NC STATE UNIVERSITY

ENGINEERING ONLINE

Lecture Notes

Course Number: CSC 513

Instructor: Dr. Singh

Lecture Number: 20

Outline

XML Representation

- Concepts
- Parsing and Validation
- Schemas

What is Metadata?

Literally, data about data

- Description of data that captures some useful property regarding its
  - Structure and meaning
  - Provenance: origins
  - Treatment as permitted or allowed: storage, representation, processing, presentation, or sharing
- Markup is metadata pertaining to media artifacts (documents, images), generally specified for suitable parsable units

ad hoc markup

Locke Brothers
123 Main St
Raleigh, NC 27695
10 February 2011

Custom...

Qty	Sku	SKU	Price	LineTotal
100	Door Locks...		$19.95	1995.00
#1	#2		#3	#4

#6	1995.00
~~Total~~ Shipping	5.00
# Taxes	160.00
#7 Total	2160.00

Terms.
Delivered: 15-Jan-11
#5

Net Due in 30 days

INDICATIVE OF THE COMMITMENTS INVOLVED

WHAT DO WE NEED TO KNOW TO INTERPRET THIS CORRECTLY?

1. QUANTITY
2. GOODS
3. UNIT PRICE
4. TOTAL PRICE
5. DELIVERY DATE
6. SHIPPING CHARGES
7. Total Due

↓
— TO INSERT INTO A DB
—

UBL UNIVERSAL BUSINESS LANGUAGE
SPECIFIES TERMINOLOGY SUCH AS THE ABOVE

Landmark

Motivations for Metadata

Mediating information structure (surrogate for meaning) over time and space

- Storage: extend life of information
- Interoperation for business
- Interoperation (and storage) for regulatory reasons
- General themes
  - Make meaning of information explicit (MORE)
  - Enable reuse across applications: *repurposing* (compare to screen-scraping)
  - Enable better tools to improve productivity

Reduce need for detailed prior agreements

Markup History

How much prior agreement do you need?

- No markup: significant prior agreement
- Comma Separated Values (CSV): no nesting
- Ad hoc tags
- SGML (Standard Generalized Markup L): complex, few reliable tools; used for document management
- HTML (HyperText ML): simplistic, fixed, unprincipled vocabulary that mixes structure and display
- XML (eXtensible ML): simple, yet extensible subset of SGML to capture *custom* vocabularies
  - Machine processible
  - Comprehensible to people: easier debugging

