1. Course Overview

Plan for Today's Lecture

- Who Are We, and Why Are We Here?
- DE and IA in the news
- Introduction to Key Concepts
- Syllabus Overview and Administrivia
Who are We and Why are We Here?

Instructor: Bob Glushko
Teaching Assistant: Zach Gillen
The Rest of You?

The Course Description

This course introduces the discipline of Document Engineering: specifying, designing, and deploying electronic documents and information repositories that enable document-centric or information-intensive applications. These applications include web services, information supply chains, single-source publishing, composite applications/virtual enterprises/portals, and so on.

The syllabus contains over 20 short case study examples from different industries, with special emphasis on business-to-business, healthcare and medical informatics, and e-government.

Document Engineering as I teach it is a superset of Information Architecture, extending its scope well beyond web site and web application design.
Document Engineering in the News

You don't yet know what Document Engineering is, but I do and we're going to start by seeing that it is in the news all the time.

What are the common issues and themes?

"Tailing Virulent Veggies" (WSJ, 3/13/07)

March 13, 2007

Tailing Virulent Veggies

Produce Industry Develops Means to Pinpoint Origin Of Contaminated Products

By JANE ZHANG
Scanning RFID Tags on Vegetable Boxes

"SPEEDy Airline Self-Service" (Hong Kong News, 1/07)
"More Clicks at the Bricks" (Business Week, 12/6/07)

The Web and other information and communications technologies have substantially affected consumer behavior.

This story analyzes some of these new patterns of information use (between people and information systems/resources, and between people).

Retail stores are offering new kinds of information services.

The Multi-channel Shopper

<table>
<thead>
<tr>
<th>THE INFORMED CONSUMER</th>
<th>Wholesale changes in the way we shop</th>
</tr>
</thead>
<tbody>
<tr>
<td>69%</td>
<td>Research products online before going to a store to make a purchase</td>
</tr>
<tr>
<td>62%</td>
<td>Have looked at least once at an online peer review before making a purchase</td>
</tr>
<tr>
<td>39%</td>
<td>Compared a product's feature and price across retail outlets online before buying</td>
</tr>
<tr>
<td>61%</td>
<td>Want to be able to scan bar codes and access information on other stores' prices</td>
</tr>
<tr>
<td>9%</td>
<td>Used a cell phone to text-message a friend or relation about a product while shopping</td>
</tr>
</tbody>
</table>
"Landstuhl to use Electronic Health Records" (Army Times, 1/16/08)

"Leaping" the Technological Divide?

With the new system, health care workers will create PDFs (portable document formats) for inpatient records, allowing them to be sent to the next hospital electronically.

Within a year, CliniComp officials hope to have all PDF records within its system at Landstuhl easily available to the Veterans Affairs Department as service members move from military to civilian status.

This has been a problem in the past; the Defense Department and VA medical records systems are not compatible.

CliniComp does not have the capability to scan and create PDFs for old medical records, so it only applies to records being created now.
"Big Employers Plan EHRs" (WSJ, 11/29/06)

Big Employers Plan Electronic Health Records

By GARY MCWILLIAMS
November 29, 2006; Page B1

Several big employers are about to deliver an electronic jolt to the U.S. health-care system.

Next week, Intel Corp., Wal-Mart Stores Inc., British Petroleum and others will disclose a plan to provide digital health records to their employees and to store them in a multimillion-dollar-data warehouse linking hospitals, doctors and pharmacies. Their goal: to cut costs by having consumers coordinate their own health care among doctors and hospitals.

Dossia Gains Momentum (dossia.org, 10/1/07)

Dossia Gains Momentum Toward Providing Employees with Personal, Private, Portable and Secure Health Records

2007 OCT 1 -- Dossia, a non-profit consortium of large employers, announced two new founding members and a healthcare advisor and technology collaborator, updating its progress toward the development of private, personal, portable and secure health records for its employees and their dependents, plus retirees.

