserer Gruppe konzentriert sich in diesem Themenfeld auf drei Gebiete:

a) Unstrukturierte Daten, wie Texte, Bild- und numerische Daten, müssen semantisch angereichert werden. Dieser Prozess, bei dem höherwertige Information aus den Rohdaten extrahiert wird, wird auch oft als semantische Annotation bezeichnet. Beispielsweise kann man bei einem Nachrichtentext die darin genannten Entitäten automatisch erkennen und vorhandenem Wissen zuordnen, etwa einem Wikipedia-Eintrag. Die so gewonnenen Annotationen eines Dokuments, deren Bedeutung und Abhängigkeiten untereinander und Beziehung zum Hintergrundwissen lassen sich in abstrakten Strukturen aus der Graphentheorie abbilden. Somit sind sie dem Computer für eine semantische Analyse zugänglich. Solche Graphstrukturen bieten sich als natürlicher Repräsentationsformalismus für solche Informationen an, da sie Objekte (Knoten) und deren Relationen (Kanten) untereinander, wie z.B. Personen und deren soziale Beziehungen, abbilden können. Graphen erlauben darüber hinaus das Darstellen reichhaltiger Metadaten und komplexer Beziehungen wie Zugehörigkeit zu abstrakten Klassen und Vererbungshierarchien.

b) Big-Data-Probleme erfordern häufig die Integration vieler heterogener Datenquellen. Hierbei ist eine semantische Repräsentation zur Zusammenführung komplexer, aber teilweise überlappender Graphen sehr hilfreich.

c) Erst danach erfolgt die eigentliche inhaltliche Analyse der Daten auf den großen, in den vorherigen Schritten gewonnenen Graphen. Typische Aufgabenstellungen sind das Erkennen von duplizierten Entitäten im Graphen, die Berechnung der semantischen Ähnlichkeiten von Subgraphen oder die Vorhersage von zusätzlichen Beziehungen.

In allen gerade genannten Schritten erfordert eine Berücksichtigung der Semantik in Big Data grundlegend neue Berechnungsmodelle. Die Mehrzahl der Big Data Berechnungsparadigmen ist Datenpunkte zentriert. Unter Beachtung der Semantik liegen die Bedeutung und damit letztendlich der Wert der Daten allerdings nicht in den Datenpunkten, sondern in der Abhängigkeit der Datenpunkte untereinander. Um in großen Graphen repräsentierte Daten effizient, also verteilt und asynchron, verarbeiten zu können, müssen Algorithmen weg von der datenparallelen Verarbeitung hin zu einer graphparallelen Verarbeitung optimiert werden.

Das Optimierungspotential für Big-Data-Analysen schöpft sich aus der Beobachtung, dass in Abhängigkeit stehende Objekte sich auch stärker beeinflussen. Die Knoten im Graph müssen somit basierend auf den Knoten in der Nachbarschaft bearbeitet werden und nicht wie bisher zeilen- oder spaltenweise wie in Datenbanktabellen hinterlegt.

**Dr. Achim Rettinger referiert auf dem 30. AIK-Symposium über den Einfluss der Struktur von Daten und zugehöriger Metadaten auf die Big-Data-Dimensionen „Volume“ und „Velocity“. Er beschäftigt sich als KIT-Nachwuchsgruppenleiter am Institut AIFB mit wissenszentrierten und lernfähigen Datenanalyseverfahren. Dies umfasst graphbasierte Datenrepräsentationen und Methoden, die sich mit zunehmenden Datenmengen kontinuierlich verbessern.**

**Professor Dr. Rudi Studer hat sich mit seiner Forschung zur Semantik internationales Renommee erworben. Er forscht und lehrt seit dem Aufkommen der semantischen Datenaufbereitung auf diesem Gebiet, war Gründungspräsident der Semantic Web Science Association und Editor-in-chief des Journal of Web Semantics: Science, Services and Agents on the World Wide Web. Er ist Professor am Institut AIFB sowie Direktor am Karlsruhe Service Research Institut (KSRI) des KIT und am FZI Forschungszentrum Informatik am KIT.**

