TECHNISCHE UNIVERSITÄT ILMENAU
Institut für Praktische Informatik und Medieninformatik
Fakultät für Informatik und Automatisierung
Fachgebiet Datenbanken und Informationssysteme

Abschlussbericht

Zuwendungsempfänger:	Technische Universität Ilmenau Fachgebiet Datenbanken und Informationssysteme
Förderkennzeichen:	03WKBD2B
Vorhabensbeschreibung:	Customer Bautronic System TP Facility Intelligence – Informationsintegration und –analyse für das Gebäudemanagement
Laufzeit des Vorhabens:	01.04.2007 – 31.03.2010
Berichtszeitraum:	01.04.2007 – 31.03.2010

Ilmenau, den 30. September 2010 Prof. Dr.-Ing. habil. Kai-Uwe Sattler

Inhaltsverzeichnis

1 Kurze Darstellung **3**

1.1 Aufgabenstellung . . . 3

1.1.1 Einordnung in das Gesamtvorhaben . . . 3

1.2 Verbundprojekt Facility Intelligence . . . 4

1.2.1 Voraussetzungen . . . 5

1.3 Planung und Ablauf des Vorhabens . . . 6

1.4 Wissenschaftlicher und technischer Stand zu Beginn des Vorhabens . . . 8

1.5 Kooperationen zur Umsetzung der Projektziele . . . 10

2 Wissenschaftlich-technische Ergebnisse **11**

2.1 Arbeitspaket 1: Datenbankmodellierung und -realisierung . . . 11

2.1.1 Konzeptuelle Datenbankmodellierung und Entwicklung eines Datenbankschemas . . . 11

2.2 Arbeitspaket 2: Datenintegration . . . 12

2.2.1 Entwicklung von Synopsentechniken zur Zeitreihenanalyse bei limitiertem Speichervolumen . . . 15

2.2.2 Änderungserkennung . . . 16

2.3 Arbeitspaket 3: Datenanalyse und Simulation . . . 18

2.3.1 Aufdecken von bisher unbekannten Zusammenhängen . . . 18

2.3.2 Trendanalyse - prediktive Techniken zur Überprüfung von Hypothesen 19

2.3.3 Erweiterung der Verfahren zur Änderungserkennung zu inkrementeller Arbeitsweise, Güteüberwachung und Materialisierung von Zwischenergebnissen . . . 20

2.3.4 Konzeption von Softwaremodulen und Umsetzung als Datenfluss Operatoren in die FI-Plattform . . . 20

2.4 Arbeitspaket 4: Rückführung von gewonnenen Informationen . . . 21

2.5 Arbeitspaket 5: Ableitung Steuerungsmodell Controller . . . 21

2.5.1 Ableitung der Steuerungsempfehlung an Gebäudeserver . . . 21

2.6 Arbeitspaket 6: Reporting/Visualisierung . . . 21

2.6.1 Bereitstellung der Analyseergebnisse . . . 21

3 Nutzen und Verwertung **23**

3.1 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse . . . 23

3.1.1 AnduIN . . . 23

3.1.2 FolloVizer . . . 24

3.2 Erfolgte Veröffentlichungen . . . 24

3.3 Vortragsreisen und Messen . . . 30

1 Kurze Darstellung

1.1 Aufgabenstellung

1.1.1 Einordnung in das Gesamtvorhaben

Mit dem gemeinsamen Leitgedanken „Gebäude werden (er)lebensWert!“ haben die Bündnispartner des regionalen Wachstumskerns „Customer Bautronic System“ (CBS) ein Gebäudeautomationssystem entwickelt, das durch die permanente Nutzerintegration und –interaktion geprägt ist. Es umfasst die Phasen der Planung, Erstellung sowie des Betriebs und der Nutzung von Gebäuden (siehe Abbildung 1).

Die Bündnispartner der Arbeitsgruppe 1 analysieren die individuellen Anforderungen von Nutzern und Betreibern in der Planungsphase, ermitteln daraus Nutzergruppen und bilden die jeweiligen Nutzerpräferenzen auf notwendige Technologien sowie deren operativen und strategischen Betrieb im Gebäude ab. Sie entwickeln somit die Basis für die Nutzerintegration in CBS und liefern entsprechende Vorgaben an die nachgelagerten Systeme, z.B. die strategischen Nutzeranforderungen an das Facility Intelligence System.

Aufgabe der Arbeitsgruppe 2 (AG 2) mit den Partnern NT Neue Technologie AG (NT.AG) und Technische Universität Ilmenau, Fachgebiet Datenbanken und Informationssysteme, war es, ein zentrales Analyse- und Optimierungsmodell zu entwickeln, das gebäudeübergreifend Gebäude-, Anlagen-, Nutzer-, und Prozessdaten während des Betriebs der Anlagen erfasst, analysiert und dabei Optimierungsmöglichkeiten entdeckt (siehe Abbildung 1, Vorhaben markiert innerhalb der roten Ellipse). Facility Intelligence (FI) wird erreicht, indem in den enormen gebäudebezogenen Datenmengen durch Strukturierung und Verknüpfung sowie Verdichtung bislang unbekannte und nützliche Zusammenhänge aufgedeckt, Trends erkannt und Prognosen erstellt werden. Die Ergebnisse werden schließlich den Nutzern und Betreibern zur Verfügung gestellt (FI-Portal), zur Optimierung in der Gebäudesteuerung genutzt und explizit in der Datenbasis für die weitere Arbeit zurückgewonnen.

Für den operativen Betrieb der Anlagen im Gebäude, dessen Erfassung sowie zur Erfassung der situativen Nutzerdaten wurden ein drahtloses Multisensorsystem, Schnittstellen zur Gebäudetechnik, der CBS-Gebäudeserver (Controller) mit dem Vorgangssystem als Datendrehscheibe und dem Parametrisierungsverfahren für die Gebäudesteuerung und konfigurierbare Mensch-Maschine-Schnittstellen durch die Bündnispartner der Arbeitsgruppe 3 entwickelt. Dabei werden die operativen Nutzeranforderungen sowie die Optimierungsvorschläge umgesetzt.

Das hochkomplexe Gesamtsystem wird durch eine gemeinsame Verwertungsgesellschaft

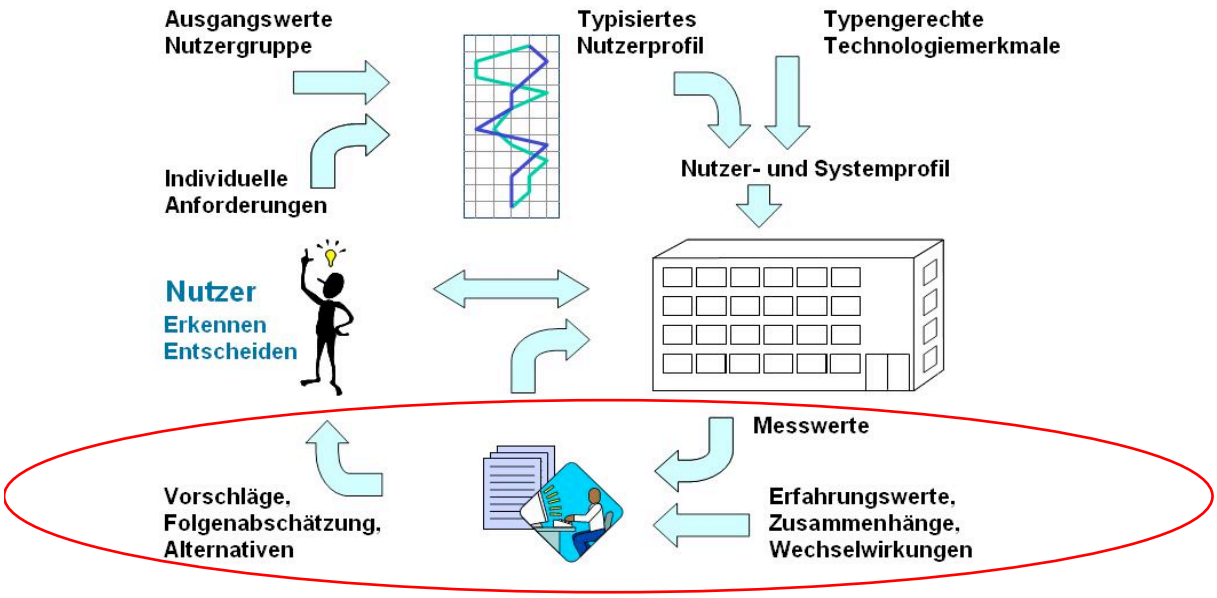

Abbildung 1: Verortung des Verbundprojekts Facility Intelligence im Customer Bautronic System (CBS)

vertrieben. Dabei bietet die Arbeitsgruppe 4 das Blended Learning-Qualifizierungstool für CBS an. Dieses ist vor allem auf die Zielgruppe der Planer, Architekten und Betreiber von Gebäuden ausgerichtet und beinhaltet Qualifizierungsangebote zur Problemlösung und zur Steigerung der Produktakzeptanz.

1.2 Verbundprojekt Facility Intelligence

Der Leistungsbeitrag der Arbeitsgruppe 2 zum innovativen regionalen Wachstumskern CBS besteht in der Informationsintegration und -analyse für das Gebäudemanagement (TU Ilmenau) und in der Entwicklung eines Wissensmanagementsystems zur Schaffung von Facility Intelligence (NT.AG). Die gemeinsame Zielstellung ist es, die nutzerzentrierte Gesamtoptimierung der Betriebsprozesse im Gebäude zu unterstützen und Entscheidungsfähigkeit zu schaffen. Unter den Maßgaben der Nutzwerterhöhung, Kostenreduzierung und Nachhaltigkeit sollen

- Daten strukturiert, verdichtet und verknüpft,
- bislang unbekannte und nützliche Zusammenhänge zur Optimierung des Gebäudebetriebs detektiert,
- Modelle zur Entscheidungsunterstützung und Kostensenkung gebildet sowie
- Nutzern und Betreibern Informationen (Reporting, Vorhersagen) innerhalb eines FI-Portals zur Verfügung gestellt werden.

