



Liebe Leserinnen und Leser,

in dieser dritten Ausgabe des DBMS Newsletters, dem Mitteilungsblatt des Lehrstuhls für Informatik, insbesondere Datenbanken und Informationssysteme am Fachbereich 4 der Universität Münster, wollen meine Mitarbeiter und ich Ihnen wieder einen Überblick über Aktualitäten und Aktivitäten geben, welche mit unserer Arbeitsgruppe und den von uns bearbeiteten Themen zu tun haben.

Ein Schwerpunkt in dieser Ausgabe ist das Thema RFID (Radio Frequency Identification) and NFC (Near Field Communication), über das man in letzter Zeit recht viel hören und lesen konnte. Es handelt sich quasi um den elektronischen Nachfolger des Barcodes, der in Form kleiner Chips

**Die Schwerpunkte
im Newsletter 04/2007:
RFID und NFC**

an Gegenständen angebracht, aber auch von Personen getragen werden kann. Mit einem Lesegerät ist die jeweilige ID des Trägers festzustellen, die dann mit einer Vielfalt von Informationen verknüpft werden kann. Joachim Schwioren berichtet über unser Projekt MoVIS (Mobile Visitor Information System), welche die Entwicklung eines intelligenten Museumsführers zum Ziel hat und in Zusammenarbeit mit der Firma Elatec GmbH aus Eichenau und dem Geologisch-Paläontologischen Museum der Universität Münster durchgeführt wird.

Aus dem Seminar, über das wir in der letzten Ausgabe des Newsletters berichtet haben, ist inzwischen ein Buchprojekt zum Thema Web 2.0 entstanden, über welches Stephan Hagemann berichtet; das Buch wird im Laufe des Sommers bei Morgan Kaufmann in San Francisco erscheinen und ist eines der ersten, das sich diesem hochaktuellen Thema umfassend widmet. Sie erfahren ferner, was aus meinem ersten Münsteraner Doktoranden wurde, wir haben die Titel der Arbeiten zusammengestellt, die in letzter Zeit entstanden sind und die wir Ihnen auf Anfrage gerne zur Verfügung stellen; wir berichten aus der Lehre sowie aus Australien. Schließlich sind die bereits geplanten Vorträge des Sommers 2007 zusammengestellt.

Weitere und stets aktuelle Informationen erhalten Sie auch über unsere Webseite unter dbms.uni-muenster.de, die in der letzten Zeit überarbeitet wurde.

Ich wünsche Ihnen viel Spaß bei der Lektüre und würde mich über ein Feedback freuen.

Münster, im April 2007

Ihr
G. Vossen
Prof. Dr. Gottfried Vossen

FOTO: GABRIELE PETERS, BOCHUM

Ankündigungen

Gastvorträge

Im Sommersemester 2007 sind folgende Veranstaltungen geplant:

Interessenten sind jeweils herzlich willkommen

Audiocodierung und ihre Anwendungen ::

Am 04. Juni 2007 spricht Harald Popp, Fraunhofer-Institut für Integrierte Schaltungen (IIS), Erlangen, Miterfinder des mp3-Formats, über mp3, AAC, MPEG Surround-Audiocodierung und ihre vielfältigen Anwendungen. Dieser Vortrag findet im Rahmen des Treffens der GI-Regionalgruppe Münsterland ab 18 Uhr im Technologiehof statt.

Gastvorträge im Rahmen der Vorlesung Universelle Datenbanksysteme ::

- 06. Juni 2007, 8:15 Uhr
Frank Weymerich, Viadee Münster
- 13. Juni 2007, 14:15 Uhr
Dr. Bodo Hüsemann, Informationsfabrik Münster
- 20. Juni 2007, 14:15 Uhr
Dr. Peter Westerkamp, WGZ Bank Düsseldorf
- 27. Juni 2007, 14:15 Uhr
Frau Dingwerth/ Herr Baum, zeb/ Münster

Überraschungen aus der Welt der Software-Reviews ::

Am 02. Juli 2007 spricht Peter Rösler (reviewtechnik.de) über Zahlen und Fakten zu Software-Reviews und widmet sich auch der Frage nach dem Grund für die große Zahl scheiternder Softwareprojekte. Dieser Vortrag findet im Rahmen des Treffens der GI-Regionalgruppe Münsterland ab 18 Uhr im Technologiehof statt.