# Mit semantischen Technologien von Big Data zu Smart Data

Bei den meisten heute auf dem Markt erhältlichen Big-Data-Lösungen steht das Volumen im Mittelpunkt. Sie unterstützen insbesondere die vertikale Skalierbarkeit. Das Hauptproblem ist aber die horizontale Skalierbarkeit. Denkt man nur an die Fülle von veröffentlichten Daten in Open-Data-Initiativen oder an die Datenmengen, die in unternehmensinternen Systemen und Datenbanken gespeichert sind, dann steht der Anwender vor zahlreichen Datenquellen, vielen verschiedenen Themen, Datenmodellen und Strukturen.

Die fluid Operations AG (fluidOps) bietet mit der Information Workbench eine Plattform für das semantische Datenmanagement, die neben dem Volume (Datenmenge) insbesondere auch die Dimensionen Variety (Heterogenität in unterschiedlichen Datentypen und Quellen) und Velocity (Menge der Datenströme und deren Verarbeitung) adressiert. Die Information Workbench unterstützt Anwender und IT-Nutzer im Big-Data-Kontext bei der Integration und Korrelation heterogener und verteilter Big Data an einer zentralen Stelle.

Hierzu liegen Ergebnisse aus aktuellen Forschungsprojekten und industriellen Anwendungen auf Basis der Information Workbench vor:

Optique (<http://www.optique-project.eu/>) ist zum Beispiel ein groß angelegtes aktuelles europäisches Forschungsprojekt, bei dem der Zugriff auf Big Data durch Endnutzer im Mittelpunkt steht. Zentraler Baustein ist hier der Ansatz von OBDA (Ontology-Based Data Access), bei dem Ontologien als konzeptuelle Modelle für den Datenzugriff zum Einsatz kommen. Dieser Ansatz ermöglicht eine durchgehende Verbindung zwischen Endnutzern und großen, verteilten, heterogenen Datenquellen. Statt auf einen IT-Experten für den Datenzugriff zurückgreifen zu müssen, werden Endnutzer in die Lage versetzt, selbst intuitive Anfragen über ihnen vertraute Vokabulare und Konzeptualisierungen zu stellen. Anfragen gegen die Ontologie werden automatisch in verteilte Anfragen über den ursprünglichen Datenquellen übersetzt und ausgewertet. Die Beziehung zwischen den Termen der Ontologie und den internen Strukturen der Datenquellen (z.B. Tabellen- und Spaltennamen in relationalen Datenbanken) werden in Form von deklarativen Mappings verwaltet.

Die Innovationen des Optique-Ansatzes betreffen:

- Komponenten für das Management der Ontologien und Mappings,
- endnutzerfreundliche Schnittstellen für das visuelle Formulieren von Anfragen,
- die automatische und effiziente Übersetzung von Anfragen mit expressiven Ontologien,
- die Anfrageoptimierung für die verteilte und skalierbare Auswertung von Anfragen unter Verwendung von dynamisch allokierten Cloud-Ressourcen,
- Methoden für die Behandlung von temporalen Anfragen und Datenströmen,
- die Integration all dieser Komponenten in einem umfassenden, flexiblen und einfach nutzbaren System: der Optique-Plattform.

Die Information Workbench ist Basis für die Entwicklung der Optique-Plattform.