Grundlage des Wissensmanagements bilden Analysen von Datenströmen aus dem Gebäude und von langfristig gespeicherten regelungsrelevanten Daten in einem mehrdimensionalen Modell. Die Analyseergebnisse sollen automatisch in den Gebäudebetrieb eingebracht (Schnittstelle zur Arbeitsgruppe 3) und/oder dem Nutzer in Form von Optimierungsvorschlägen und Informationen zur Verfügung gestellt werden können. Der Nutzer kann damit seine Entscheidungen auf eine umfangreiche Wissensbasis stützen. Die jeweilige Ausprägung des Facility Intelligence Systems soll abhängig von den strategischen Nutzeranforderungen gestaltet werden können (Schnittstelle zur Arbeitsgruppe 1).

1.2.1 Voraussetzungen

Das Verbundprojekt „Wissensmanagement für die Gebäudewirtschaft zur Schaffung von Facility Intelligence“ wird seit April 2007 im Rahmen des innovativen regionalen Wachstumskerns Customer Bautronic Services durch das BMBF gefördert und zeichnet sich durch die enge Zusammenarbeit des Unternehmens NT Neue Technologie AG mit dem Fachgebiet Datenbanken und Informationssysteme der TU Ilmenau aus. Die gemeinsame Forschungs- und Entwicklungsarbeit fußt auf den jeweiligen Kernkompetenzen der Partner.

Die NT Neue Technologie AG arbeitet im Bereich Informations- und Wissensmanagement. Dabei reicht das Portfolio von der Softwareentwicklung bis hin zu Managed IT- und Hosting Services.

Über die Unternehmensapplikationen für Dokumenten-, Kontakt- und Contentmanagement hinaus, werden gleichfalls Individualanwendungen nach Kundenwünschen entwickelt. Erfahrene Berater und Projektmanager helfen bei der Analyse und Strukturierung von Anforderungen und erarbeiten mit dem Kunden gemeinsam Strategien zur Optimierung von Geschäftsprozessen und deren Unterstützung durch IT-Lösungen. Ein besonderer Schwerpunkt liegt auf in der Thematik Kundenbeziehungsmanagement/CRM.

Weiterhin ermöglicht das Rechenzentrum der NT.AG nicht nur Web-, Mail- und Applicationhosting, sondern auch Serverhosting und Colocation. Für den mobilen Zugriff auf Groupwarefunktionen wer-den in Zusammenarbeit mit T-Mobile und Vodafone Blackberry-Enterpriseserver gehostet. Darüber hinaus wird sowohl den Unternehmenskunden als auch öffentlichen Institutionen professioneller Soft-ware-Support über den NT.AG-Helpdesk geboten. Schließlich steht das Unternehmen gleichfalls als kompetenter und zuverlässiger Outsourcing-Partner zur Verfügung.

Seit ihrer Gründung ist die NT.AG bestrebt, über das Tagesgeschäft hinaus Know-how und Kapazitäten in die Entwicklung zu investieren. Damit soll die eigene Wettbewerbsfähigkeit gestärkt und mithin ein gesundes Wachstum ermöglicht werden, mit dauerhaften

Arbeitsplätzen in der Region. So konnte das Team des Unternehmens in den letzten 10 Jahren plattformübergreifend einen wachsenden Erfahrungsschatz vor allem im Bereich Wissensmanagementlösungen aufbauen.

Das Fachgebiet Datenbanken und Informationssysteme gehört zur Fakultät für Informatik und Automatisierung der TU Ilmenau und wird von Prof. Dr.-Ing. habil. Kai-Uwe Sattler geleitet. Zu den Forschungsschwerpunkten zählen neben Fragestellungen von Datenbankmanagementsystemen (autonomes Management und Tuning, verteilte Verarbeitung, Architekturen) insbesondere Data Analytics- und Data-Mining-Techniken für Data Warehousing und Sensordaten sowie Fragen der Informationsintegration und semantischer Technologien. Die Arbeiten des Fachgebiets werden von der DFG, dem Freistaat Thüringen, dem BMBF sowie der Industrie (SAP AG, Ingres Germany GmbH) gefördert.

Speziell für das Gebiet der Energieoptimierung von Gebäuden konnte die Siemens Building Technology GmbH & Co OHG als F&E-Fremddienstleister gewonnen werden. Ihre Aufgabe bestand in der Unterstützung der Partner des Verbundprojekts in den Bereichen der Analyse, Simulation und Optimierung des Wärmeenergieverbrauchs von Räumen und Gebäuden.

1.3 Planung und Ablauf des Vorhabens

Die Arbeitsziele wurden in die die nachfolgend skizzierten sieben Arbeitspakete gegliedert, die innerhalb des genehmigten Arbeits- und Zeitplans durchgeführt wurden:

1. Datenbankmodellierung und –realisierung:
    - Spezifikation und Modellierung aller für das Gebäude- und Energiemanagement sowie die nutzerzentrierte Optimierung relevanten Gebäude-, Nutzer- und Prozessdaten sowie von externem Expertenwissen
    - Realisierung der modellierten Datenstruktur
2. Datenintegration
    - Konzipierung geeigneter Schnittstellen für die Datenaufnahme in die Datenbank
    - Realisierung von Import- und Transformationsoperatoren
3. Datenanalyse und Simulation
    - Erforschung und Entwicklung von Techniken zum Data Mining (Assoziationsanalyse, Clusteranalyse), zur energetischen Simulation, zur Trendanalyse sowie zur Plausibilitätsprüfung von sich kontinuierlich ändernden Prozessdaten

(Sensor- und Aktordaten) in Verbindung mit den strukturiert vorliegenden Stammdaten (Gebäude- und Nutzerdaten) sowie mit Hintergrundwissen (Expertenwissen)

- Ausprägung für den Kontext der Gebäudewirtschaft mit dem Ziel der langfristigen, nutzerzentrierten Optimierung des Betriebs von Gebäuden

4. Rückführung der gewonnenen Informationen
   - Übernahme der Analyse- und Simulationsergebnisse aus als neues entscheidungsrelevantes Wissen
   - Änderungspropagation an den Gebäudeserver (Schnittstelle AG 3)
5. Ableitung Steuerungsmodell Controller
   - Erarbeitung der Form der Zielvorgaben/Optimierungsvorschläge an den Gebäudeserver gemeinsam mit den Bündnispartnern der Arbeitsgruppe 3
   - Konzeptionierung und Realisierung entsprechender Abläufe und Szenarien innerhalb von zeit- und/oder ereignisgesteuerten Operatoren
6. Reporting/Visualisierung
   - Schaffung einer webbasierten Nutzerschnittstelle (FI-Portal) zur Darstellung der Erkenntnisse und Analyseergebnisse zu Verbrauchs- und Betriebsdaten
   - Export von Berichten
   - Ausgestaltung gemäß des durch die Arbeitsgruppe 1 ermittelten strategischen Nutzerprofils
7. Systemausprägung für Prototypen
   - Erstellung eines Funktionsmusters des Wissensmanagements inkl. webbasiertem FI-Portal
   - Zusammenführen der Einzelkomponenten zu einem Labormuster und Durchführen von Last- und Funktionstests des CBS-Gesamtsystems in Zusammenarbeit mit allen anderen Arbeitsgruppen
   - Möglichkeit der Ausprägung für öffentliche Gebäude, Seniorenheime sowie Bürogebäude

Die Inhalte und Ergebnisse der einzelnen Arbeitspakete werden im nächsten Kapitel ausführlich dargestellt. Der Arbeits- und Zeitplan konnte während der dreijährigen Förderphase weitestgehend eingehalten werden. Es mussten lediglich Verschiebungen von Arbeitspaketen aufgrund inhaltlicher Abhängigkeiten oder verspäteter Zuarbeiten in Kauf

genommen werden. So konnte z.B. das Simulationsmodell erst später als geplant fertiggestellt werden. Um nicht im Gesamtplan in Verzug zu geraten, wurden andere Arbeitspakete zeitlich vorgezogen.

Die Planung beinhaltete zudem die folgenden Meilensteine:

- Meilenstein 1 (03.2008): Modul Datenintegration und –aufbereitung
- Meilenstein 2 (03.2009): Datenanalyse und Prozessoptimierung
- Meilenstein 3 (09.2009): Facility Intelligence Plattform
- Meilenstein 4 (02.2010): Prototypen

Die erzielten Ergebnisse der Arbeitsgruppe zu den Meilensteinterminen sowie die weitere strategische Ausrichtung des Wachstumskerns wurden auf den jeweiligen Projekt- und Strategieworkshops dargestellt. Zudem fanden monatliche Treffen der Projektkoordinatoren der vier Verbundprojekte unter Federführung des INIT e.V. statt. Inhalt dieser Treffen waren jeweils die Einhaltung der Arbeits- und Zeitpläne, notwendige Absprachen zwischen den Arbeitsgruppen sowie die Vorbereitung der Workshops und der öffentlichkeitswirksamen Veranstaltungen.

Als Basis für den Informationsaustausch und die Koordinierung der Verbundprojekte konnte eine IBM Quickplace-Umgebung für alle Beteiligte des Wachstumskerns durch die NT.AG realisiert und administriert werden. Projektkoordinator des Verbundprojekts 2 war in der Phase der Antragstellung und zu Beginn der Forschungs- und Entwicklungsarbeiten Herr Matthias Maier, der die NT.AG leider zum 31.08.2007 verließ. Im Anschluss übernahm Herr Torsten Mucke vorübergehend die Koordination der Arbeitsgruppe. Ab 01.01.2009 konnte Herr Matthias Markgraf, der bis dahin Erfahrungen im Rahmen der CBS-Gesamtprojektkoordination im INIT-Projektbüro sammeln konnte, bei der NT.AG eingestellt und als Projektkoordinator der Arbeitsgruppe 2 gewonnen werden.

