IST 516
Introduction

Dongwon Lee, Ph.D.

Penn State / IST
Class Webpage

https://online.ist.psu.edu/ist516/ or via ANGEL
Welcome Message sent from Dongwon Lee

IST 516 Syllabus
https://online.ist.psu.edu/ist516/syllabusfall2011

Course Content
https://online.ist.psu.edu/ist516/

The Question Cafe
Have a question about the course that might be of interest to the rest of the class? Post it here!

MediaSite and Live Question Tool
Use these tools to view the live lecture, ask questions from a distance during the live lecture, and view the recorded lecture

Adobe Connect
Use this tool for online Office Hours and to meet with other students online.

Discussion Activity

Team Project - Reports and Presentations (Team Assignments)
Introduction

- Instructor
  - Dongwon Lee, Ph.D.
  - Associate Professor @ IST

- TA
  - TBD

- Discussion Activity #1 in AGNEL
  - Briefly (1-2 paragraphs) introduce yourself
  - When students form teams for Project, this introduction can be used to find team members
  - Will NOT be graded
Office Hour

- 2 Office hours:
  - Dongwon Lee: WED Noon – 2PM
  - TA: TBD

- Use Adobe Connect

- Email
  - Use ANGEL to send/receive emails
  - Send to BOTH me and TA so that any of us can answer fast
  - Will get answers within 24 hours at most
The Question Café

Have a question about the course that might be of interest to the rest of the class? Post it here!

- Settings  Reports  Utilities  Submissions  Delete

[Threaded View]

[-] Directions

This is a public discussion forum for you to post questions either to the instructor(s) or other students in the course. Remember—there are no stupid questions, and they may also be used by the instructor(s) to ask you questions about how you think the course is going.

Test your webcam at this link! Suggest EVERYONE use Firefox not IE!

http://www.testwebcam.com/

This link helped me prove my new webcam was working and you should be prompted if your Adobe Flash is missing or out-of-date too.

One last bit of advice to all. The PSU IT Support person I spoke with suggested that Angel "works better" using Mozilla Firefox than Internet Explorer. Try it better for me.

Lastly, from Angel's Home menu, don't forget to update your personal information in the Person (last nav button top left) is where!

Good Luck!
Do NOT Broadcast Emails

- Like to use broadcast mechanism mainly for class-wide announcement by Instructors
- Use ANGEL Question Café instead
  - Questions and replies are archived and threaded → better management
  - We monitor and answer them regularly
  - Too many broadcast emails → students tend to ignore instructors’ broadcast emails
- Do NOT send emails to entire class members in ANGEL
Grading Weights

- 20% : 6 Labs (Individual)
  - Short exercise to master course materials
  - One lowest score will be discarded
- 22% : 2 Homework (Individual)
  - Longer more in-depth assignment
- 25% : 1 Project (Team)
  - Design and implementation assignment
- 3% : 1 Presentation (Individual)
- 30% : 1 Final Exam (Individual)
  - Open-book written exam covering all materials
Grading Scale

Use standard PSU grading scale

- A : 93% to 100%
- B+ : 90% to 92.9%
- B  : 87% to 89.9%
- B- : 83% to 86.9%
- C+ : 77% to 79.9%
- C  : 70% to 76.9%
- D  : 60% to 69.9%
- F  : below 60%
Eg, Fall 2012 Grade Distribution

<table>
<thead>
<tr>
<th>Grades</th>
<th>Distribution (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9</td>
</tr>
<tr>
<td>A-</td>
<td>13.3</td>
</tr>
<tr>
<td>B+</td>
<td>13.3</td>
</tr>
<tr>
<td>B</td>
<td>24.4</td>
</tr>
<tr>
<td>B-</td>
<td>13.3</td>
</tr>
<tr>
<td>C</td>
<td>11.1</td>
</tr>
<tr>
<td>C-</td>
<td>4.4</td>
</tr>
<tr>
<td>D</td>
<td>2.2</td>
</tr>
<tr>
<td>F</td>
<td>2.2</td>
</tr>
</tbody>
</table>
What is IST 516?

- Roughly 2 sections
  - First half: The Web and its Standards
  - Second Half: Information Retrieval

- Objectives
  - Understand fundamental concepts, theories, methods, and techniques of modern Information Retrieval and Web Techniques.
  - Understand the design and basic functionalities of modern information retrieval system and search engines systems.
Part #1: The Web

The XML Family of Specifications: The Big Picture

Last Updated: April 16, 2002

Copyright (c) 2002 Kenneth B. Sall. All Rights Reserved.
Part #1: IR and SE

- IR Basics
- Vector Space Model
- Evaluation
- Relevance Feedback
- Probabilistic IR (if time permits)
- Clustering/Classification
Format of Class

- Consists of Instructor's presentation and discussion
- Materials
  - Materials will be drawn from academic literature, the Web, and two textbooks
  - NO required textbooks
  - WD: Web Dragon
  - IIR: Intro. To Information Retrieval
Team Project

- A team based project: 250 points
  - 1 planning report: 10% (25 points)
  - 1 mid-term progress report: 30% (50 points)
  - 1 final report and demo: 60% (175 points)
    Pre-recorded and played during class
  - All members get the same score for projects, but presentations are graded separately

- To be assigned around Week #7

- Based on other students’ introduction, form your own team of 3-4 members till week #7
  Email me your team members when formed
Design and implement a useful/interesting web app that leverages on several techniques to be covered in class:

- XML/XML Schema
- CCS/RSS
- XSLT/XPath
- Web Service / RDF
- Cloud Computing
- Search Engine, etc
Assignments

- Check the schedule at syllabus page
  - Due is **SUN 11:55PM** by default
  - About 10 days to work (2 weekends)
- Assignments must be TYPED and dropped to proper ANGEL drop boxes
  - hand-drawn figures are OK → scan and drop
  - Late assignments allowed till TUE 11:55 PM
  - **25% deduction** for every 12 hours late
Pre-requisite

- Requires basic **coding** and primitive college-level **math** skills
- Several assignments involve simple to medium-level coding (eg, 50-100 lines)
  - Project requires some traditional programming (eg, PHP, Java, or JavaScript)
    - We do **NOT** teach this
  - Labs/Homework require coding in web standard (eg, XML Schema, CSS, XSLT, XPath)
    - We do teach this
Pre-requisite

• Requires basic **coding** and primitive college-level **math** skills
  • 2nd half materials on IR use some (pre)-calculus
    Function, cosine, binomial, factorial, logarithm, …

\[
= \sum_{t \in q \cap d} (1 + \log tf_{t,d})
\]

• Little derivative / integral
• Some Probabilities
  \[ [P(A) - P(E)] / [1 - P(E)] \]
Pre-requisite

- If you
  - Hate coding, and/or
  - Don’t have basic Math covered

- Do **NOT** take this course
  - Take other stepping courses in IST first and try this course later times