Für den Einsatz der Plattform gibt es konkrete Anwendungsszenarien aus der Industrie, speziell im Bereich Energie:

- In einem Szenario bei Siemens erfassen beispielsweise Diagnose-Ingenieure in Kraftwerken Ereignisse in zeitgestempelten Sensordaten. Die Diagnose-Ingenieure können aber nur dann mit ihren Visualisierungs- und Trenderkennungstools arbeiten, wenn sie Sensor- und Ereignisdaten über verschiedene und verteilte Datenbanken hinweg abfragen können. Sensordaten können die Größe von mehreren Terabytes haben, Ereignisdaten – wie z. B. „Alarm ausgelöst zum Zeitpunkt X“ – die Größe von mehreren Gigabytes. Mit einem täglichen Zuwachs von 30 Gigabytes übersteigt die Gesamtmenge der Rohdaten das Limit dessen, was sie derzeit speichern können.
- In einem weiteren Szenario bei Statoil entwickeln Geologen und Geophysiker stratigraphische Modelle unerforschter Gebiete auf der Grundlage von Daten aus früheren Untersuchungen in der nahegelegenen Umgebung. Damit sie die Daten in ihr leistungsstarkes Visual Analytics Tool speisen können, müssen sie relationale Daten in der Größe von über 1000 Terabytes abfragen. Die Daten sind unterschiedlich strukturiert und beinhalten über 2000 Tabellen in unterschiedlichen und verteilten Datenbanken.

**Unser Gastautor Dr. Peter Haase leitet seit 2009 bei fluidOps die Forschung und Entwicklung. Zuvor war er von 2003 bis 2009 am Institut AIFB, wo er 2006 in der Gruppe von Professor Studer promovierte. Seine Interessen und Expertise umfassen semantische Technologien, Linked Data, Ontologien und Cloud Computing. In seinem Vortrag im Rahmen des 30. AIK-Symposiums wird er die Information Workbench sowie Forschungsprojekte und Anwendungsszenarien in diesem Zusammenhang näher vorstellen.**



## Semantik in Big Data 30. AIK-Symposium



## Highlights 2013/14 Erfreuliches aus dem Institutsleben



# Lehre + Forschung

Lehre und Forschung am Institut AIFB von außen positiv bewertet zu sehen, gehört zu den schönsten Ereignissen im Institutsleben. Im Berichtszeitraum 2013/14 durften wir uns gleich mehrfach freuen. Sieben Institutsmitglieder bekamen im Rahmen der uns sehr wichtigen Lehr-evaluation durch die Studierenden der KIT-Fakultät für Wirtschaftswissenschaften ein „Zertifikat für gute Lehre“ ausgestellt. Wir sind stolz darauf.

Für ihre Vorlesungen und Übungen zu „Grundlagen der Informatik II“ wurde die Gruppe von Hartmut Schmeck mit dem Lehrpreis der Fakultät ausgezeichnet. Lukas König, Sanaz Mostaghim und Friederike Pfeiffer-Bohnen gestalten sie mit Hartmut Schmeck gemeinsam. Sie unterstützen die Studierenden beim Erarbeiten des Wissens unter anderem durch neuartige Lehr- und Übungsbücher, die eLearning und Präsenzlehre verbinden. Das Buchkonzept für „100 Übungsaufgaben zu Grundlagen der Informatik“ (Oldenbourg Verlag / De Gruyter, 2013) hat die Forschungsgruppe kooperativ entwickelt und umgesetzt. Aus jedem Buchkapitel führt ein Link zur audiovisuellen Aufzeichnung der entsprechenden Vorlesung. Die Übungsaufgaben sind mit einem Diskussionsforum und dort direkt mit dem Diskussionsstrang zu dieser Aufgabe verknüpft. Der Ansatz bewährt sich hervorragend. Die Studierenden empfinden ihn als Bereicherung, weil sie dadurch Fragen sofort selbständig lösen oder zumindest weiterbearbeiten können.