1.4 Wissenschaftlicher und technischer Stand zu Beginn des Vorhabens

Der Markt für Gebäudesteuerung bot zu Beginn des Vorhabens keine multifunktionalen, gewerkeübergreifenden und gleichzeitig die individuelle Nutzerpräferenz in den Mittelpunkt stellenden Steuerungssysteme an. Vielmehr war der Markt geprägt von einzelzellenbezogenen Entscheidungssystemen (Smartboxes), wie „Hometronic“ von Honeywell oder „Assisto“ von Techem. Damit kann der Nutzer z.B. Einfluss auf die durch Zeitprofile geregelte Raumtemperatur in einzelnen Wohn- oder Gewerbeeinheiten nehmen. Darüber hinaus vereinfachen sie die Bedienung von einigen anderen Gebäudesteuerungssystemen,

z.B. der Rollläden, an einem zentralen Bedienpunkt. Allerdings werden hierbei dem Nutzer keineswegs alle notwendigen Informationen für eine Gesamtoptimierung des individuellen Nutzwertes, z.B. unter Einbezug von Kostenauswirkungen, unter Berücksichtigung von Interdependenzen zwischen den einzelnen gebäudetechnischen Anlagen und den Nutzwert beeinflussenden Systemen geboten.

Bei der Vielfalt zu regelnder Systeme und unter Berücksichtigung externer Einflussfaktoren und Informationen zu Kosten und Nachhaltigkeit ist der normale Nutzer, dessen Arbeit oder Kerngeschäft nicht im Bereich der Gebäudeautomatisierung liegt, grundsätzlich überfordert damit, selbst eine nutzeroptimierte Gesamtstrategie herbeizuführen. Oft ist zu beobachten, dass Einstellungen an den Anlagen einmal bei der Inbetriebnahme bzw. des Bezugs des Gebäudes vorgenommen werden, danach aber nicht mehr an eine geänderte Nutzung oder sonstige neue Anforderungen angepasst werden. Das liegt z.T. auch in der Vielzahl der möglichen Einstellungen, der einhergehenden Unübersichtlichkeit und schlechten Bedienbarkeit begründet. Da Kostenzusammenhänge und Parameter in den einzelnen Unterlagen nur unzureichend erklärt werden, ist eine umfassende Einarbeitung in die Thematik oder fachmännisches Personal für optimierende Anpassungen und zum Ausschöpfen der vollen Funktionalität notwendig. Abhängigkeiten zwischen verschiedenen Systemen werden in Handbüchern i.A. nicht betrachtet.

Die beschriebene allgemeine Situation birgt ein enormes Potential für Einsparungen sowie für Optimierungen bzgl. der Nutzwerte. Über das Gesamtsystem CBS soll dieses Potential aufgedeckt und genutzt werden. Durch eine nutzerzentrierte Planung und den nutzerzentrierten Anlagenbetrieb soll zudem die Zufriedenheit der Nutzer gesteigert werden. Das CBS-Wissensmanagement mit gebäudeübergreifenden Analyse- und Optimierungsmodellen stellt dabei eine vollkommen neue und innovative Komponente für Gebäudeautomationssysteme dar.

Speziell die automatische Detektion von neuen und nützlichen Zusammenhängen aus einem fortlaufenden Datenstrom, wie er z.B. von Sensoren im Gebäude erzeugt wird, ist ein aktuelles Problem der Forschung und Entwicklung. Entsprechend waren für das beantragte Vorhaben auch Arbeiten im Bereich Frequent Pattern Mining relevant. So sollen anhand von Assoziationsanalysen Korrelationen innerhalb von Transaktionen, also z.B. zwischen gleichzeitig auftretenden Sensormesswerten, entdeckt werden. Die ersten Analysemethoden in diesem Bereich waren Apriori-basiert [4], neuere Ansätze sind FP-Growth-basiert [2]. Da der Datenbestand fortlaufend um neue Transaktionen wächst und eine komplette Neuberechnung zeit- und ressourcenaufwändig ist, beschäftigt sich die Forschung nun auch mit inkrementellen Data-Mining-Verfahren, die bisher berechnete Ergebnisse mit neu eingehenden Daten aktualisieren. Gerade in diesem Bereich besteht noch immer ein hoher Bedarf – nicht nur in der Grundlagenforschung, sondern auch in

der anwendungsorientierten F&E. Das Verbundprojekt „Facility Intelligence“ des Wachstumskerns CBS, speziell das Teilprojekt der TU Ilmenau, soll hier einen Beitrag in Form von Online-Mining-Algorithmen auf Sensordatenströmen liefern.

Die Nutzung von Wissensmanagement hat sich für unterschiedlichste Problemstellungen überall dort als entscheidender und neuer erfolgreicher Lösungsansatz bewiesen, wo auf Basis vielfältiger sich dynamisch ändernder Datenmassen Entscheidungen im Sinne von Gesamtoptimierungen getroffen werden müssen. In der Gebäudewirtschaft wurde dieser Lösungsansatz allerdings noch nicht umfassend umgesetzt. Die verschiedenen Data-Mining-Verfahren, die im kaufmännischen Bereich Zusammenhänge und Parameter für Business Intelligence liefern und darstellen, wurden für Gebäude und deren Nutzer ebenfalls noch nicht weiterentwickelt oder adaptiert. Der Innovationssprung des CBS Systems besteht in der Kombination von Nutzerpräferenzmodell und technische Gebäudesteuerung unter der Regie eines übergreifenden Analyse- und Optimierungsmodells, welches das Nutzerverhalten antizipiert, den Nutzer über Zusammenhänge, Kosten und Behaglichkeit informiert und die Gebäudesteuerung entsprechend der Analyseergebnisse und der Präferenzen der Nutzer anpasst.

1.5 Kooperationen zur Umsetzung der Projektziele

Die einzelnen Pakete wurden von der NT Neue Technologie AG und dem Fachgebiet Datenbanken und Informationssysteme der TU Ilmenau bearbeitet. Während der gesamten Projektlaufzeit gestaltete sich die enge Zusammenarbeit zwischen beiden Partnern als äußerst angenehm und fruchtbar. Die Siemens Building Technologies GmbH & Co OHG fungierte als F&E Fremddienstleistungspartner der NT.AG. Wir danken an dieser Stelle auch den Bündnispartnern der anderen Arbeitsgruppen für die gute Zusammenarbeit, vor allem bei der Konzeptionierung der Schnittstellen zwischen den Verbundprojekten sowie beim Aufbau des gemeinsamen Labormusters.

2 Wissenschaftlich-technische Ergebnisse

In den weiteren Abschnitten sollen die wesentlichen, durch die TU Ilmenau beigebrachten, wissenschaftlich-technischen Ergebnisse aufgezeigt werden. Die Präsentation folgt dabei der durch die Arbeitspakete vorgegebenen Struktur.

2.1 Arbeitspaket 1: Datenbankmodellierung und -realisierung

2.1.1 Konzeptuelle Datenbankmodellierung und Entwicklung eines Datenbankschemas

Ziel des ersten Arbeitspaketes war die Modellierung und Realisierung (Datenbankentwurf) des Data Warehouses zur Datenspeicherung in der FI-Plattform. Die Umsetzung des Data Warehouses erfolgte als relationale Datenbank im Microsoft SQL Server.

Aus Performance-Gründen wurde während der Entwurfsphase des Data Warehouses ein kombiniertes Star/Snowflake-Schema entwickelt. Die Faktenrelation dient dabei sowohl zur Aufnahme der kontinuierlich produzierten Sensordaten als auch zur Aufnahme der durch die Aktoren generierten Daten. Eine Differenzierung der Daten erfolgt erst zum Analysezeitpunkt.

Im Kern weist das entwickelte Schema die drei hierarchischen Dimensionen Zeit, Sensor und Raum auf. Die Integration von Gebäudedaten in das Data Warehouse erfolgt auf auf Basis der VDI-Richtlinie 6021 (Datenaustausch für die thermische Lastberechnung von Gebäuden), der Ergebnisse aus INNOSEG[1] sowie der Energieeinsparverordnung (EnEV). Mit diesem Vorgehen konnte vorab sichergestellt werden, dass für die spätere Analyse alle relevanten Gebäudeinformationen enthalten sind. Die Modellierung orientiert sich an den Industry Foundation Classes (IFC), da es sich hierbei um einen offenen Standard für die Kommunikation im Bauwesen handelt, der bereits in der Planungssoftware vieler führender Hersteller umgesetzt wurde. Die Vorarbeit in Form der gesetzlichen Normen und Regularien hierfür lieferte die NT.AG. Die Integration der Gebäude erfolgte als dimensionale Attribute über der Klasse Raum.

Die Nutzerdaten wurden entsprechend der Vorlage von Arbeitsgruppe 1 in das Data Warehouse integriert.