Neue Veröffentlichungen der Arbeitsgruppe ::

Service-Orientierung

GOTTFRIED VOSSEN:

From Processes via Workflows to Services; Transactions of the Society for Design and Process Science (TSDPS), December 2006, 1-9.

GOTTFRIED VOSSEN UND PETER WESTERKAMP:

Why Service-Oriented could Make E-Learning Standards Obsolete; eingereicht beim International Journal of Technology Enhanced Learning.

SIMON SCHWINGEL, GOTTFRIED VOSSEN UND PETER WESTERKAMP:

Securing the Infrastructure for Service-Based E-Learning Environments; wird erscheinen in Vijayan Sugumaran (ed.), Advanced Topics in Intelligent Information Technologies, Volume II

K. VIDYASANKAR UND GOTTFRIED VOSSEN:

Multi-Level Modeling of Web Service Compositions with Transactional Properties; zur Veröffentlichung eingereicht.

Data Warehousing

S. RIZZI, A. ABELLÓ,

J. LECHTENBÖRGER UND J. TRUJILLO:

Research in Data Warehouse Modeling and Design: Dead or Alive? Proc. 9th ACM International Workshop on Data Warehousing and OLAP (DOLAP), November 2006, 3-10.

J.-N. MAZÓN, J. TRUJILLO

UND J. LECHTENBÖRGER:

A Set of QVT Relations to Assure the Correctness of Data Warehouses by Using Multidimensional Normal Forms; Proc. 25th International Conference on Conceptual Modeling (ER), LNCS 4215, November 2006, 385-398.

TILL HASELMANN, JENS LECHTENBÖRGER UND GOTTFRIED VOSSEN:

Data Warehouse Detective: Schema Design Made Easy; Datenbanksysteme in Business, Technologie und Web (BTW 2007), Proceedings, März 2007, Aachen, 606-608.

E-Learning

DANIEL DAHL UND GOTTFRIED VOSSEN:

Added Value for e-Learning Repositories through Social Tagging of Learning Objects; Datenbanksysteme in Business, Technologie und Web (BTW 2007), Workshop Proceedings, März 2007, Aachen, 270-279.

DANIEL DAHL UND GOTTFRIED VOSSEN:

share.loc – Transparenz und Integration von Lernobjekten an der Universität Münster; Working Paper No. 9 des E-Learning Kompetenzzentrums am European Research Center for Information Systems, Münster 2007.



DANIEL DAHL UND GOTTFRIED VOSSEN:

Evolution of Learning Folksonomies: Social Tagging in e-Learning Repositories; eingereicht beim International Journal of Technology Enhanced Learning.

Web 2.0

GOTTFRIED VOSSEN

UND STEPHAN HAGEMANN:

From Version 1.0 to Version 2.0: A Brief History of the Web; Discussion Paper 2007-01, Dept. of Management Systems, University of Waikato Management School, Hamilton, New Zealand.

RFID-Technologie und Anwendungen

JOACHIM SCHWIEREN

UND GOTTFRIED VOSSEN:

Implementing Physical Hyperlinks for Mobile Applications using RFID Tags; in Vorbereitung.

Allgemeine Datenbank-Themen

GOTTFRIED VOSSEN:

Database Processing; wird erscheinen in J.G. Webster (Hrsg.): Wiley Encyclopedia of Electrical and Electronics Engineering, Volume 4 (2nd edition), John Wiley & Sons, New York, 2007.

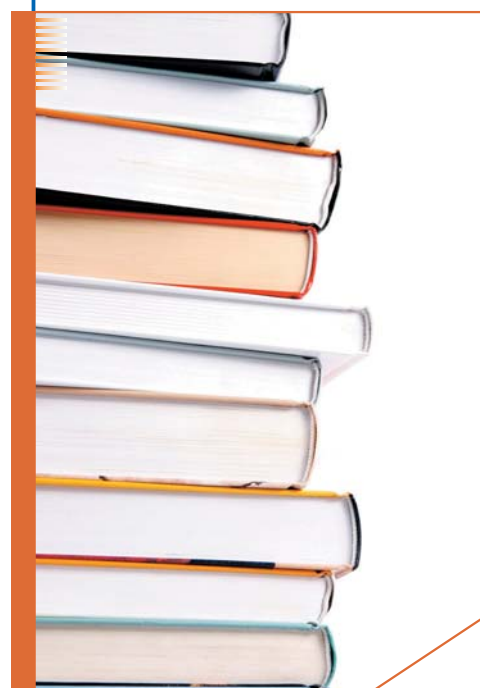
GOTTFRIED VOSSEN UND GERHARD WEIKUM:

Transactions; wird erscheinen in J. Hammer, M. Schneider (eds.): Handbook of Database Technology, 2007 (verfügbar als Discussion Paper 2007-02, Dept. of Management Systems, University of Waikato Management School, Hamilton, New Zealand).