Hochwertige Lehre mit persönlicher Beratung und Betreuung, umgesetzt mit innovativen Methoden und Werkzeugen, betrachten wir als Grundlage eines guten Studienkonzeptes, das den Bedürfnissen und Erwartungen der Studierenden entspricht. Um diesem Anspruch an uns selbst dauerhaft gerecht werden zu können, beteiligen wir uns regelmäßig an Forschungsprojekten zur Weiterentwicklung der Lehre. Vielversprechende Arbeiten finden dazu derzeit im Rahmen der Projekte IFS und nuKIT statt. IFS steht für ein „Integriertes Feedback-System“, welches die Gruppe von Andreas Oberweis als Teil des vom Bundesministerium für Bildung und

Forschung (BMBF) finanzierten KIT-Gesamtprojektes „KIT-Lehre hoch Forschung“ aufbaut. nuKIT ist eine Smartphone-Anwendung aus der Forschungsgruppe von Hartmut Schmeck, die aktive Direktkommunikation zwischen Studierenden und Vortragenden möglich macht. Es können Fragen gestellt, Antworten gegeben und Vortragsparameter bewertet werden. nuKIT ist am KIT bereits in zahlreichen Vorlesungen im Einsatz.

Auch in Rufen und Berufungen aus der wissenschaftlichen Gemeinschaft drückte sich Aufmerksamkeit und Anerkennung für unsere Arbeit aus. Stefan Tai, der im November 2007 vom IBM Thomas J. Watson Research Center, New York, zu uns kam, wechselt nach Berlin. Sieben Jahre forschte und lehrte er am Institut AIFB. Jetzt hat er einen Ruf an die Fakultät für Elektrotechnik und Informatik der TU Berlin angenommen. Er baut dort das Fachgebiet „Wirtschaftsinformatik – Information Systems Engineering“ auf. Wir danken ihm für sein Engagement am Institut AIFB und wünschen ihm alles Gute! Christian Janiesch aus dem Team von Stefan Tai hat einen Ruf als Juniorprofessor für Wirtschaftsinformatik an die Universität Würzburg erhalten. Sanaz Mostaghim, seit Oktober 2013 im Rahmen einer Dorothea-Erxleben-Gastprofessur als Professorin an der Universität Magdeburg, fand sich in der luxuriösen Lage, zwischen drei Rufenden wählen zu können. Entschieden hat sie sich für eine Professur an der Universität Magdeburg. Wir wünschen ihr an der neuen Wirkungsstätte viel Erfolg!

Andreas Oberweis wurde in den Vorstand der Gesellschaft für Informatik (GI) e.V. gewählt. Er steht damit als einer von drei Vizepräsidenten neben dem Präsidenten an der Spitze der mit rund 20.000 Mitgliedern größten Vertretung von Informatikerinnen und Informatikern im deutschsprachigen Raum. Agnes Koschmider wurde als eine der ersten Nachwuchswissenschaftlerinnen von der GI als Juniorfellow ausgezeichnet. Juniorfellows bietet die GI einen Rahmen, eigenverantwortlich Ideen zur Gestaltung der Informatik in Wissenschaft und Gesellschaft zu entwickeln und umzusetzen.



Achim Rettinger, Projektleiter in der Forschungsgruppe Wissensmanagement, ist zum Leiter der KIT-Nachwuchsgruppe „Adaptive Data Analytics“ ernannt worden. Die Nachwuchsgruppe beschäftigt sich mit der semantischen Anreicherung unstrukturierter Daten (multilingualer Text, multiple Modalitäten) und statistischer Inferenz auf strukturierten Daten.

Seit Jahren sind Professoren des Instituts AIFB dem FIZ Karlsruhe, Leibniz-Institut für Informationsinfrastruktur, in vielfältiger Weise verbunden, sei es durch die Mitgliedschaft im Aufsichtsrat, im wissenschaftlichen Beirat oder durch gemeinsame Forschungsprojekte. In Zukunft wird sich die Zusammenarbeit weiter vertiefen. Im Zuge der Verstärkung der Forschungsaktivitäten am FIZ Karlsruhe werden in Kooperation mit dem KIT zwei neue Professuren eingerichtet. Eine Professur mit der Widmung „Immaterialgüterrechte in verteilten Informations-Infrastrukturen“, entsteht in Zusammenarbeit mit der KIT-Fakultät Informatik. Die zweite Professur mit der Widmung „Information Service Engineering“ wird mit dem Institut AIFB eingerichtet.