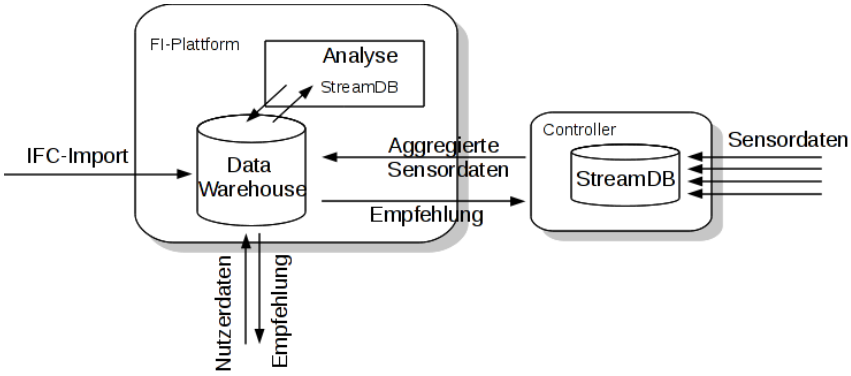

Abbildung 2: Datenintegration FI-Plattform

2.2 Arbeitspaket 2: Datenintegration

Im Folgenden soll auf die Integration der Daten aus den verschiedenen Quellen eingegangen werden. Prinzipiell müssen die folgenden drei Formen von Datenquellen unterschieden werden (siehe Abbildung 2):

- *Gebäudedaten:* Diese Daten beschreiben alle wesentlichen statischen Eigenschaften eines Gebäudes (Fläche pro Raum, Wandstärke, Fenster, Türen, Wärmedurchgangskoeffizienten etc.). Gebäudedaten müssen im Allgemeinen lediglich einmal übernommen werden.
- *Nutzerdaten:* Die vom Nutzer spezifizierten Eigenschaften, welche das Gebäude aufweisen muss (zum Beispiel die Raumtemperatur). Nutzerdaten werden zur Gebäudeinstanziierung erfasst, können über die Zeit aber regelmäßig geändert werden.
- *Prozessdaten:* Die von den in das Gebäude eingebrachten Sensoren erfassten Messwerte (zum Beispiel Temperatur, Luftfeuchtigkeit, CO2-Gehalt) und die aktuellen Regelstellungen der Aktoren. Die aktuellen Aktoreinstellungen werden für das Ableiten von Regelmengen benötigt. Beide Datenquellen, Sensoren und Aktoren, liefern kontinuierlich Daten an das Data Warehouse (je nach Sensor/Aktor treffen die Daten im Abstand von wenigen Sekunden oder mehreren Stunden ein).

Gebäudedaten Im Allgemeinen finden der Import von Gebäudedaten einmalig statt. Bereits beim Erzeugen des Data Warehouse wurde auf eine Integration des IFC-Standard geachtet. Zum einfachen Datenaustausch existiert mit IFCxml ein entsprechendes XML-basiertes Format. Für die Integration der Gebäudedaten wurde eine entsprechende IFC-Komponente als Operator für den Microsoft-SQL-Server-Integration-Service entwickelt. Mit Hilfe des Integrationswerkzeuges können von den Architekten exportierte Gebäude-

[1]http://www.fh-erfurt.de/fhe/ibit/projekte0/innoseg/

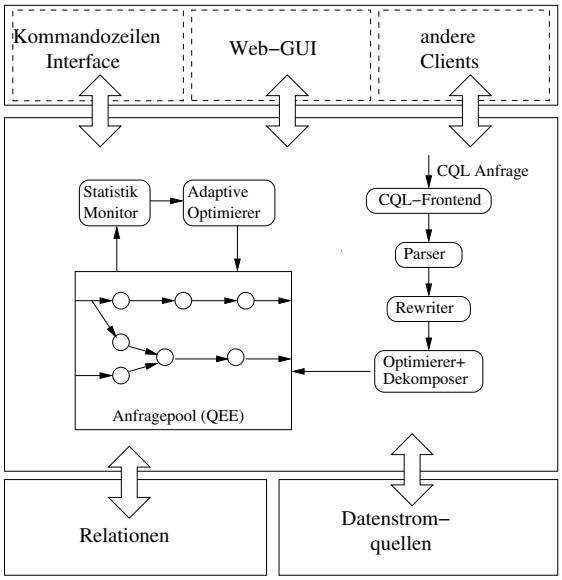

Abbildung 3: Grundlegende Architektur von StreamDB

informationen ohne zusätzliche manuelle Eingriffe direkt in die FI-Plattform importiert werden.

Nutzerdaten Die Erfassung der Nutzerinformationen erfolgt durch das von Arbeitsgruppe 3 entwickelte NMI (Nutzer-Maschinen-Interface). Diese Informationen werden der FI-Plattform durch das gemeinsam von den Arbeitsgruppen 2 und 3 entwickelte XML-Austauschformat bereitgestellt. Die abschließende Integration war Aufgabe der NT.AG und wurde wiederum in Form neues Moduls für den Microsoft-SQL-Server-Integration-Service umgesetzt.

Prozessdaten Zu den wesentlichen Herausforderungen des Projektes zählt die u.U. große Menge an Daten von den in den Gebäuden untergebrachten Sensoren und Aktoren. Je nach Anzahl an Sensoren/Aktoren und deren Messhäufigkeit, können pro Stunden mehrere hundert Megabyte oder sogar Gigabyte anfallen, welche in die FI-Plattform importiert werden müssen. Da ein Import der vollständigen Daten nur bedingt möglich bzw. sinnvoll ist (die Menge an verwalteten Daten im Data Warehouse ist entscheidend für die Dauer der Analysen) wurde ein System zu Datenaggregation entwickelt. Mit Hilfe des entwickelten Systems *StreamDB* lässt sich die Datenmenge signifikant reduzieren. Die aggregierten und in das Data Warehouse integrierten Daten repräsentieren dabei die wesentlichen statistischen Eigenschaften der von den Sensoren/Aktoren erfassten Daten. Im Weiteren soll eine Beschreibung von StreamDB und deren Komponenten im Detail gegeben werden.

StreamDB wurde als Datenstrom-Managementsystem (DSMS) umgesetzt. Das System

bietet die Möglichkeit einer vollständig deklarativen Aufgabenbeschreibung. Es ist somit auf einfache Weise möglich neue Verarbeitungsprozesse zu beschreiben. Wie die definierten Ziele erreicht werden, wird vollautomatisch durch StreamDB bestimmt. Das entwickelte System weist dabei alle, von traditionellen Datenbank-Managementsystemen bekannten, Verarbeitungsschritte auf. Vom System entgegengenommene Anfragen werden analysiert, transformierte, optimiert und anschließend zur Ausführung gebracht. Abbildung 3 zeigt den groben Aufbau des entwickelten Datenstrom-Managementsystems (DSMS) und alle alle wesentlichen Komponenten.

Für die Datenaggregation wurden generische Operationen zum Filtern von Daten, sowie verschiedene statistische Aggregationsverfahren integriert (zum Beispiel min, max, avg, sum, ...). Weiterhin wurden Operatoren zum Verknüpfen von Datenströmen und zum Gruppieren bezüglich verschiedener Prädikate implementiert.

Zur Optimierung der Verarbeitungsgeschwindigkeit wurde die StreamDB komplett in der objektorientierten Sprache C++ entwickelt. Die Verarbeitung der Daten durch die StreamDB findet vorrangig auf dem von der Arbeitsgruppe 3 entwickelten Controller statt (siehe Abbildung 2), da die Stromverarbeitung innerhalb der FI-Plattform technisch sehr aufwändig und kostenintensiv ist. Zudem wurden mit diesem Schritt die auf Veränderungen im Datenstrom um die sonst vorhandene Zeitverzögerung beschleunigt. Für die Kommunikation der FI-Plattform mit StreamDB wurde ein entsprechender Client entwickelt und als Operator in den Microsoft-SQL-Server-Integration-Service integriert.

Da die Daten im Data Warehouse schnell sehr umfangreich werden können ist ein vollständiges Laden in den Hauptspeicher mit anschließendem Verarbeiten oftmals nicht möglich. Ein einfacher Lösungsansatz ist das sukzessive Lesen und auswerten der Daten. Da es sich bei den so erzeugten Daten wiederum um Datenströme handelt kann an dieser Stelle u.a. StreamDB zur Analyse eingesetzt werden (siehe Abbildung 2).

Für den Datenaustausch zwischen StreamDB (auf dem Controller) und der FI-Plattform wurden zwei Modelle integriert:

- *Synchrone Kommunikation*: StreamDB sendet die Ergebnisse direkt nach deren Fertigstellung an die FI-Plattform. Existiert zu diesem Zeitpunkt keine Verbindung zwischen beiden Systemen, dann gehen diese Ergebnisse verloren.
- *Asynchrone Kommunikation*: Ergebnisse werden von StreamDB in vordefinierten Ringpuffern zwischengespeichert. Die FI-Plattform kann anschließend zu einem beliebigen Zeitpunkt die vollständige im Puffer enthaltene Ergebnismenge anfordern. Um einen Überlauf des Puffers zu verhindern, werden alte Ergebnisse kontinuierlich überschrieben, d.h. zu alte Daten gehen an dieser Stelle verloren.

Im Rahmen des Projektes kann eine permanente Verbindung zwischen FI-Plattform und dem Controller nicht in jedem Fall sichergestellt werden. Zudem ist eine dauerhafte Internetverbindung mit zusätzlichen Kosten verbunden. Da die Auswertung der Daten durch die FI-Plattform nur zum Beispiel 1 mal pro Tag oder Woche stattfindet bietet sich der Ansatz der asynchronen Kommunikation an.

Da zu Projektbeginn keine Ausreichende Datenbasis vorhanden war wurde zunächst ein entsprechender Datengenerator entwickelt. Dieser erlaubte sowohl das Erzeugen von synthetischen Daten mit definierten Eigenschaften, als auch das Laden von den später zur Verfügung stehenden Messdaten. Für eine möglichst praxisnahe Entwicklung wurden an der TU Ilmenau Sensoren angeschafft, welche über den Datengenerator ausgelesen werden könnten.

2.2.1 Entwicklung von Synopsentechniken zur Zeitreihenanalyse bei limitiertem Speichervolumen

Da die komplette Speicherung von Datenströmen aus verschiedenen Gründen (beschränkter Speicherplatz bzw. Ergebnisse in Echtzeit) nicht möglich ist, wurden in den letzten Jahren verschiedene Lösungsansätze entwickelt. Prinzipiell können hierbei zwei Herangehensweisen unterschieden werden:

- Verfahren, welche repräsentative Werte aus dem Datenstrom für die Analysen einsetzen und
- Verfahren, welche die kompletten Daten in Form von Datenzusammenfassungen (*Synopsen*) aufbewahren.