Gender-Forschung

ERIC DEAKINS UND GOTTFRIED VOSSEN:

Women Leading and Learning through Present-Day Web Linking Technology; zur Veröffentlichung eingereicht (verfügbar als Discussion Paper 2007-03, Dept. of Management Systems, University of Waikato Management School, Hamilton, New Zealand).



Forschung im Fokus :: RFID-Technologie und mobile Informationssysteme

RFID ist die Abkürzung für Radio-Frequenz-Identifikation und steht für eine Technologie, mit der sich Objekte kontaktlos identifizieren oder orten lassen. In der letzten Zeit taucht das Schlagwort RFID immer häufiger in der Öffentlichkeit, in Publikationen und in der Presse auf. Auch wenn diese Technologie schon einige Jahre alt ist, erfährt sie doch im Moment einen enormen Boom, da ein wirtschaftlicher Einsatz in immer mehr Anwendungsgebieten erst jetzt durch sinkende Preise der RFID-Komponenten möglich wird. Darüber hinaus eröffnen sich zahlreiche neue Potenziale für innovative Einsatzszenarien von RFID-Technologie; so wird im Moment am Lehrstuhl der Prototyp eines PDA-basierten Besucherinformationssystems für Museen entwickelt, welches dem Besucher durch den Einsatz von RFID-Technologie einen einfacheren und intuitiven Zugriff auf Informationen zu Exponaten ermöglicht.

Grundlagen der Radio-Frequenz-Identifikation (RFID)

RFID-Technologie ermöglicht das kontaktlose Auslesen von Informationen aus speziellen Microchips (so genannten Transpondern oder Tags) mittels Funkwellen. Ein Lesegerät (auch Reader genannt) erzeugt dabei ein Induktionsfeld oder ein elektromagnetisches Feld, welches die Transponder in Reichweite über eine integrierte Antenne mit einem schwachen Strom versorgt. Dieser Strom reicht aus, um den Chip des (meist batterielosen) Transponders dazu zu veranlassen, die auf ihm gespeicherten Daten oder Identifikationscodes an den Leser wiederum per Funk zu übertragen (Backscatter).

Für RFID-Technologie gibt es sehr vielfältige Einsatzmöglichkeiten. Diese reichen von relativ simplen Anwendungen wie Diebstahlsicherungs-Etiketten in Kaufhäusern über Wegfahrsperrungen für PKW, Zugangskontrollsystemen zu Gebäuden, Zeiterfassung von Arbeitnehmern oder Tier-Identifikation mittels Transpondern in Ohrmarken im landwirtschaftlichen Bereich. Aber auch der Schutz von Markenartikeln oder Medikamenten vor Plagiaten kann mit Hilfe von RFID erreicht werden. Die Hauptanwendungsfelder finden sich zurzeit im Bereich der Logistik, z.B. bei der Lieferketten-Automatisierung oder dem Waren- und Bestandsmanagement. Hier kommt RFID häufig als Ersatz für Barcode-Systeme zum Einsatz, da RFID-Tags deutlich robuster und weniger anfällig für Beschädigungen als Barcode-Label sind und auch unter besonders aggressiven Umweltbedingungen (Hitze, Feuchtigkeit, etc.) problemlos funktionieren können.

Aber auch im medizinischen Bereich gewinnt RFID immer mehr an Bedeutung, da sich durch die zweifelsfreie und automatisierte Identifikation von Patienten, Betten, Medikamenten oder medizinischen Geräten Abläufe optimieren lassen und vermeidbare Behandlungsfehler reduziert werden können. RFID-Technik lässt sich in den unterschiedlichsten Einsatzbereichen sinnvoll verwenden. Doch nicht immer stehen nur die technischen Details der Funktechnik im Vordergrund: Häufig stellen sich bei der Anbindung an bestehende Informationssysteme erst die wirklichen Herausforderungen.