Freuen durften wir uns einmal mehr auch über den Zuschlag für mehrere europäische und nationale Verbundforschungsprojekte. Hartmut Schmeck und seine Mitarbeiter im Institut für Angewandte Informatik (IAI) des KIT und im AIFB waren wesentlich an der erfolgreichen Antragstellung für die nächste fünfjährige Förderperiode (2015-2019) der Helmholtz Programme „Supercomputing und Big Data“ sowie „Speicher und vernetzte Infrastrukturen“ beteiligt.

## 4 Projekte

stehen beispielhaft für aktuelle Forschungs- und Entwicklungsvorhaben am Institut AIFB. Als Teil der Aktivitäten der Knowledge and Innovation Community InnoEnergy des EIT European Institute of Innovation & Technology arbeitet die Gruppe von Hartmut Schmeck im Projekt **Hybrid Energy Grid Management (HEGRID)** an einem Energieträgerübergreifenden Rahmen für ein interoperables, integriertes Energiemanagement für Strom-, Gas- und Wärmenetze. Die Basis von HEGRID ist das Open Energy Management Framework, eine Plattform, die vielfältige Ansätze für das Stromnetz der Zukunft unterstützt.

Andreas Oberweis und sein Team arbeiten an einem **IT-basierten Feedbacksystem (IFS)**, das organisatorische und prozessorientierte Qualitätsverbesserung der Studien-, Lehr- und Lernbedingungen auf vier Ebenen adressiert: Anwendung, Feedback-Kanäle, Dokumentation, Administration. Moderne Mobiltechnologie und Social Networks wie Twitter, Facebook werden einbezogen. IFS ist Teil des Gesamtprojektes „KIT-Lehre hoch Forschung“, das vom Bundesministerium für Bildung und Forschung finanziert wird.

Im Projekt **Cross-lingual Cross-media Knowledge Extraction (xLiMe)** wagt sich die Forschungsgruppe von Rudi Studer im Verbund mit europäischen Forschungspartnern aus Universitäten und Unternehmen an das große Problem der medien- und sprachübergreifenden Auswertung von öffentlich zugänglichen Informationsquellen. Maschinenlesbare Textdokumente, Videos, Audios, multimediale Nachrichtenstreams usw., die in verschiedenen Sprachen im Internet veröffentlicht werden, sollen in nahezu Echtzeit analysiert und mit Wissensbasen verknüpft werden. Auf diese Weise will man in Zukunft z.B. Fernsehkonsumenten Zusatzinformationen zu einer Nachrichtensendung liefern, die sie gerade sehen. Die Wissenschaftler verbinden dafür Methoden der Verarbeitung von natürlich-sprachlichen Texten, maschinellem Lernen und semantischen Technologien. Das zweite beispielhafte Vorhaben der Gruppe, gefördert als Konsortialprojekt im EU FP7-Programm im Bereich Transport, ist nicht weniger anspruchsvoll. **iVision** will für Piloten ein virtuelles Cockpit bauen, das ihnen wie in einem Simulator alle zum Fliegen notwendigen Informationen vor Augen hält; und zwar exakt zu der jeweiligen Flugsituation, in der sich die Maschine gerade befindet.

<http://www.aifb.kit.edu/web/Projekte>

## 5 Forschungsgruppen

am Institut AIFB untersuchen und entwickeln Methoden, Konzepte und Verfahren, um grundlegende Forschungserkenntnisse der Informatik für die professionelle Gestaltung von computergestützten technischen Systemen zu nutzen.

Zentrales Thema der Forschungsgruppe **Effiziente Algorithmen** ist die Entwicklung von Methoden für den wirtschaftlichen Einsatz moderner IT-Infrastrukturen. Besonderes Interesse gilt der Beherrschbarkeit und effizienten Nutzung vielfältig vernetzter, adaptiver Systeme mit der Fähigkeit zur Selbstorganisation. Schwerpunkte der Forschung liegen aktuell auf der Nutzung und Weiterentwicklung von Konzepten des Organic Computing für die effektive Gestaltung zukünftiger dezentraler Energiesysteme sowie dem Aufbau der neuen Disziplin Energieinformatik. Daneben werden naturinspirierte Optimierungsverfahren, insbesondere für multikriterielle und dynamisch veränderliche Problemstellungen, weiterentwickelt.