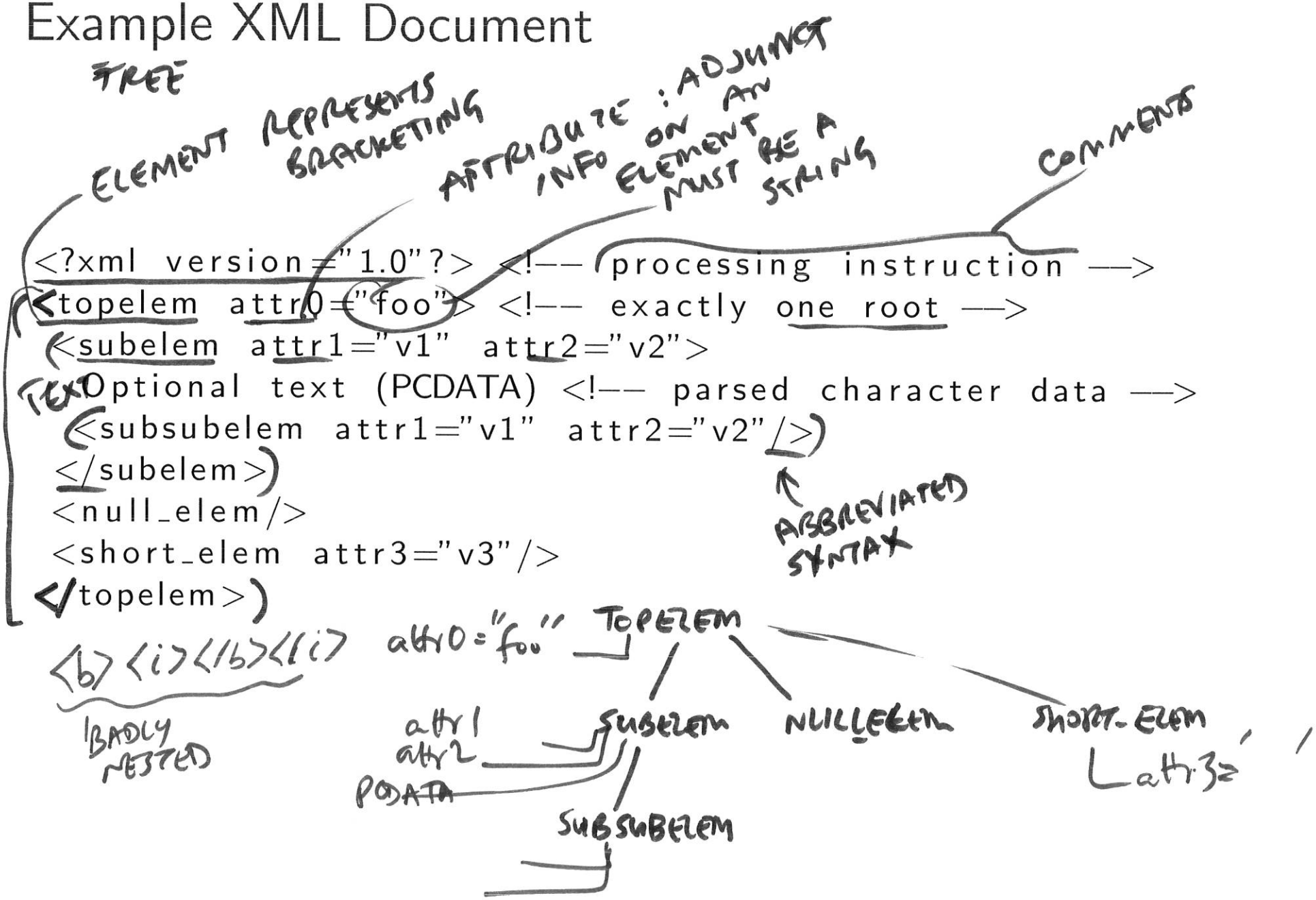
Uses of XML

time
space

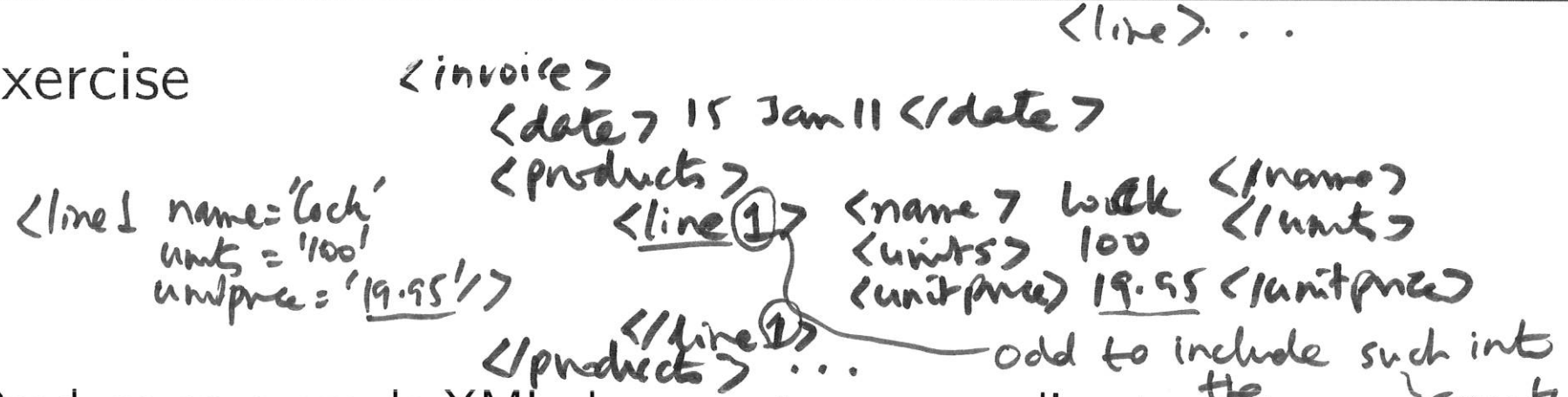
Supporting arms-length relationships

- Exchanging information across software components, even within an administrative domain
- Storing information in nonproprietary format
- Representing semistructured descriptions:
  - Products, services, catalogs
  - Contracts
  - Queries, requests, invocations, responses (as in SOAP): basis for Web services

Example XML Document



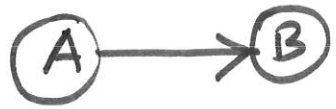
Exercise

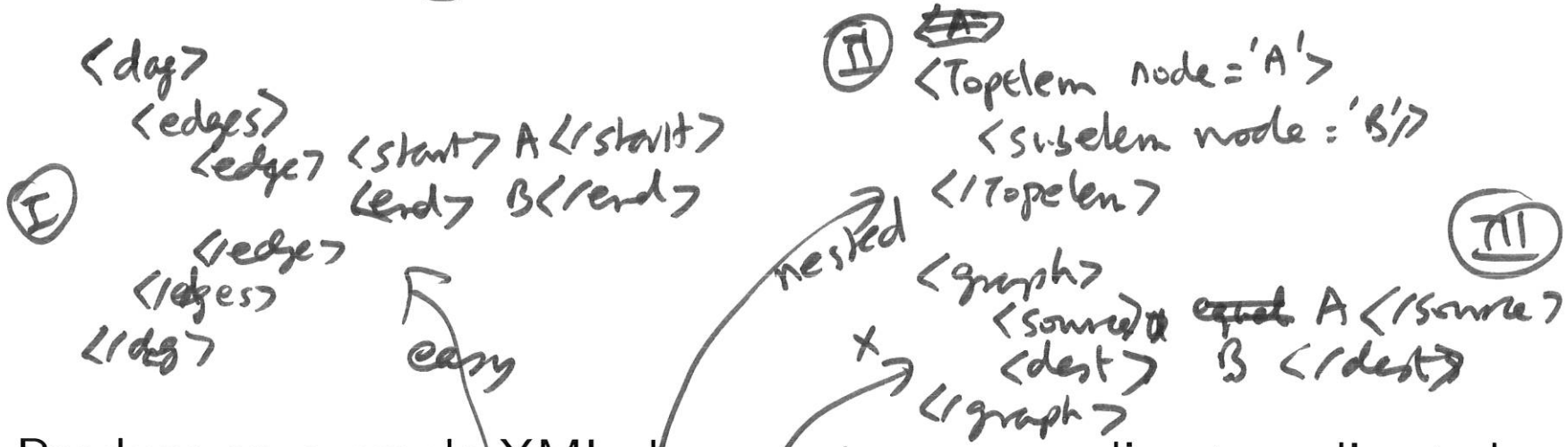


Produce an example XML document corresponding to

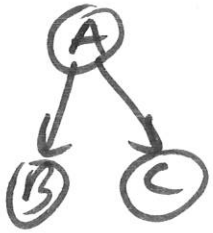
- An invoice from Locke Brothers for 100 units of door locks at $19.95, each ordered on 15 January and delivered to Custom Home Builders
- Factor in certified delivery via UPS for $200.00 on 18 January
- Factor in addresses and contact info for each party
- Factor in late payments

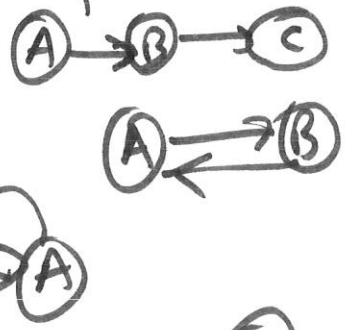
Exercise





Produce an example XML document corresponding to a directed graph





: II is impossible

: I is impossible