Potentiale von RFID

in Verbindung mit mobilen Informationssystemen

Häufig wird RFID als Brücke zwischen der physischen Welt einerseits und der „virtuellen“ Welt eines Informationssystems andererseits gesehen, denn es lassen sich Medienbrüche, die früher unumgänglich waren, verhindern. Besondere Möglichkeiten ergeben sich hier im Bereich der mobilen Informationssysteme, also solcher Systeme, die auf einem PDA (Personal Digital Assistant), einem Smart Phone, einem UMPC (Ultra Mobile PC) oder sonstigem mobilen Endgerät laufen.



SCHON BALD REALITÄT: MOVIS IM GEO-MUSEUM IN MÜNSTER

Da diese Systeme ohnehin losgelöst sind vom stationären Einsatz, ergeben sich hier besonders viele Möglichkeiten in Verbindung mit RFID-Technologie. Im Bereich von Krankenhäusern oder ambulanten Pflegediensten können PDA-basierte Informationssysteme mit integriertem RFID-Reader viele Prozesse verschlanken, Medienbrüche reduzieren, zeitraubende manuelle Datenerfassung einsparen und mehr Sicherheit für die Patienten schaffen, da z.B. das Risiko der Abgabe eines falschen Medikaments reduziert, wenn nicht sogar ausgeschlossen werden kann. Durch RFID werden Patienten und Medikamente automatisch identifiziert, und diese Informationen können dann direkt mit der digitalen Patientenakte abgeglichen werden.

RFID schafft eine Verbindung zwischen der physischen Welt und Informationssystemen

Aber auch in ganz anderen Bereichen, etwa bei Wartungsmaßnahmen von Objekten wie Feuerlöschern, Brandschutzklappen oder Heizungsanlagen kann RFID-Technik

in Verbindung mit einem mobilen Informationssystem dem Wartungstechniker nicht nur viel Zeit und Arbeit sparen, sondern auch mehr Sicherheit schaffen, da ein Techniker tatsächlich am zu wartenden Objekt gewesen sein muss. Ansonsten ist keine Abarbeitung des Wartungsauftrags im System möglich. Zusätzlich kann das System viel einfacher kontextabhängige Informationen zur Verfügung stellen (etwa Wartungsanleitungen oder technische Details), ohne dass der Benutzer diese über umständliche Suchmechanismen und den limitierten Interaktionsmöglichkeiten des PDA aufrufen muss.

Doch auch im persönlichen Umfeld können mobile RFID-basierte Informationssysteme eine wertvolle Hilfe sein. Neben Verbraucherinformationssystemen, die direkt am Point-of-Sale Zugriff auf detaillierte Informationen zu Produkten ermöglichen (insbesondere wenn, wie in Zukunft

geplant, beim Item-Level-Tagging alle Consumer-Produkte und Verpackungen mit einem RFID-Tag versehen sein werden) und so den Käufer z.B. über Inhaltstoffe, Herkunft oder mögliche Unverträglichkeiten informieren. Ein weiteres großes Einsatzfeld stellen touristische Informationssysteme wie mobile Städte-Guides oder PDA-Führer für Museen und Parks dar.

Besucherinformationssystem für Museen mit RFID-Unterstützung

Aktuell wird am Lehrstuhl im Rahmen einer Kooperation mit dem Geologisch-Paläontologischen Museum der Universität Münster und der u. a. auf RFID-Technologie spezialisierten Firma Elatec GmbH aus Eichenau ein Prototyp für das mobile, PDA-basierte Besucherinformationssystem MoVIS mit RFID-Unterstützung



entwickelt. Das Ziel ist es dabei, dem Museums-Besucher einen möglichst einfachen und intuitiven Zugang zu relevanten Informationen über Exponate zu eröffnen. Dabei werden die Exponate mit RFID-Tags ausgestattet, welche dann von einem in den PDA integrierten RFID-Leser durch den Besucher ausgelesen werden. Damit wird automatisch eine Verknüpfung zu den für dieses Exponat relevanten Inhalten wie

Texten, Audio-Beiträgen, Bilder oder Videos hergestellt. Ein lästiges Suchen nach den richtigen Informationen oder das umständliche Eintippen von Nummern-Codes entfällt.