Verfahren der ersten Gruppe reduzieren lediglich die Anzahl der Elemente des Datenstromes (zum Beispiel durch geeignetes verwerfen), wobei die Originalwerte erhalten bleiben. Im Gegensatz dazu versucht man bei Datenzusammenfassungen lediglich die Eigenschaften der Daten des Stromes zu erfassen. Die entstehenden Synopsen selbst enthalten dabei zumeist keine Elemente des Originalstromes mehr. Die Zusammenfassungen sind das Ergebnis der Anwendung von Funktionen auf den Datenstrom. Typische Funktionen sind einfache Aggregatfunktionen (zum Beispiel Summe, Durchschnitt, Minimum, Maximum, Variance) oder komplexere Funktionen wie zum Beispiel Histogramme oder Wavelets. Aufgrund des zusammenfassenden Charakters von Synopsen, können Anfragen nur noch approximativ beantwortet werden.

Im Rahmen des Projektes wurden verschiedene Untersuchungen im Bereich der Synopsen durchgeführt. So beschäftigte sich u.a. die Diplomarbeit „Synopsen-Sharing zur effizienten Anfrageverarbeitung in Datenströmen“ mit der Analyse und Bewertung bestehender Syn-

opsen. Existierende Synopsenlösungen wurden in dieser Arbeit auf ihre Eignung zur Wiederverwendung in mehrerer Anfragen untersucht. Zudem wurden mit dem Dyadic Grid ein neuartiges Histogramm zur effizienten Speicherung von Häufigkeitswerten präsentiert.

Weiterhin wurden im Rahmen der entwickelten komplexen Operatoren Konzepte existierender Synopsen erweitert (zum Beispiel im Fall der Erkennung von Burst) oder wie beim Verfahren der Erkennung von Mustern über quantitativen Attributen weitestgehend neu entwickelt.

2.2.2 Änderungserkennung

Neben der Integration von Daten in die FI-Plattform war die Erkennung von Änderungen der Datencharakteristiken einer der Entwicklungschwerpunkte in Arbeitspaket 2 für die TU Ilmenau. Bei der Erkennung von Änderungen wurde sich auf das Finden von möglichen Fehlersituation beschränkt. In diesem Zusammenhang konnten die nachstehenden Probleme als für das Projekt wesentlich bestimmt werden:

- Ausreißererkennung
- Identifikation von Anormalem Verhalten (Burst-Erkenung)
- Erkennung von Sensorinaktivität (im Fall von Sensorausfällen)

Im weiteren soll auf die Lösung beider Probleme im Detail eingegangen werden.

Ausreißererkennug Beim Erfassen von Daten durch die Sensoren und dem Transfer zum Controller bzw. zu StreamDB kann es zu fehlerhaften Werten kommen, welche im weiteren Einfluss auf die Analyse in StreamDB oder die FI-Plattform haben können. Mögliche Fehlerquellen sind zum Beispiel falsch erfasste Werte (durch defekte Sensorhardware) oder ein fehlerhafter Datenstransfer. Die Erkennung solcher kurzzeitigen Fehler wird unter dem Begriff der Ausreißererkennung zusammengefasst. In der Literatur lassen sich hierzu verschiedene Detektionsverfahren finden.

Prinzipiell lassen sich diese in zwei Klassen einteilen. Verfahren, welche versuchen anhand von vorkategorisierten Trainingsdaten Charakteristika zu erlernen und statistische Verfahren, welche die Eigenschaften der Sensordaten, wie zum Beispiel deren Distanz oder Dichte in Relation zueinander setzen. Da ein vorkategorisieren von Trainingsdaten nicht auf einfache Weise möglich war wurde für die Ausreißererkennug ein einfacher distanzbasierter Ansatz gewählt und implementiert. Für eine möglichst frühzeitige Erkennung wurde der entsprechende Operator in StreamDB integriert. Der entwickelte Operator stellt dabei gleichzeitig alternative (geschätzte) Sensorwerte zur Verfügung.

Burst-Erkennung Im Gegensatz zu einmalig auftretenden fehlerhaften Werten handelt es sich bei Bursts um länger dauernde andersartige Ereignisse, deren zeitliche Länge vor der eigentlichen Analyse i.A. nicht bekannt ist. Typische Beispiel für Burst sind Feuer oder Überschwemmung.

Zu den größten Herausforderungen im Zusammenhang mit der Identifikation von Burst zählt deren möglichst frühzeitige und korrekte Erkennung. Werden Bursts zu spät oder gar nicht identifiziert, dann können eventuell notwendiger Maßnahmen nicht mehr rechtzeitig eingeleitet werden. Kommt es wiederum zu einem häufigen Auslösen von *falsch-positiven* Anomalien, dann kann dies in einer hohen Systemlast und u.U. in teuren Folge-Maßnahmen resultieren. Im Bereich der Gebäudeautomation kann die Burst-Erkennung zum Beispiel zur Brand- bzw. Einbruchsdetektion eingesetzt werden. Beim feststellen entsprechender Ereignisse werden Nachrichten an die FI-Plattform gesendet, welches anschließend Gegenmaßnahmen wie zum Beispiel die Benachrichtigung von Feuerwehr oder Polizei einleiten kann. Ein nicht Erkennen des eingetretenen Ereignisses hat ebenso negative Folgeerscheinung wie ein falsches Auslösen. Weiterhin galt es zu verhindern, dass die im Data Warehouse enthaltenen Daten durch kurzzeitig auftretende Ereignisse verfälscht und somit die anschließenden Analysen erschwert werden.

Die Erkennung von Bursts geschieht üblicherweise unter Verwendung von fensterbasierten Verfahren. Bei diesen wird geprüft inwieweit die im aktuellen Fenster vorliegenden Daten von den Daten früherer Fenster abweichen. Für die Bestimmung der Abweichung wird sowohl auf statistische Verfahren, als auch auf Verfahren aus den Bereichen Signal Analyse oder künstliche Intelligenz zurückgegriffen. Die Detektion von Bursts über (nahezu) beliebige Fenstergrößen wird als *elastic burst detection* bezeichnet. In [6, 3] präsentierten die Autoren einen Ansatz zur effizienten Detektion von Bursts mit variablen Zeitfenstern. Mit der als *shiftet aggregation tree* bezeichneten Datenstruktur wurde in [5] eine effiziente, einfach zu implementierende Lösung zur Burst Detektion über Datenströmen vorgestellt. Aufgrund der verwendeten Schwellwertstrategie ist der vorgestellte Ansatz nur für stationäre Daten geeignet deren zeitlicher Verlauf keine Trends oder Perioden aufweist. Die im Projekt anfallenden Sensor/Aktordaten weisen allerdings sowohl saisonale Schwankungen als auch Trends auf. Betrachtet man zum Beispiel Außensensoren für Temperatur oder Luftfeuchtigkeit, so unterliegen diese saisonalen Schwankungen wie zum Beispiel Tag und Nacht oder den Jahreszeiten. Entsprechend führen Trends bzw. periodische auftretende Schwankungen zu falsch-positiven Alarmen.

Der im Rahmen des Projektes entwickelte Ansatz berücksichtigt sowohl stationäre als auch nicht-stationäre (trend- bzw. saisonbehaftete) Daten. Hierzu wurde das bekannte Verfahren der elastive Burst Erkennung mit Vorhersagetechniken für Zeitreihen kombiniert. Für eine frühzeitige Erkennung eventuell abnormalen Verhaltens wurde das entwi-

ckelte Verfahren als Operator in das DSMS StreamDB integriert. Die Benachrichtigung der FI-Plattform erfolgt, wie oben beschrieben, umgehend beim erkennen von fehlerhaftem Verhalten über ein synchrone Kommunikation.

Erkennung von Sensorausfällen Im Projektverlauf hat sich gezeigt, dass die von den Sensoren/Aktoren bereitgestellten Daten weniger zuverlässig sind, als zunächst erwartet. Da die von den Sensoren/Aktoren u.a. für die Aktivierung von Prozessen notwendig sind (Event-basierte Verarbeitung) hat der Ausfall von neuen Daten einen signifikanten Einfluss insbesondere auf die Vorverarbeitung durch StreamDB.

Zum Zweck der Erkennung von Sensor/Aktorausfällen wurde ein entsprechender Temporaler Operator in StreamDB. Dieser überwacht den eingehenden Datenstrom auf Aktivitäten von Sensoren/Aktoren und benachrichtigt die FI-Plattform, wenn ein Sensor/Aktor über einen vorab definierten Zeitraum keine Daten an das System gesendet hat.

2.3 Arbeitspaket 3: Datenanalyse und Simulation

2.3.1 Aufdecken von bisher unbekannten Zusammenhängen

Korrelationsanalyse Das Finden von Korrelationen wurde in die beiden Teilbereiche Frequent-Pattern-Mining und Association-Rule-Mining unterteilt. Beide Verfahren wurden als Operatoren in das DSMS StreamDB integriert.

Zu den wesentlichen Herausforderung zählte die Entwicklung eines Verfahrens, welches sowohl für die Detektion von häufigen Mustern (Frequent-Pattern) über quantitativen Attribute als auch für die Verarbeitung von Datenströmen geeignet war. Existierende Lösungsverfahren, wie zum Beispiel das in [4] präsentierte Apriori-basierte Verfahren oder der in [1] präsentierte FP-Stream lösen lediglich eines der Probleme. Ein Verfahren für das Finden häufiger Mustern über quantitativen Attributen in Datenströmen waren zum Zeitpunkt der Bearebeitung nicht bekannt, so dass mit dem $FP^2 - Stream$ ein entsprechendes Verfahren neu entwickelt wurde.

Häufige Muster werden beim entwickelten Verfahren in einem Präfixbaum (ähnlich dem FP-Tree von Han et al. [2]) verwaltet. Statt diskreter Werte wurden in die Knoten des Präfixbaumes Intervalle eingebettet, welche einer kontinuierlichen Verfeinerung unterliegen. Durch entsprechende Verfeinerungs- und Rekombinationsphasen wird die Synopse kontinuierlich an die sich ändernden Daten der Sensoren/Aktoren angepasst. Für eine Optimale Verfeinerung der Intervalle wurden zudem Techniken aus dem Bereich der multidimensionalen Histogramme integriert.