Dossia was established by major U.S. employers Applied Materials, BP America Inc., Cardinal Health, Intel Corporation, Pitney Bowes Inc. and Wal-Mart to create a Web-based system that will enable employees to gain access to their own personal health data, which is now largely inaccessible to them. Dossia will use a Web-based infrastructure to empower individuals to manage their own health care, improve communications with their doctors, and provide more complete and accurate information for healthcare providers than the current system, which continues to be fragmented and still partially paper-based.
"Insurers Lack Governance" (Gartner, 1/9/08)

Insurers and Reinsurers Lack Governance for XML Initiatives, says Gartner

Gartner, Inc. - January 9, 2008

While XML adoption continues to rise year over year among insurers and reinsurers, many companies lack the governance and management structure to optimize its use, according to a joint survey conducted by Gartner and the Association for Cooperative Operations Research and Development (ACORD), a global, nonprofit insurance association whose mission is to facilitate the development and use of standards for the insurance, reinsurance and related financial services industries.

What Are the Common Themes in These News Items?

Enormous amounts of existing (paper) documents and legacy processes would benefit from automation, process re-engineering, transformation to SOA

New business processes are created / coordinated / choreographed via the management and exchange of electronic documents

Standards / patterns for documents and business processes are essential

Information technology and business processes are co-evolving with many ways to create business value

But projects can be challenging, and their success depends on many factors besides technology
So Document Engineering Isn't (Just) About XML

XML is a useful technology for Document Engineering, but using XML doesn't make you a document engineer.

The *best* thing about XML is the ease with which you can create a new vocabulary for a particular type of document.

XML is just the syntax in which we encode document models... what really matters is how we modeled the documents.

Creating Models is Easy, But Creating GOOD Models is Hard

The *worst* thing about XML is the same as the best thing – the ease with which you can create a new vocabulary.

No way around the classical problems of classification and naming we know from philosophy, linguistics, cognitive psychology, and information science.

XML is NOT "self-describing"

- The same content will inevitably be described using different names, and different content will be given the same names.

There are often multiple vocabularies for the same or related domains and especially for the common information models that are used in more than one domain.
The Document Exchange Pattern

Businesses / governments have long dealt with each other / their people through documents

Halfat's clay pot receipt for taxes is certainly one of the oldest documents that record a business transaction (355 BCE)

On the 16th of Tammuz, year 4 of Artaxerxes, Halfat brought barley: 1 kor, 12 seah, 3 qab; wheat: 1 kor, 5 seah, 4 qab.

The Document Exchange Pattern (continued)

Very natural thing to do

- the simplest case is "here's my catalog, do you want to buy anything" and the exchanged document being "here's my order"

We use concepts like "supply chains" and "distribution channels" as metaphors for the coordinated or choreographed flow of information and materials/products between businesses

These are complex patterns composed from the document exchange pattern
Document Exchange: The Mother of All Patterns

Document exchange is the "mother of all patterns" for business models, business processes, and business information

- **Business model or organizational**
  - patterns: marketplace, auction, supply chain, build to order, drop shipment, vendor managed inventory, etc.

- **Business process**
  - patterns: procurement, payment, shipment, reconciliation, etc.

- **Business information**
  - patterns: catalog, purchase order, invoice, etc. and the components they contain for party, time, location, measurement, etc.

Document Engineering and Web Services / SOA

"Web services" and "Service Oriented Architecture" are important concepts

The idea is simple – encapsulate or "wrap" some specific and discrete unit of functionality to hide its implementation and make it a reusable service (by sending it an XML message, to which it replies with an XML message)

Many business patterns like supply chains or virtual enterprises are a natural fit for web services, easy to see idea of service composition

But exchanging information does no good if the information can't be understood by the parties (or applications) doing the exchanging.

Document Engineering will ensure that the documents / services can be understood
Modeling Documents {and,vs,or} Modeling Processes

A document exchange -- or any web-based service -- consists of both the documents and the processes that produce and consume them.

By understanding the information in the documents, we learn what kinds of processes (or services) are possible.

By understanding the processes (or services), we learn what kinds of information are needed.