Darüber hinaus können die Informationen speziell an die individuellen Bedürfnisse des Besuchers angepasst werden. So erhalten Kinder andere Informationen als Erwachsene und Experten wiederum andere Informationen als interessierte Laien. Neben der einfachen Handhabung für den Besucher ergeben sich auch für den Museumsbetreiber neue Möglichkeiten: Durch die erfassten Zugriffsdaten auf die Exponate lassen sich Rückschlüsse darüber ziehen, wie stark bestimmte Exponate frequentiert werden und wo Besucher noch mehr oder anders geartete Informationen benötigen könnten.

Abgerundet wird die Anwendung von der Möglichkeit für den Besucher, nach dem Museumsbesuch noch einmal auf die Informationen über einen speziellen Webzugang zugreifen zu können und so den eigenen Museumsbesuch noch einmal Revue passieren zu lassen. Nach der Entwicklung und Fertigstellung des Prototyps ist gegen Ende des Jahres ein Feldtest im Museum geplant.

<http://dbms.uni-muenster.de/projects/movis/>
<http://www.elatecworld.com/>

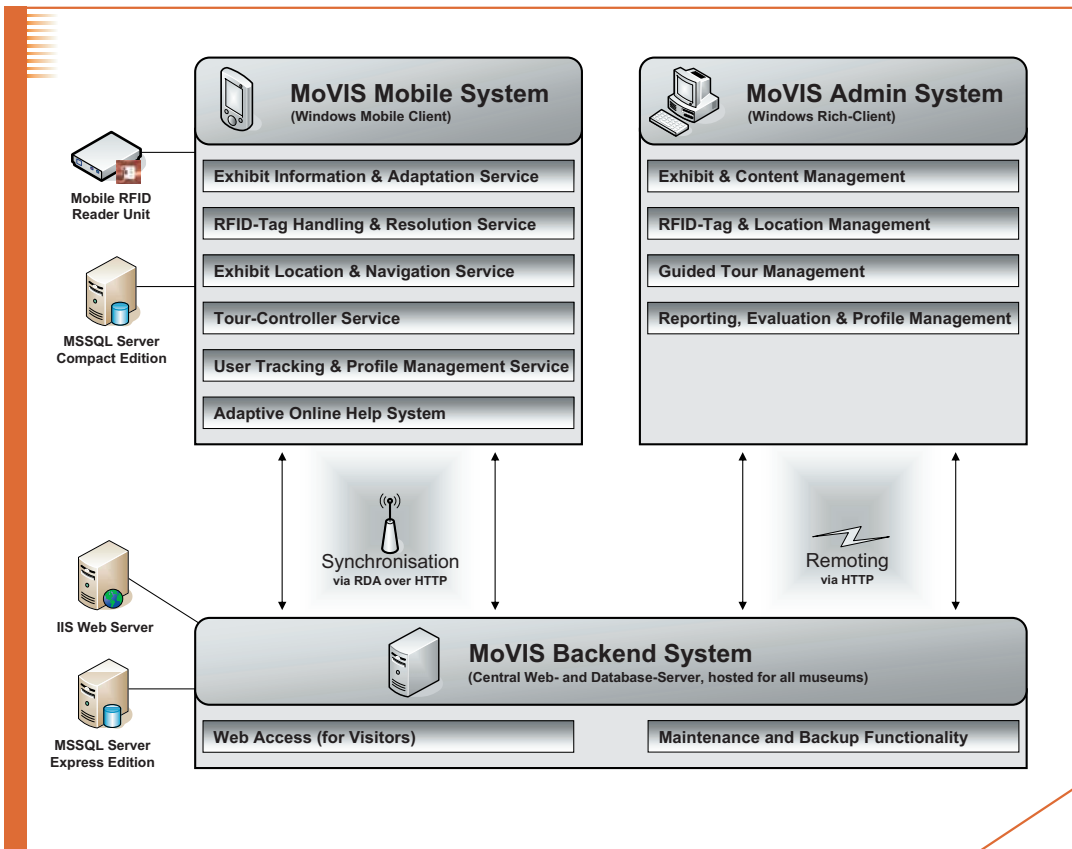
Was macht eigentlich... Dr. Fahrner?



CHRISTIAN FAHRNER promovierte im Jahr 1996 zum Thema „Schematransformationen in Datenbanken“ am Lehrstuhl für Informatik bei Professor Dr. Gottfried Vossen.