Die Forschungsgruppe **Betriebliche Informationssysteme** beschäftigt sich mit Methoden, Vorgehensmodellen und Werkzeugen für das Business Process Engineering und das serviceorientierte Workflow-Management, unterschiedlichen Non-Standard-Anwendungen betrieblicher Informationssysteme sowie der strategischen Informatik-Planung und -Organisation in Unternehmen. Grundlage für die Konzeption und Implementierung innovativer betrieblicher Informationssysteme sind servicebasierte IT-Infrastrukturen, in denen Geschäftsprozesse kollaborativ, flexibel, sicher und zuverlässig ausgeführt werden können.

Die Forschungsgruppe **Wissensmanagement** arbeitet an Methoden zur Unterstützung von Wissensmanagement in Unternehmen, an der Entwicklung von Methoden und Werkzeugen zur Verwirklichung der Idee des Semantic Web sowie an Fragestellungen zu Service Science. Eine zentrale Rolle spielen Informations- und Applikationsintegration, die automatische Ableitung von neuem Wissen sowie der intelligente Zugriff auf das vorhandene Wissen. Grundlegende methodische Basis ist die semantische Repräsentation von Wissen durch Ontologien und Metadaten sowie maschinelle Lernverfahren. Das Themenfeld Big Data gewinnt zunehmend an Bedeutung. Hier geht es vor allem um die Verwaltung und Analyse von heterogenen Datenquellen. Ein weiteres wichtiges Forschungsthema ist die Kopplung von Linked Data und REST Services zu so genannten Linked Services im Kontext von Webarchitekturen.

Der zentrale Arbeitsschwerpunkt der Forschungsgruppe **Komplexitätsmanagement** ist die Untersuchung struktureller Ursachen für das Auftreten von hoher Komplexität und die daraus abgeleitete Entwicklung effizienter algorithmischer Methoden zur Lösung komplexer Probleme. Auf der Basis graphentheoretischer, analytischer und logischer Ansätze werden Beiträge zum besseren Verständnis komplexer Systeme und komplexer Probleme geleistet.

Die Forschungsgruppe **eOrganisation** untersuchte das junge Forschungsfeld „Cloud Service Engineering“. Aufbauend auf Grundlagen des Software Service Engineering, Web-Engineering und der verteilten Systeme, erforschte und entwickelte die Gruppe Modelle, Methoden und Middleware-Technologien für Cloud Computing und Service-oriented Computing. Einen Arbeitsschwerpunkt bildete die Gestaltung und Bewertung von Qualitätseigenschaften von Cloud-Diensten sowie Modellierungs- und Engineering-Werkzeugen für die Dienste-Komposition.

<http://www.aifb.kit.edu/web/Forschungsgruppen>

## 59 Mitarbeiterinnen und Mitarbeiter

arbeiten am Institut AIFB. Sie stehen den Professoren und Studierenden im Lehrbetrieb zur Seite, bearbeiten die Forschungsprojekte bzw. halten den Forscherinnen und Forschern den Rücken frei von Verwaltungsarbeiten und sorgen für eine leistungsfähige Infrastruktur. 1 Honorarprofessor und 7 Lehrbeauftragte bereichern das Lehrangebot zusätzlich. Zu den Forschungsgruppen gehören ferner 5 Stipendiaten, 20 Doktoranden, die am FZI Forschungszentrum Informatik arbeiten, sowie 13 externe Doktoranden, die in anderen Unternehmen beschäftigt sind. 4 Auszubildende absolvieren ihre Berufsausbildung am Institut AIFB, und, je nach Semester, sind zwischen 90 und 110 studentische und wissenschaftliche Hilfskräfte in den Projekten bzw. als Tutoren für Lehrveranstaltungen aktiv. Insgesamt sind das 210 Personen – manchmal ein paar mehr, manchmal ein paar weniger.