Das entwickelte Frequent-Pattern-Mining-Verfahren wurde als Operator in das DSMS StreamDB integriert, so dass die Detektion der häufigen Muster nahezu in Echtzeit über den nichtaggregierten Daten stattfinden kann. Die so erzeugten häufigen Muster dienen als Grundlage für das finden von entsprechenden Regelmengen. Zu diesem Zweck wurde zusätzlich ein entsprechendes Association-Rule-Mining-Verfahren in das DSMS StreamDB integriert. Dieses erhält die zuvor identifizierten häufigen Muster und deren Häufigkeitswerte als Eingabe und erzeugt aus diesen eine kontinuierlich angepasste Regelmenge, welche an die FI-Plattform weitergeleitet wird.

Das Frequent-Pattern-Mining-Verfahren wurde zunächst in Form eines nicht-inkrementellen Verfahrens entwickelt, dessen Ausgabemenge zu jedem definierten Zeitpunkt die vollständige Menge aller gefundenen häufigen Muster darstellte. Im Weiteren wurde eine effiziente inkrementelle Variante des Verfahrens entwickelt, welche sich auf die Ausgabe von Veränderungen in den häufigen Itemsets beschränkte.

Raum-Belegungsprofile Weiterhin wurde im Rahmen von Arbeitspaket 3 ein Verfahren zur Identifikation von Raum-Belegungsprofilen entwickelt. Ziel des Verfahrens war das Finden von Regeln, welche definieren, wann ein Raum belegt bzw. wann dieser Raum frei ist. Basierend auf diesen Regeln können anschließend zum Beispiel Zeitprogramme für Temperaturwerte oder die Beleuchtung bestimmt werden.

Das entwickelt Verfahren basiert wiederum auf dem FP-Stream-Algorithmus nach Giannella et. al [1]. Eine der wesentlichen Herausforderungen bei dieser Problemstellung war es die von den Sensoren zur Verfügung gestellte Zeitdimension in geeigneter Weise für das Verfahren vorzubereiten. Mit dem entwickelten Verfahren werden die von den Sensoren gelieferten Zeitstempel in Items verschiedener Granularität zerlegt. Zusammen mit den von den Sensoren erfassten Messwerten dienten diese als Eingabe für das verwendete FP-Stream-Verfahren.

Das entwickelte Verfahren wurde als Operator in den Microsoft-SQL-Server-Integration-Service. Die Evaluierung des Verfahrens erfolgte aus Basis von Messwerten eines Bewegungssensors an der TU Ilmenau.

2.3.2 Trendanalyse - prediktive Techniken zur Überprüfung von Hypothesen

Ziel dieses Arbeitspakets war es, Verfahren und Techniken zu entwickeln und zu integrieren, die die Vorhersage von zukünftigem Verhalten ermöglichen. Ein Teilbereich dieser Aufgabe ist die Erkennung von Trends und Perioden. In diesem Zusammenhang bedeutet ein (linearer) Trend zum Beispiel, dass sich die Sensorwerte kontinuierlich erhöhen. Andererseits können Sensordaten jedoch auch periodische Schwankungen aufweisen, die von

geeigneten Verfahren erkannt und entsprechend behandelt werden müssen

Zur Erkennung von Trends und Perioden mittels Zeitreihenvorhersagen wurde die exponentielle Glättung (1.-3. Ordnung) als geeignet befunden. Ein wichtiges durch dieses Verfahren erfülltes Kriterium ist die One-Pass-Eigenschaft, d.h. jeder Datenwert wird nur ein einziges Mal gelesen, was eine wesentliche Voraussetzung zur Echtzeitverarbeitung der Daten darstellt. Prinzipiell sind auch andere ARMA-Modelle (AutoRegressive-Moving Average) denkbar. Aufgrund der Einfachheit in der Implementierung und der Berechnung bei weiterhin bekannt guten Ergebnissen wurde dieses Verfahren gewählt.

Das Verfahren wurde in Form eines entsprechenden Operators in das DSMS StreamDB integriert. Somit konnte eine Änderungserkennung direkt auf den Original-Daten und in Echtzeit vorgenommen werden. Der Vergleich der Vorhersagewerte mit den tatsächlich erfassten Werten erfolgt über Filteroperatoren mit entsprechenden Prädikaten. Überschreitet die Differenz zwischen zuvor vorhergesagten Werte und gemessenen Werten einen vorab definierten Schwellwert (zum Beispiel die Varianz oder die Standardabweichung), dann passieren die gemessenen Werte diesen Filter und dienen zugleich der FI-Plattform als Benachrichtigung. Werte, welche den definierten Schwellwerten genügen lösen keine Reaktion am Filteroperator aus.

2.3.3 Erweiterung der Verfahren zur Änderungserkennung zu inkrementeller Arbeitsweise, Güteüberwachung und Materialisierung von Zwischenergebnissen

Im Rahmen der Entwicklung der Data-Warehouse-Komponente in der FI-Plattform hat sich gezeigt, dass Aufgrund des zu erwartenden hohen Datenvolumens der zu erfassenden und zu analysierenden Prozessdaten dem physischen Datenbank-Entwurf und Tuning eine besondere Aufmerksamkeit zu widmen ist. Ein viel versprechender Ansatz hierfür ist die Präaggregation und Materialisierung häufig ausgeführter Anfragen für Reports etc., wobei jedoch die Auswahl der Präaggregationen eine Herausforderung darstellt. Zu diesem Zweck wurde ein Ansatz entwickelt, welcher ausgehend vom Anfrage-Workload eine optimale Auswahl von Materialisierungen ermittelt. Diese Auswahl wird dynamisch während des Betriebs durch kontinuierliches Monitoring anpasst.

2.3.4 Konzeption von Softwaremodulen und Umsetzung als Datenfluss Operatoren in die FI-Plattform

Alle oben beschriebenen Verfahren wurden entweder als Teil des DSMS StreamDB implementiert und evaluiert oder als Teil Paket für den MSIS zur Verfügung gestellt. Für

die Kommunikation des DSMS mit dem Vorgangssystem und der FI-Plattform wurden entsprechende Anbindungen geschaffen. Aufgrund der Anbindung der StreamDB an die FI-Plattform stehen alle von der TU Ilmenau entwickelten Verfahren der FI-Plattform in vollem Umfang zur Verfügung.

2.4 Arbeitspaket 4: Rückführung von gewonnenen Informationen

Für das Arbeitspaket 4 waren keine Aufgaben für die TU Ilmenau vorgesehen.

2.5 Arbeitspaket 5: Ableitung Steuerungsmodell Controller

2.5.1 Ableitung der Steuerungsempfehlung an Gebäudeserver

Neben dem im Arbeitspaket 3 entwickelten Verfahren zur Erkennung häufiger Muster (FP^2-Stream) in Datenströmen wurde zusätzlich ein Operator geschaffen werden, welcher die Muster aus der Datenstruktur auf geeignete Weise ausließt und eine entsprechende Regelmenge ableitet (Assoziation Rule Mining). Hierbei war zu beachten, dass nicht alle häufigen Muster in entsprechenden Regeln ausgedrückt werden sollte, da die Menge der gefundenen Muster einen sehr großen Umfang aufweisen kann. Das entwickelte Verfahren zum Ableiten der Assoziationsregeln wurde als Operator in das DSMS StreamDB integriert. Die Ergebnisse des Operators werden in Form von Regeln an die FI-Plattform weitergeleitet, welche aus diesen Steuerungsempfehlungen Gebäudeserver entwickelt.

2.6 Arbeitspaket 6: Reporting/Visualisierung

2.6.1 Bereitstellung der Analyseergebnisse

Die von dem DSMS StreamDB ermittelten Ergebnisse werden entweder in Form einfacher Textnachrichten oder als XML-Nachrichten an die FI-Plattform gesendet. Desweiteren wurde das StreamDB Interface so konzipiert, dass sich prinzipiell beliebig viele Clients gleichzeitig mit dem Server verbinden und Datenströme abonnieren können. Mit diesem Interface ist es möglich, sowohl Daten an die FI-Plattform zu senden, als auch besondere Ereignisse getrennt zu betrachten und u.U. sofortige Maßnahmen einzuleiten (z.B. beim erkennen von Feuer einen Alarm auszulösen). Eine Ausschließliche Auswertung über die FI-Plattform könnte hierbei zu träge sein.