Im Juni 1996 hat Dr. Fahrner seine Laufbahn bei der heutigen Aareal Bank (vormals Teil der Depfa-Bank) in Wiesbaden begonnen und seitdem die Bereiche Treasury, Unternehmensrechnung und Information Technology in verschiedenen Positionen durchlaufen. Die Aareal Bank AG mit Hauptsitz in Wiesbaden ist eine der führenden internationalen Immobilienbanken, aktiv in über 20 Ländern, und im MDAX gelistet. Seit 2006 leitet Dr. Fahrner als Managing Director den Bereich Organisation in der Aareal Bank. Hierzu gehört neben der Verantwortung für die Infrastruktur, den Einkauf und verschiedenen Shared Services auch die Verantwortung für das Organisationsmanagement und das Inhouse-Consulting. Ein weiterer Aufgabenschwerpunkt liegt zurzeit in der Koordination von Standortgründungen in Asien und Nordamerika. Dem Lehrstuhl ist Dr. Fahrner verbunden geblieben u. a. durch ein gemeinsames Projektseminar im Sommersemester 2005.

MoVIS-ARCHITEKTURÜBERSICHT



Buchankündigung

Im Sommer 2007 wird das Buch *Unleashing Web 2.0 – From Concepts to Creativity* (Arbeitstitel) von Gottfried Vossen und Stephan Hagemann erscheinen. Das Thema Web 2.0 zeichnet sich, auch mehr als zwei Jahre nach seiner Einführung, insbesondere durch die Verwirrung aus, die es produziert. Dieses Buch versucht mit dieser Unordnung aufzuräumen. Dazu strukturiert das Buch die Entwicklungsströme des Internets, die seit seiner Entstehung Anfang der 90er Jahre zu beobachten sind. Diese lassen sich grob unterteilen in Änderungen der Anwendungen, der Technologie und des Nutzerverhaltens. Neuerungen wie Rich Internet Applications, Mash-ups und Tagging haben eine neue Qualität ins Internet gebracht, und zwar insbesondere bei Nutzerinteraktion und Informationsintegration. Das Schlagwort „Sozialisierung“ wird auf seine Auswirkungen auf Software, Suche und andere Anwendungen hin untersucht. Weitere Themen betreffen die Einflüsse, die heute von Web 2.0 Techniken auf das Web ausgehen und die Zusammenführung von Ideen des Semantischen Web mit denen von Web 2.0.

Aktuelles Projektseminar :: Web 2.0 – Community-getriebenes Lernen

Im Sinne des Blended Learning, der Mischform von Präsenzlehre und elektronischem Lernen, werden Lehrende wie Lernende in vielen Bereichen heute von mehr oder weniger komplexen Plattformen unterstützt: angefangen bei der Konzeption von Vorlesungen und Seminaren über die Verteilung von Lerninhalten bis hin zur teilweise automatisch korrigierbaren Abschlussprüfung im PC-Pool der Universität. Bislang wenig bis keine Beachtung findet die intensive Aufbereitung von Lernmaterialien (bspw. vor Klausuren), die bis heute in der Regel auf Ausdrucken in Papierform stattfindet. Mit dem Ziel, aus der „realen Welt“ bekannte und vertraute Funktionalitäten wie das Markieren und Hervorheben von Inhalten, das Anlegen von Notizen oder die Strukturierung von Inhalten in einem Softwaresystem abzubilden, entwickelten fünf Studierende in unserem Projektseminar des Wintersemesters 2006/07 die Plattform Learnr.



Angelehnt an populäre Web 2.0-Anwendungen setzt sich Learnr auf technischer Seite zum Ziel, den von Desktop-Applikationen bekannten Bedienungskomfort auf eine Webanwendung zu übertragen. Auf inhaltlicher Seite steht das Community-getriebene Lernen im Vordergrund: Während beispielsweise in der bekannten Anwendung Flickr (<http://www.flickr.com>) Internetnutzer Fotos zwecks Auffindbarkeit und Kategorisierung mit Schlagwörtern („Tags“ genannt) versehen, werden in Learnr vorlesungsbegleitende Foliensätze durch Lernende auf ähnliche Weise annotiert. In der Folge entwickelt sich in der Gemeinschaft ein die Inhalte auszeichnendes und einordnendes Vokabular. In Konsequenz dieses Lerner-zentrierten Ansatzes spricht das System nicht mehr die Sprache der Entwickler oder des lehrenden Dozenten, sondern primär die seiner Benutzer. Im kommenden Sommersemester 2007 wird Learnr erstmals vorlesungsbegleitend am Lehrstuhl für Informatik eingesetzt. Es gibt vielfältige Einsatzgebiete über die Hochschule hinaus: Weiterführende Schulen, Fortbildungen im Unternehmen oder das private Lernen sind nur einige zu nennende potentielle Szenarien.