<http://www.aifb.kit.edu/web/Personen>

## 38 Vorlesungen

mit bis zu 700 Zuhörern sowie 27 Seminare und Praktika wurden vom Institut angeboten. 1185 Rechnerpraktika wurden durchgeführt. Diese bei den Studierenden äußerst beliebte Lehrform ist sehr aufwändig, da in Kleingruppen mit in der Regel zwei bis vier Teilnehmern gearbeitet wird. Sie werden dabei von uns intensiv betreut.

[http://www.aifb.kit.edu/web/Lehrangebot\\_des\\_Instituts\\_AIFB](http://www.aifb.kit.edu/web/Lehrangebot_des_Instituts_AIFB)

## 4309 Prüfungen

wurden im Wintersemester 2012/13 und im Sommersemester 2013 am Institut AIFB abgenommen.

## 132 Studien- und Abschlussarbeiten

legten Studierende vor. 35 Diplomarbeiten, 38 Masterarbeiten, 56 Bachelorarbeiten und drei Studienarbeiten wurden geschrieben.

<http://www.aifb.kit.edu/web/Abschlussarbeiten>

## 7 Zertifikate für gute Lehre

wurden an Mitarbeiter und Lehrbeauftragte des Instituts nach der Evaluation durch Studierende erteilt, die von der Fakultät für Wirtschaftswissenschaften regelmäßig durchgeführt wird. Im Sommersemester 2013 erhielten **Professor Dr. Stefan Tai** für die Pflichtvorlesung „Angewandte Informatik II: Informatiksysteme für den eCommerce“, **Dr. Daniel Sommer** für die Vorlesung „Datenbanksysteme“, **Professor Dr. Stefan Klink** für die Vorlesung „Dokumentenmanagement und Groupware-systeme“ sowie **Dr. Roland Schätzle** für die Vorlesung „Management von Informatik-Projekten“ die Auszeichnung. Im Wintersemester 2013/14 zeichneten die Studierenden die Übungen von **Timm Caporale** (Datenbanksysteme und XML), sowie von **Kaibin Bao** und **Fabian Rigoll** (Algorithms for Internet Applications) und von **Andreas Drescher** (Modellierung von Geschäftsprozessen) mit Zertifikaten für gute Lehre aus.

## 103 Publikationen

wurden im Jahr 2013 aus dem Institut AIFB veröffentlicht. 2 Bücher, 2 Buchbeiträge und 12 Veröffentlichungen in Zeitschriften stammen von Angehörigen des Instituts AIFB. Über 60 Beiträge in Tagungsbänden sowie zahlreiche weitere Publikationen zeugen von der aktiven Teilnahme am wissenschaftlichen Austausch.

<http://www.aifb.kit.edu/web/Veroeffentlichungen>

## 1 Start-Up

haben die beiden Alumni **Dr. Daniel Herzig** und **Dr. Günter Ladwig** aus der Forschungsgruppe Wissensmanagement von **Professor Dr. Rudi Studer** gegründet. Für ihr neues Unternehmen **SearchHouse** gewannen Sie das begehrte EXIST Stipendium. SearchHouse bietet eine neue Suchtechnologie an, die Eingaben semantisch interpretiert. Sie kann auf strukturierte Daten aus beliebigen Datenbanken angewendet werden.