Zusätzlich wurde eine Web-basierte Benutzerschnittstelle zu StreamDB geschaffen, welche sowohl die einfache Erstellung neuer komplexer Anfragen erlaubt (siehe Abbildung 4),

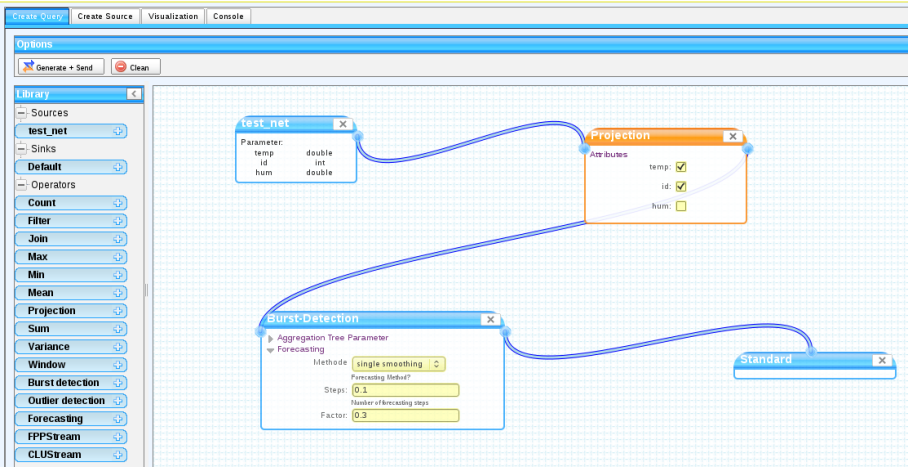

Abbildung 4: Grafische Anfragebeschreibung

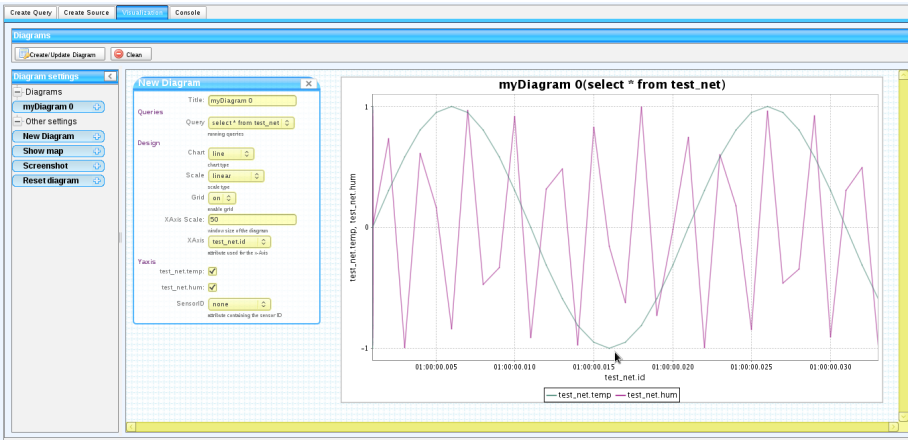

Abbildung 5: Viszualisierung von Anfrageergebnissen

als auch die Visualisierung aktuell laufender Datenströme und Berechnungsergebnissen in Echtzeit ermöglicht (siehe Abbildung 5). Die Oberfläche ist vorrangig als technisches Informationsoberfläche zu sehen, welches der Erfassung von Datenstromcharakteristika in Echtzeit dient. Zur einfachen Integration in die FI-Plattform wurde die Visualisierungskomponente als Web-Applikation entwickelt und basiert auf gängigen Standards (Java, Javascript, etc).

3 Nutzen und Verwertung

3.1 Voraussichtlicher Nutzen und Verwertbarkeit der Ergebnisse

Das im Rahmen des Projektes entwickelte DSMS StreamDB bildet eine sehr gute Grundlage für Projekte im Bereich der Datenstromanalyse und des Complex Event Processing. Mit den Projekten AnduIN und FolloVizer wurden bereits erste Projekte abgeleitet, welche im Folgenden kurz beschrieben werden sollen.

3.1.1 AnduIN

Bei Sensoren handelt es sich oftmals um mehr als einfache Datenquellen. Drahtlose Sensornetzwerke, wie sie zum Beispiel beim Projektpartner IMMS zum Einsatz kommen, sind üblicherweise selbstorganisierende Verbunde von Sensorknoten. Jeder dieser Knoten ist für sich genommen ein Computer mit stark beschränkten Ressourcen. Im Regelfall verfügen diese über eine Sensorik zur Erfassung von Realweltdaten (zum Beispiel zum Messen von Luftfeuchtigkeit, Temperatur, etc.), einer leistungsschwachen CPU, einer oftmals stark beschränkten Speichermenge für Daten und Programme, einer Batterie zur Energieversorgung und einem drahtlosen, zumeist funkbasierten Kommunikationsmodul.

Das im Rahmen einer Dissertation entstandene Projekt *AnduIN* erweitert das StreamDB Projekt in Richtung dieser drahtlosen Netzwerke mit dem Ziel einer energieeffizienten Datenvorverarbeitung auf den einzelnen Sensorknoten. Die Verarbeitung der Daten ist dabei zweigeteilt. Soweit möglich und sinnvoll findet eine Verarbeitung der Daten auf den Sensorknoten innerhalb des Netzwerkes statt. Zwischen- bzw. Endergebnisse werden an eine zentrale Instanz weitergeleitet, welche eine finale Auswertung der Daten vornimmt und die Ergebnisse eventuell abhängigen Anwendungen bereitstellt. Die zentrale Instanz von *AnduIN* basiert in weiten Teilen auf dem StreamDB Projekt, wobei wesentlichen Teile der Anfrageoptimierung von StreamDB um die Möglichkeit der Integration von Sensorknoten in den Analyseprozess angepasst wurden.

Bisher befindet sich das *AnduIN* Projekt in einem prototypischen Stadium. Das Sensornetwerk beschränkt sich auf eine kleine Zahl an Sensorknoten. Für die Zukunft ist es geplant, praxisnahe Studien mit diesem Projekt durchzuführen. So ist gegenwärtig zum Beispiel der Einsatz bei einem Projektpartner auf einem mittelgroßen Sensornetzwerk mit mehreren Hundert Sensorknoten angedacht.

Abbildung 6: Anwendung FolloVizer

3.1.2 FolloVizer

Soziale Netze wie zum Beispiel Facebook, StudiVZ oder Twitter gehören Heute zum Alltag vieler Internetteilnehmer. Neben persönlichen Websites bieten diese Plattformen die Möglichkeit des einfachen Kontaktes mit anderen interessierten Nutzern. Ein, in den letzten Jahren zusehends aufstrebendes soziales Netzwerk, ist Twitter. Bei Twitter können Benutzer Kurznachrichten, sogenannte „Tweets“, in ihrem persönlichen Micro-Blog hinterlassen.

Eine Besonderheit von Twitter ist die Möglichkeit der „Tweets“-Abfrage in Echtzeit - Twitter kann somit als eine „Sensorquelle“ angesehen werden, welche von einem DSMS verarbeitet und analysiert werden kann. Zu diesem Zweck wurde StreamDB um Techniken zur Text- und Tag-Verarbeitung erweitert. Weiterhin wurden Operationen zu Analyse von Geo-Informationen integriert.

Basieren auf diesen Erweiterungen wurde mit dem FolloVizer-Projekt eine Anwendung geschaffen, welche dem Nutzer die Möglichkeit der Überwachung und Echzeitauswertung von Twitter-Datenströmen erlaubt (Abbildung fig:foll). Zusätzlich bietet die entwickelte Anwendung die Möglichkeit aus den Nachrichten extrahierte Informationen mit zusätzlichen Informationen aus externen Datenquellen, wie zum Beispiel Wikipedia oder Facebook, zu verknüpfen.

3.2 Erfolgte Veröffentlichungen

Im Rahmen des Projektes entstanden die folgenden Publikationen:

- *K. Hose, M. Karnstedt, D. Klan, K. Sattler, J. Quasebarth: Incremental Mining*

for Facility Management. KDML 2007: Knowledge Discovery, Data Mining, and Machine Learning, Sept. 2007, Halle (Saale), S. 183-190.

Modern buildings are equipped with high-tech systems that take care of several fundamental aspects, e.g., air-conditioning, heating and water supply. The requirements posed on facility management by such buildings are challenging. Modern techniques implement adaptive control systems to achieve this, in which decisions are preferably based on the results of (multiple correlated) mining tasks on recently gathered sensor data. In this work, we discuss the general relationship between such control systems and the underlying mining tasks. We exemplary choose change detection in the context of pattern analysis as a representative, because this mining task involves general requirements known from stream processing like the need for incremental algorithms, but also poses specific challenges like in-time detection. We present three concrete approaches for this and an according evaluation.

- *K. Hose, J. Quasebarth, K. Sattler.: Interoperability in Peer Data Management Systems. VLDB International Workshop on Semantic Data and Service Integration (SDSI), Wien, Sept. 2007.*

  Interoperability plays an important role for a variety of applications. One of them are Peer Data Management Systems, where autonomous data sources (peers) interact with each other based on semantic mappings between their schemas. The building blocks that enable interoperability and thus the main challenges in such systems are mapping representation, query rewriting, and efficient query processing. While most approaches regard these aspects in separate this paper presents a comprehensive study of the interactions between these blocks. Our considerations try to provide a holistic view on semantic interoperability in distributed environments such as PDMS. We discuss techniques for distributed query processing and rewriting that consider high-level query operators such as top-N and skyline. Furthermore, we discuss how to increase efficiency by applying routing indexes and relaxation of result completeness/correctness.

- *K. Hose, C. Lemke, K. Sattler, D. Zinn: A Relaxed but not Necessarily Constrained Way from the Top to the Sky. CoopIS 2007, Vilamoura, Algarve, Portugal, Nov. 28-30, 2007, S. 399-407.*

  As P2P systems are a very popular approach to connect a possibly large number of peers, efficient query processing plays an important role. Appropriate strategies have to take the characteristics of these systems into account. Due to the possibly large number of peers, extensive flooding is not possible. The application of routing indexes is a commonly used technique to avoid flooding. Promising techniques

to further reduce execution costs are query operators such as top-N and skyline, constraints, and the relaxation of exactness and/or completeness. In this paper, we propose strategies that take all these aspects into account. The choice is left to the user if and to what extent he is willing to relax exactness or apply constraints. We provide a thorough evaluation that uses two types of distributed data summaries as examples for routing indexes.

- *D. Klan, K. Sattler, K. Hose, M. Karnstedt: Decentralized Management of Replication Objects in Massively Distributed Systems, EDBT 2008: Workshop on Data Management in Peer-to-peer systems (DaMaP'08), Nantes, France, March 2008, pp. 19-26*

  Data replication is a central technique to increase availability and performance of distributed systems. While offering many advantages it also requires more effort for ensuring data consistency in case of updates. In the research literature various approaches for replication management in distributed databases have been presented, but they are mostly limited either in scalability or in the consistency guarantees they provide. On the other hand, P2P systems usually provide replication support but ignore the update problem. In this paper we present a new approach for managing replicated data in wide area distributed networks. Our solution is orthogonal to the underlying infrastructure and managed in a decentralized manner. It guarantees single-master consistency and allows updates at any node of the system by combining traditional replication techniques with ideas known from P2P systems.