<http://dbms.uni-muenster.de/teaching/>

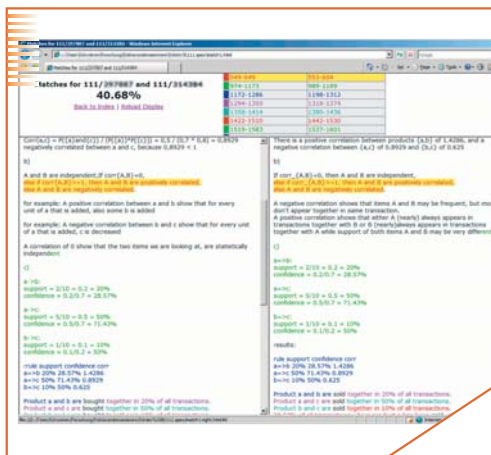
xLx-Plag :: Erkennung von Plagiaten im xLx-E-Learning-System

Web-basierte E-Learning-Systeme zur Unterstützung von vorlesungsbegleitenden Übungen, wie das am Lehrstuhl entwickelte und seit Jahren erfolgreich eingesetzte xLx-System, bieten sowohl den Lernenden als auch den Tutoren zahlreiche Vorteile. Dazu zählen z.B. mehr Flexibilität, größere Unabhängigkeit und letztlich eine effizientere Durchführung von Übungen im Lehrbetrieb. Die Versuchung kann für Studenten groß sein, Lösungen von anderen Übungs-Teilnehmern zu übernehmen oder aus dem Internet zu kopieren. Für die Tutoren ist die Erkennung solcher Plagiate in einem E-Learning-System jedoch aufgrund der hohen Teilnehmerzahl und der fehlenden Möglichkeit, alle Lösungen nebeneinander zu stellen, bisher sehr schwierig bis unmöglich.

← xLx-Plag findet identische Textpassagen.

Mit xLx-Plag, der neusten Innovation des xLx-Systems, die im Rahmen einer Diplomarbeit entstanden ist, lassen sich solche Plagiate jetzt automatisiert aufspüren. Dazu kommen spezielle Algorithmen zum Vergleich von Zeichenketten und zur Bestimmung von Ähnlichkeitsmaßen in Texten zum Einsatz. Durch den Einsatz des Systems konnten bereits erste Erfolge erzielt werden, da schon die bloße Ankündigung des Einsatzes dieses Tools einen signifikanten Rückgang der Abschreibversuche bewirkt hat. So sichert xLx-Plag eine faire Bewertung von Übungsaufgaben vom xLx-E-Learning-System.

<https://dbms.uni-muenster.de/xLx/>



Forschungsaufenthalt :: Besuch der University of Sydney

Im Rahmen seines Aufenthalts auf der Südhalbkugel hat Prof. Vossen Ende Oktober letzten Jahres die University of Sydney in Australien besucht. Dabei traf er in der Business School mit dem auch in Münster bekannten Marketing-Professor Charles Areni zusammen und stellte am 23.10. im Business Information Systems (BIS) Department im Rahmen eines Vortrags mit dem Titel „Information Systems Research from a European Perspective“ die Konzeption, die Ziele und aktuelle Arbeiten des Münsteraner ERCIS vor. Am 24.10. besuchte Prof. Vossen dann die School of IT, führte Gespräche u.a. mit Prof. Alan Fekete und Dr. Uwe Röhm und hielt einen Vortrag über das Thema „Ontology Engineering from a Database Perspective“. Die University of Sydney ist auch bei Münsteraner Studierenden bereits als Ziel eines Auslandsaufenthalts beliebt; dies dürfte nach dem jetzigen Besuch weiter zunehmen, da insbesondere mit BIS Folgeaktivitäten geplant sind.

