Class Announcements

- Database Assignment 1 Due 10/20
- (regular) Assignment 3: Due 10/25
- Midterm: 10/27

Alibaba Discussion

Networked Computing in direct Procurement

- History predates Internet
- Electronic Data Interchange (EDI)
  - Exchange order information between firms involved in direct procurement
  - Usually large firms who could afford proprietary communication links
  - Initially order and invoice
  - Existed since 70's

In Class Exercise

- Use a laptop or phone and log in to Alibaba.com
- Pretend you are a small business in the US needing cheap products. Do a search, and say what your "story" is.
  - E.g. "We pretend to be a retailer aiming to sell American flags at marked up prices. We'll search for American flags."
  - E.g. "We pretend to be a contractor looking for cheap supplies. We'll search for "toilet."
  - E.g. "We pretend to be a beach shop selling tourist goods. We'll search for "sunglasses."
- How do the prices you get compare to retail prices? Are there minimum quantities to make an order?
- Are the first listings from "gold" suppliers?
- Do these suppliers seem to be large companies or small to medium sized?
Data and information

by

David G. Messerschmitt

Key concept

The key commodity manipulated by information technology is information.

To be manipulated in a computing/networking environment, information must be represented by data.

What is information?

Information

From a user (human) perspective...

...recognizable patterns that influence you in some way (perspective, understanding, behavior...)

In the computing infrastructure, information has a somewhat different connotation as structure and interpretation added to data.

Data

A bit is “0” or “1” — the atom of the information economy.

Data is a collection of bits, like

- “01011011010110”
- “0000011”
- “11101110101101101101101011010”

Note: the terms data and information are not always used consistently!

Representation

- Take the place of the original
- Equivalent to, in the sense that the original can be reconstructed from its representation
- Often the original can only be approximately reconstructed, although it may be indistinguishable to the user
  - e.g. audio or video
ASCII

<table>
<thead>
<tr>
<th>Alphabet</th>
<th>Hex</th>
<th>Binary</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;7&gt;</td>
<td>x37</td>
<td>00110111</td>
</tr>
<tr>
<td>&lt;8&gt;</td>
<td>x38</td>
<td>00111000</td>
</tr>
<tr>
<td>&lt;9&gt;</td>
<td>x39</td>
<td>00111001</td>
</tr>
<tr>
<td>&lt;&gt;</td>
<td>x3a</td>
<td>00111010</td>
</tr>
<tr>
<td>&lt;=</td>
<td>x3b</td>
<td>00111011</td>
</tr>
<tr>
<td>=&gt;</td>
<td>x3c</td>
<td>00111100</td>
</tr>
<tr>
<td>&lt;=&gt;</td>
<td>x3d</td>
<td>00111101</td>
</tr>
<tr>
<td>&lt;*&gt;</td>
<td>x3e</td>
<td>00111110</td>
</tr>
<tr>
<td>&lt;At&gt;</td>
<td>x3f</td>
<td>00111111</td>
</tr>
<tr>
<td>&lt;A&gt;</td>
<td>x40</td>
<td>00000000</td>
</tr>
<tr>
<td>&lt;B&gt;</td>
<td>x41</td>
<td>00000001</td>
</tr>
<tr>
<td>&lt;C&gt;</td>
<td>x42</td>
<td>00000010</td>
</tr>
<tr>
<td>&lt;D&gt;</td>
<td>x43</td>
<td>00000011</td>
</tr>
</tbody>
</table>

Note that this representation is not unique…
…this one happens to be a standard (ANSI X3.110-1983)

A picture

This picture conveys information

This information is represented in this computer, but how?

Representation of picture: image

Expanding a small portion of the picture, we see that it is represented by square pixels…
…300 tall by 200 wide…
…with a range of 256 intensities per pixel

300 • 200 • 8 bits = 480,000 bits (but it can be compressed)

Color picture

A color picture can be represented by three monochrome images…

At the expense of three times as many bits

Terminology

Communication data to another user or organization

Representation needs to be standardized

If the representation is not standardized, the information is garbled!

Communicate data to another user or organization
**Regeneration**

- Make a precise copy of the data (copy bit by bit)
- If you know the representation, this is equivalent to making a precise copy of the information
- Each such precise copy is called a generation
- The process is called regeneration

**Replication of information**

Anything that can be regenerated can be replicated any number of times.

This is a blessing and a curse.

**Analog information cannot be regenerated**

Analog information can be copied, but not regenerated.

We will never know exactly what the original of this Rembrandt looked like.

**Discrete information can be regenerated**

Regeneration can preserve data (but not its original physical form).

Regeneration is possible for information represented digitally (which is tolerant of physical deterioration).

\[ 0 + \text{noise} \rightarrow 0 \]
\[ 1 + \text{noise} \rightarrow 1 \]

**Replication of information requires knowledge of representation**

Replication of information also presumes knowledge of its representation.

Replication preserves the integrity of the data, but that is not sufficient.

**Implications**

Digitally represented information can be preserved over time or distance in its precise original form by occasional regeneration.

- digital library
- digital telephony

Replication of data is easy and cheap.
Implications (con’t)

- Replication of information requires knowledge of the structure and interpretation
  - Standardization or some other means
- Extreme supply economies of scale
- You can give away or sell and still retain
- Unauthorized replication or piracy relatively easy