- *K. Hose, D. Klan, M. Marx, K. Sattler: When is it Time to Rethink the Aggregate Configuration of Your OLAP Server?, VLDB Demonstrations Track, Auckland, New Zealand, Aug. 2008.*

  OLAP servers based on relational backends typically exploit materialized aggregate tables to improve response times of complex analytical queries. One of the key problems in this context is the view selection problem: choosing the optimal set of aggregation tables (called configuration) for a given workload. In this paper, we present a system that continuously monitors the workload and raises a quantified alert, when a better configuration is available. We address the tasks of query monitoring and view selection at the OLAP level instead of the SQL level, which simplifies the containment checks as well as rewriting and in this way helps to reduce the complexity of the backend system. At the demo we plan to show how our system works, i.e., how the system reacts upon arbitrary (interactive) workloads and how the user is alerted that a better configuration is available.

- *D. Klan, M. Karnstedt, Ch. Pölitz, K. Sattler: Towards Burst Detection for Non-*

Stationary Stream Data, KDML 2008: Knowledge Discovery, Data Mining, and Machine Learning, Oct 2008, Würzburg, Germany, pp. 57-60

Detecting bursts in data streams is an important and challenging task, especially in stock market, traffic control or sensor network streams. Burst detection means the identification of non regular behavior within data streams. A specifically crucial challenge on burst detection is to identify bursts in the case of non-stationary data. One approach is to apply thresholds to discover such bursts. In this paper, we propose a new approach to dynamically identify suitable thresholds using techniques known from time series forecasting. We present fundamentals and discuss requirements for threshold-based burst detection on stream data containing arbitrary trends and periods.

- *M. Karnstedt, D. Klan, Ch. Pölitz, K. Sattler, C. Franke: Adaptive Burst Detection in a Stream Engine, 2009 ACM Symposium on Applied Computing (SAC'09), Apr 2009, Honululu, USA*

  Detecting bursts in data streams is an important and challenging task. Due to the complexity of this task, usually burst detection cannot be formulated using standard query operators. Therefore, we show how to integrate burst detection for stationary as well as non-stationary data into query formulation and processing, from the language level to the operator level. Afterwards, we present fundamentals of threshold-based burst detection. We focus on the applicability of time series forecasting techniques in order to dynamically identify suitable thresholds for stream data containing arbitrary trends and periods. The proposed approach is evaluated with respect to quality and performance on synthetic and real-world sensor data using a full-fledged DSMS.

- *K. Hose, D. Klan, K. Sattler: Online Tuning of Aggregation Tables for OLAP, Workshop on Self Managing Database Systems (SMDB'09 icw. ICDE 2009), Shanghai, China*

  Materializing results from complex aggregation queries helps to significantly improve response times in OLAP servers. This problem is known as the view selection problem: choosing the optimal set of aggregation tables (called configuration) for a given workload. In this paper we present an online approach for adjusting the configuration dynamically to the current workload. This approach is implemented as part of an open source OLAP server and acts on the level of multidimensional MDX queries. The work presents the details of cost estimation and optimization of the system demonstrated in [10] and extends it by an online tuning strategy.

- *D. Klan, M. Karnstedt, Ch. Pölitz, K. Sattler: Towards Burst Detection for Non-*

Stationary Stream Data, KDML 2008: Knowledge Discovery, Data Mining, and Machine Learning, Oct 2008, Würzburg, Germany, pp. 57-60.

Detecting bursts in data streams is an important and challenging task, especially in stock market, traffic control or sensor network streams. Burst detection means the identification of non regular behavior within data streams. A specifically crucial challenge on burst detection is to identify bursts in the case of non-stationary data. One approach is to apply thresholds to discover such bursts. In this paper, we propose a new approach to dynamically identify suitable thresholds using techniques known from time series forecasting. We present fundamentals and discuss requirements for threshold-based burst detection on stream data containing arbitrary trends and periods.

- *C. Franke, M. Karnstedt, D. Klan, M. Gertz, K. Sattler, W. Kattanek: In-Network Detection of Anomaly Regions in Sensor Networks with Obstacles, BTW 2009, Münster, Germany*

  In the past couple of years, sensor networks have evolved to a powerful infrastructure component for monitoring and tracking events and phenomena in many application domains. An important task in processing streams of sensor data is the detection of anomalies, e.g., outliers or bursts, and in particular the computation of the location and spatial extent of such anomalies in a sensor network. In this paper, we present an approach that facilitates the efficient computation of such anomaly regions from sensor readings. We propose an algorithm to derive spatial regions from individual anomalous sensor readings, with a particular focus on obstacles present in the sensor network. We improve this approach by proposing a distributed in-network processing technique where the region detection is performed at the sensor nodes. We demonstrate the advantages of this strategy over a centralized processing strategy by utilizing a cost model for real sensors and sensor networks.

- *C. Franke, M. Karnstedt, D. Klan, M. Gertz, K. Sattler, E. Chervakova: In-Network Detection of Anomaly Regions in Sensor Networks with Obstacles, Computer Science - Research and Development, Special Issue "BTW 2009 Best Papers", Springer Verlag.*

- *K. Hose, C. Lemke, K. Sattler: Maintenance Strategies for Routing Indexes, Distributed and Parallel Databases - Data Management in Peer-to-Peer Systems, 2009.*

  Query processing in large-scale unstructured P2P networks is a crucial part of operating such systems. In order to avoid expensive flooding of the network during query processing so-called routing indexes are used. Each peer maintains such an index for its neighbors. It provides a compact representation (data summary) of

data accessible via each neighboring peer. An important problem in this context is to keep these data summaries up-to-date without paying high maintenance costs. In this paper, we investigate the problem of maintaining distributed data summaries in P2P-based environments without global knowledge and central instances. Based on a classification of update propagation strategies, we discuss several approaches to reduce maintenance costs and present results from an experimental evaluation.

- *D. Klan, K. Hose, K. Sattler: Developing and Deploying Sensor Network Applications with AnduIN, 6th International Workshop on Data Management for Sensor Networks (DMSN) icw. VLDB 2009, 6 Seiten. pp. 1511-1515.*

  Wireless sensor networks have become important architectures for many application scenarios, e.g., traffic monitoring or environmental monitoring in general. As these sensors are battery-powered, query processing strategies aim at minimizing energy consumption. Because sending all sensor readings to a central stream data management system consumes too much energy, parts of the query can already be processed within the network (in-network query processing). An important optimization criterion in this context is where to process which intermediate results and how to route them efficiently. To overcome these problems, we propose AnduIN, a system addressing these problems and offering an optimizer that decides which parts of the query should be processed within the sensor network. It also considers optimization with respect to complex data analysis tasks, such as burst detection. Furthermore, AnduIN offers a Web-based frontend for declarative query formulation and deployment. In this paper, we present our research prototype and focus on AnduIN's components alleviating deployment and usability.

- *Elena Chervakova , Daniel Klan , Tobias Rossbach, Energy-optimized Sensor Data Processing, 4th European Conference on Smart Sensing and Context (EuroSSC) 2009, pp. 35-38*

  We consider the problem of efficient data aggregation and processing within wireless sensor networks (WSN). The system we pro- pose allows monitoring and sensor data processing by injecting bytecode into the network. Additionally we present AnduIN, a system whose goal it is to decide which parts of a given query should be processed within the sensor network.

Weiterhin entstanden im Projektzeitraum eine Vielzahl an Studien- und Diplom bzw. Bachelorarbeiten, welche hier nicht im einzelnen aufgeführt werden sollen.

3.3 Vortragsreisen und Messen

Als Teil des wissenschaftlichen Austauschs, präsentierten Mitarbeiter das Projekt auf folgenden nationalen und internationalen Veranstalltungen:

- 11. EDBT (International Conference on Extending Database Technology) in Nantes / Frankreich 2008
- 34. VLDB (International Conference on Very Large Data Bases) in Auckland / Neuseeland 2008
- Workshop „Knowledge Discovery, Data Mining, Maschinelles Lernen“ als Teil der LWA 2008 in Würzburg / Deutschland
- GI-Fachtagung Datenbanksysteme für Business, Technologie und Web (BTW 2009) 2.–6. März 2009 in Münster / Deutschland
- 24th Annual ACM Symposium on Applied Computing (ACM SAC'09) 9.–12. März 2009 in Honolulu / USA
- 4th International Workshop on Self Managing Database Systems (SMDB 2009) 29. März 2009 Shanghai / China
- 6th International Workshop on Data Management for Sensor Networks (DMSN 2009) am August 24, 2009 in Lyon / Frankreich
- 26th IEEE International Conference on Data Engineering, 1-6. März, 2010, Long Beach, California / USA
- The Second International Workshop on Information Quality and Quality of Service for Pervasive Computing in Conjunction with IEEE PERCOM 2010, Mannheim / Germany, March 29, 2010

Die TU Ilmenau war durch Mitarbeiter auf folgenden öffentlichkeitswirksamen Veranstaltungen vertreten:

- Cebit 2008
- Cebit 2009
- CBS Projekt- und Strategieworkshop 2009
- Innovationstags Thüringen in Erfurt 2009

Literatur

[1] C. Giannella, J. Han, E. Robertson, and C. Liu. Mining Frequent Itemsets over Arbitrary Time Intervals in Data Streams. Technical report, Indiana University, 2003.

[2] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In *SIGMOD*, pages 1–12, 2000.

[3] D. Shasha and Y. Zhu. *High Performance Discovery in Time Series: Techniques and Case Studies*. Springer, 2004.

[4] R. Srikant and R. Agrawal. Mining quantitative association rules in large relational tables. *SIGMOD Rec.*, 25(2):1–12, 1996.

[5] Xin Zhang and Dennis Shasha. Better burst detection. In *ICDE '06*, pages 146–149, 2006.

[6] Y. Zhu and D. Shasha. Efficient Elastic Burst Detection in Data Streams. In *KDD 2003, Washington, D.C., USA*, pages 336–345, 2003.