A Process-Centric Depiction

```
Product Catalog --> Order --> Invoice

Buyer Creates Order
Seller Processes Order & Creates Invoice
```
A Document-Centric Depiction

The Fundamental Difference is Visibility
Benefits of a Document-Centric Modeling Approach

Documents are more tangible than processes, easier to analyze and communicate

SOA emphasizes documents as the public interfaces to private processes

David Cohn: 100,000 nouns enable us to understand the meanings of 10,000 verbs

Document Component Architecture

Product Catalog → Order → Invoice

- Party
- Item ID
- Price
- Address
- Item Name
- Quantity
- Item Description
Overlapping Components in Business Processes

Component-based User or Application Interfaces
Course Syllabus

We've just touched on almost every topic in the course

Models of Business Organization and Business Processes
Models of Business Information; XML Vocabularies
Models of Business Architecture; Web Services
Analyzing and Modeling Business Processes, Documents and Information Components
Model-Based Applications
Management and Strategy Issues, Case Studies

Required Readings


All readings will be available online or as paper handouts; do we need a course reader? Some copyright clearance issues to work out
Course Deliverables

6 assignments throughout the semester. These assignments are designed to develop and reinforce practical skills in analysis, modeling, and implementation of document-centric and model-based applications

Team Project during the second half of the semester

There is no final exam or midterm

Course Grading

Assignments are 50% of final grade

Team project 40%

Class participation (in class, list serve, blog, etc.) 10%
Team Project

Individual assignments teach separate skills that you'll bring together in a 2-4 person team project

Last half of the semester; pick project by early March, incremental reports up to final presentation and report at semester end

For ISchool 2008ers, could be component of MIMS project

For ISchool 2009ers, could be incubator for MIMS project (or summer internship, or 2008-9 GSR appt.)

Some Past Document Engineering Projects
[1]

Course Approval System -- analysis and redesign of system by which new courses are born (primary clients: UCB Academic Senate, IS&T)

System Map -- interactive inventory and visualization of campus IT systems, precursor to campus-wide Data Dictionary (primary client: Central Computing Services)

Digital Chemistry -- data modeling to enable content sharing across delivery platform (primary client: Chemistry department [Mark Kubinec])

Event Calendar Network -- replace hodge-podge of calendars that can't share events with repository and syndication/reuse network (primary client: public affairs [Jeff Kahn])

Single-source reuse of Genentech process control documents

Personal health record data model
Some Past Document Engineering Projects [2]

- Bio/Bib - reuse of information for faculty reviews
- Advancing to PhD Candidacy Process and Documents
- Construction Project Management and Collaboration
- Class Chat data modeling and analysis
- Data model for "multimodal phone dialogs"
- iNaturalist "data entry spectrum"
- SF General Hospital EHR information flow
- State of California e-services feasibility

Class Participation

- Demonstrating that you're keeping up your part of the bargain
- You are not expected to be equally interested in and expert on every topic and every reading
- You are expected to be passionately interested in something – and to show it – or what are you doing in this course?
- Participation on the list serve can substitute for or complement your participation in class
My Biases and Expectations

Document Engineering is practical but also intellectually challenging.

I'm not a formal person and will be as accessible as I can to all of you – I will have two official office hours at least twice on M,T, or W.

But my informality doesn't mean I'm casual about what goes on in my class.

Course Web Sites and List Serves

Sylvia is familiar to SIMS students.

Sign up for "i243" list serve

- e-mail to majordomo@ischool.berkeley.edu
- Subject: Leave blank
- Body of message: subscribe i243
Readings for 28 January

part of Chapter 16 of Document Engineering [Textbook, 554-571]

"Tailing Virulent Veggies" -- Wall Street Journal (March 13, 2007)

"More Clicks at the Bricks" -- Business Week (December 6, 2007)

"Landstuhl to use Electronic Medical Records" -- Army Times (January 16, 2008)

"Why XForms", Elliotte Rusty Harold, (October 31, 2006)

"Accelerating RosettaNet" -- E-Commerce World (November 2001)

"Department of Homeland Security's Efforts Promise to Be the Biggest Change Management Job of All Time" -- CIO (December 2002)