



**INLS 490-154: Introduction to Information
Retrieval System Design & Implementation
Fall 2008**

IR on Web 2.0

“Some business, some buzz”

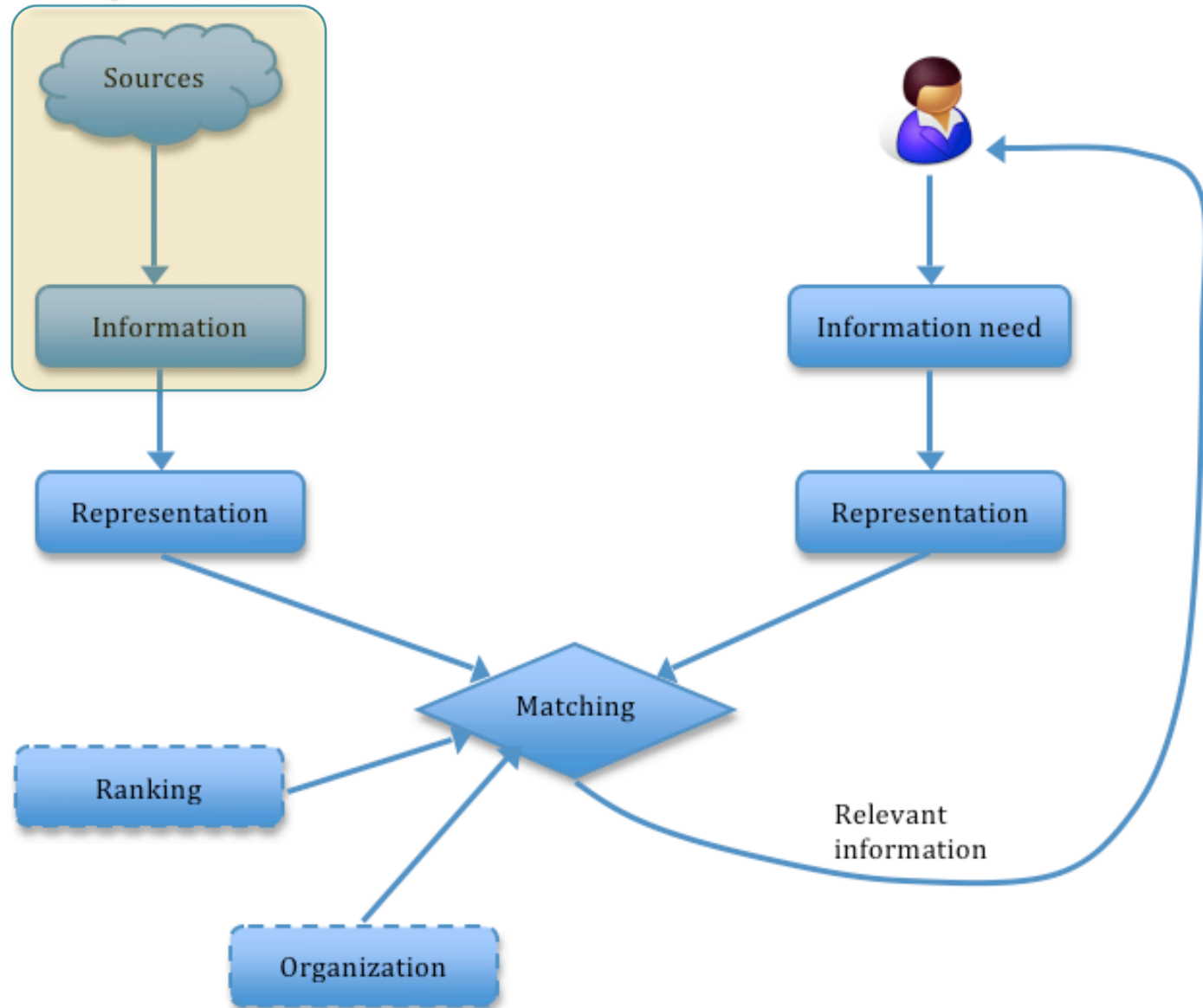
October 29, 2008

Chirag Shah

School of Information & Library Science (SILS)

UNC Chapel Hill

Today's lab





Today's lab

- Introduce REST
- Learn to parse XML with PHP
- Issue REST service calls



REST

- REST = REpresentational State Transfer
- An architecture style, not a standard.
- Guideline to define and address resources.
- Each resource is identified with a URI (in case of HTTP).
- REST typically deals with exchanging resources and concepts, while RPC deals with exchanging methods.

Parsing an XML file with PHP

- Three processes:
 1. A function to handle the start tags
 2. A function to handle the data between the tags
 3. A function to handle the end tags
- Steps:
 1. Create an XML parser object.
 2. Assign handlers for the start and the end of an element.
 3. Assign handler for the data between the tags.
 4. Parse XML file and store elements and values in an array.
 5. Free XML parser memory.



Web-based service calls

- Yahoo! REST requests for the web search
<http://developer.yahoo.com/search/rest.html>

Web feeds

- Web feed = data feed = news feed
- Feed is a document, often XML-based
- Formats of feeds: RSS, Atom
- RSS = Really Simple Syndication
- Content distributors *syndicate*
- Users *subscribe*
- *Aggregation* = a collection of web feeds available at one place
- Aggregator = feed reader = news reader





Summary

- Web 2.0 about effectively sharing information, and building communities and facilitating collaboration around information.
- Concepts introduced: REST, XML, feeds, syndication, aggregation



Next time

- More “traditional” ways of collecting data off the web.
- Designing and building a web crawler.