<http://searchhaus.net/>

## 18 Dissertationen

wurden im Berichtszeitraum 2013/14 am Institut betreut. Die frischgebackenen Doktorinnen und Doktoren, ihre Promotionsthemen und ihre Betreuer:

**Birger Becker:** „Interaktives Gebäude-Energiemanagement“ (Hartmut Schmeck)

**David Bermbach:** „Benchmarking Eventually Consistent Distributed Storage Systems“ (Stefan Tai)

**Stefanie Betz:** „Petri-Netzbasierte Modellierung und Analyse von Risikoaspekten in Geschäftsprozessen“ (Andreas Oberweis)

**Eugenie Giesbrecht:** „Distributional Tensor Space Model of Natural Language Semantics“ (Rudi Studer)

**Julia Hoxha:** „Cross-domain Recommendations based on semantically-enhanced User Web Behavior“ (Rudi Studer)

**Martin Junghans:** „Methods for Efficient and Accurate Discovery of Services“ (Rudi Studer)

**Björn Keuter:** „Bidirektionale Abbildung zwischen Geschäftsprozessmodellen und IT-Kommunikationssystemen“ (Andreas Oberweis)

**Lukas König:** „Towards Complex Behavior in Evolutionary Robotics“ (Hartmut Schmeck)

**Yongtao Ma:** „Effective Instance Matching for Heterogeneous Structured Data“ (Rudi Studer)

**Anees ul Mehdi:** „Epistemic Reasoning in OWL 2 DL“ (Rudi Studer)

**Sabrina Merkel:** „Building Evacuation with Mobile Devices“ (Hartmut Schmeck)

**Marc Mültin:** „Das Elektrofahrzeug als flexibler Verbraucher und Energiespeicher im Smart Home“ (Hartmut Schmeck)

**Carolin Mund:** „Identification of Emerging Scientific Topics in Bibliometric Databases“ (Rudi Studer)

**Roland Stühmer:** „Web-oriented Event Processing“ (Rudi Studer)

**Andreas Josef Wagner:** „Rank-aware, Approximate Query Processing on the Semantic Web“ (Rudi Studer)

**Johannes Winter:** „Intelligentes Energiemanagement in kleinen und mittleren Unternehmen“ (Hartmut Schmeck)

**John Erik Wittern:** „Modeling and Selection of Software Service Variants“ (Stefan Tai)

**Yongchun Xu:** „Efficient Context-aware Real-time Processing of Personal Data“ (Rudi Studer)

**aifb**

[www.aifb.kit.edu](http://www.aifb.kit.edu)

Institut für  
Angewandte Informatik  
und Formale  
Beschreibungsverfahren

**Herausgeber:**  
Andreas Oberweis  
Hartmut Schmeck  
Detlef Seese  
Rudi Studer

September 2014

**aik**

[www.aik-ev.de](http://www.aik-ev.de)



**AIFB. Das Informatik-Institut der KIT-Fakultät für  
Wirtschaftswissenschaften.**

Partnerschaften zwischen Wissenschaft, Wirtschaft, Politik  
und Verwaltung fördern den Technologie- und Wissenstransfer.  
Auf allen Seiten!

**Wollen Sie nicht auch mit uns zusammenarbeiten?**

Prof. Oberweis	<a href="mailto:andreas.oberweis@kit.edu">andreas.oberweis@kit.edu</a>
Prof. Schmeck	<a href="mailto:hartmut.schmeck@kit.edu">hartmut.schmeck@kit.edu</a>
Prof. Seese	<a href="mailto:detlef.seese@kit.edu">detlef.seese@kit.edu</a>
Prof. Stucky	<a href="mailto:wolffried.stucky@kit.edu">wolffried.stucky@kit.edu</a>
Prof. Studer	<a href="mailto:rudi.studer@kit.edu">rudi.studer@kit.edu</a>

**Institut AIFB  
KIT-Campus Süd  
76128 Karlsruhe**

**Redaktionelle Bearbeitung:**  
Dr. Daniel Sommer, Institut AIFB  
[daniel.sommer@kit.edu](mailto:daniel.sommer@kit.edu)  
Vera Münch, Hildesheim  
[vera-muench@kabelmail.de](mailto:vera-muench@kabelmail.de)

**Gestaltung:**  
Studio Quitta, München  
[www.studio-quitta.de](http://www.studio-quitta.de)

**Druck:**  
Heinrich Fischer  
Rheinische Druckerei GmbH, Worms