Aktuelles Schlagwort :: Data Provenance

Die Einfachheit, mit der man heute Daten im Web kopieren und transformieren kann, macht es zunehmend schwieriger festzustellen, woher ein bestimmter Datensatz stammt. Anders ausgedrückt: Bei Daten im Web hat man häufig keinerlei Information, wie diese Daten an den betreffenden Ort oder auf die betreffende Webseite gelangt sind. Sie könnten z. B. dorthin kopiert worden sein, aus einer früheren Kopie stammen sowie im Laufe mehrerer Kopierprozesse editiert und verändert worden sein. Sucht man etwa nach günstigsten Preisen für eine bestimmte Ware, nach einer aktuellen Nachricht oder nach einer Medienbewertung, so erhält man auf diese Weise u. U. inkorrekte Informationen, und man hat sich ja inzwischen sogar daran gewöhnt, Daten aus dem Web nicht unbedingt und vollständig zu trauen. Der Begriff Data Provenance bzw. Data Lineage, zu Deutsch etwa Daten-Herkunft bzw. Daten-Abstammung, bezeichnet den Prozess des Rückverfolgens von Daten, des Erfassens ihrer Ursprünge und ihrer

In dieser Rubrik stellen wir in jeder Ausgabe einen aktuellen Begriff aus dem Umfeld von Datenbanken und Informationssystemen vor.

Bewegungen zwischen Quellen, insbesondere zwischen unterschiedlichen Datenbanken. Data Provenance ist in wissenschaftlichen und anderen Anwendungen, in denen die Genauigkeit und der Entstehungszeitpunkt von Daten zentral sind (z. B. in der Bioinformatik), von besonders hoher Bedeutung; auch im Data Warehouse-Bereich ist Provenance nicht unwichtig.

Offensichtlich ist die Analyse der Herkunft eines Dokuments kein neues Problem, sondern hat Historiker, Textkritiker und andere Wissenschaftler schon Jahrhunderte lang beschäftigt. Im Kontext von Datenbanken ist „Herkunft“ ein neueres und größeres Problem, da man unterschiedliche Granularitätsebenen unterscheiden muss: vom einzelnen Bildpunkt einer Image-Datei bis zu einem vollständigen Datenbankinhalt. Herkunft lässt sich offensichtlich dadurch dokumentieren, dass man Annotationen an Teilen einer Datenbank anbringt (im Stile von Marginalien an einem Text). Allerdings haben Datenbanken im Allgemeinen eine feste Struktur, welche insbesondere kein Hinzufügen beliebiger irregulärer Daten



PROF. STEVE ELLIOT (CHAIRMAN BUSINESS INFORMATION SYSTEMS)
UND PROF. GOTTFRIED VOSSEN

erlaubt. Daher arbeitet man in diesem Kontext an weniger rigiden Strukturformen für Daten sowie an entsprechender Software zum Speichern, Archivieren und Anfragen annotierter Daten. Ein technischer Ansatz zur Beschreibung von Herkunftsproblemen bedient sich etwa der Anfrageumkehrung (Query Inversion), die sich dafür interessiert, welche Datensätze in einer Datenbank zu einem gegebenen Output bzw. Anfrageergebnis beigetragen haben. Exemplarische Projekte, die sich mit Provenance und Lineage befassen, sind z.B. DBNotes zur Entwicklung eines Annotations-Managementsystems der University of California, Santa Cruz oder Trio-One des InfoLab der Stanford University.

Ein offensichtlich verwandtes Problem ist das der Accountability von Daten, was u. a. dadurch erreicht wird, dass man Daten mit einer zuverlässigen Historie versieht, so dass zurückliegende Ereignisse auf Datenverfälschung untersucht werden können. Hiermit befasst sich z.B. die Datenbankgruppe an der University of Massachusetts in Amherst, USA, deren Leiter Gerome Miklau erst kürzlich mit einem Faculty Early Career Preis der National Science Foundation ausgezeichnet wurde, der ihm weitere Arbeiten in diesem Bereich ermöglicht.

DBNotes: <http://www.cs.ucsc.edu/~laura/dbnotes/>

Trio: <http://infolab.stanford.edu/trio/>

Impressum

Herausgeber:

Prof. Dr. Gottfried Vossen

Lehrstuhl für Informatik

Universität Münster

Leonardo-Campus 3 | 48149 Münster

fon +49 251 83 38150

fax +49 251 83 38159

