DATA CENTER

Make sense of your data

David Brock, Founder and Director
Data Center
Massachusetts Institute of Technology
<CallsPerDay>
  2575
</CallsPerDay>
PROBLEM

“CallsPerDay”

?  

That’s not even a word
ANOTHER PROBLEM

<CompanyData>
  <CompanyName>
    Fidelity Employer Services Company
  </CompanyName>
  <Location>
    Merrimack
  </Location>
  <CallData>
    <RecordDate>
      Tue Aug 11, 2004
    </RecordDate>
    <CallsPerDay>
      2575
    </CallsPerDay>
  </CallData>
</CompanyData>
MORE PROBLEMS

<CompanyData>
  <CompanyName>
    Fidelity Employer Services Company
  </CompanyName>
  <Location>
    Merrimack
  </Location>
  <CallData>
    <RecordDate>
      Tue Aug 11, 2004
    </RecordDate>
    <CallsPerDay>
      2575
    </CallsPerDay>
  </CallData>
</CompanyData>

<CorporateRecords>
  <Company>
    Fidelity Employer Services Company
  </Company>
  <Records Data="Tue Aug 11, 2004">
    <Calls Units="PerDay">2575</Calls>
  </Records>
</CorporateRecords>
OBSERVATION

<CompanyData>
  <CompanyName>
    Fidelity Services
  </CompanyName>
  <Location>
    Merrimack
  </Location>
  <CallData>
    <RecordDate>
      Tue Aug 11, 2005
    </RecordDate>
    <CallsPerDay>
      2575
    </CallsPerDay>
  </CallData>
</CompanyData>
PROPOSAL

Vocabulary + Grammar

Words without structure  Structure without words
INTEGRATE
INTEGRATE
DICTIONARY ENTRY

KEY

call.5

DEFINITION

call  n.  a telephone connection.

SEMANTIC LINKS

Synonyms: phone_call.1, telephone_call.1
Type of: telephone.2, telephony.1
Part of:

SCHEMA

Data: ^[+][0-9]\d{2}-\d{3}-\d{4}$
Attributes: party.5, duration.1, telephone_number.1

LOCALIZATION

Data: 电话, telefoongesprek, 전화, ТЕЛЕФОННЫЙ А ВЫЗОВ
DICTIONARY DEVELOPMENT

- Oxford English Dictionary
- National Library of Medicine
- Unified Medical Language System
- United States Department of Defense
- Princeton University, WordNet
- American Chemical Society
- Chemical Abstracts Service
- West Law Publishing
- Black’s Law Dictionary
- Acronym Finder
- Acronym Dictionary
DICTIONARY DEVELOPMENT

• Web accessible
• Web editable
• Web community
• Staged approval
  • Proposal – Universal accessible and editable
  • Draft – Universal accessible and limited editable
  • Pre-approval – Universal accessible and limited comments
  • Recommendation – Universal accessible
Grammar
purchase_order.1

customer.1

name.2

"Bill Buckram"

credit_card.1

ID.3

quantity.1

name.1

price.1

description.1

2938 8488 4001 5439; Jun 2007

item.1

name.1
1. Phrases – root words plus modifiers
2. Sections – document chapters, sections and subsections
3. Key-value pairs – tax forms, medical records, receipts, etc.
4. Tables – enumerations, tables, spreadsheets
NOUN PHRASES

AuthorizedPricingInformation

authorized.1_pricing.1_information.1

CurrentAccountBalance

current.1_account.2_balance.6

UnitPrice

unit.1_price.2

PaymentMethod

payment.2_method.2
<customer_receipt>
  <merchant>Core Communications</merchant>
  <date>14-Sep-2005 09:25:35 PM</date>
  <invoice_number>47189</invoice_number>
  <amount2>$9.95</amount2>
  <first_name>John</first_name>
  <last_name>Smith</last_name>
  <phone>978-224-9981</phone>
</customer_receipt>
<table>
<thead>
<tr>
<th>Name</th>
<th>County</th>
<th>DateOfLicensure</th>
</tr>
</thead>
</table>

**EMS Agencies Table**

- **Agency**
  - **Name**: Dixie County EMS
  - **County**: Dixie
  - **DateOfLicensure**: 2/1/2002
<table>
  <tag1>data</tag1>
  . . .
  <heading>data</heading>
  . . .
  <row>
    <column>data</column>
    . . .
  </row>
</table>
Applications
WEB SERVICE INTEGRATION

- XML Schema A
  - Database 1
  - Conflict
- XML Schema B
  - Database 2
WEB SERVICE TRANSLATION

XML Schema A
M-XML
Database 1

M-XML
XML Schema B
Database 2
<M>
  <version>1.0</version>
  <channel4>
    <title>CNN.com</title>
    <link>http://www.cnn.com/rssclick/?section=cnn_topstories</link>
    <description>
      CNN.com delivers up-to-the-minute news and information on the latest top stories, weather, entertainment, politics and more.
    </description>
    <language>en-us</language>
    <publication_date>Tue, 20 Sep 2005 18:01:37 EDT</publication_date>
    <managing_editor>editor@cnn.com</managing_editor>
    <webmaster>webmaster@cnn.com</webmaster>
    <item>
      <title>Rita.7 pounds.2 Florida_Keys.0 and.0 spawns.2 tornadoes.0</title>
      <description>
        Hurricane Rita battered South Florida and the Keys with heavy rain and strong winds Tuesday after strengthening to a Category 2 storm. Gov. Jeb Bush warned residents to stay vigilant as the storm -- with maximum sustained winds of 100 mph -- passed through the Straits of Florida without so far making official landfall. Radar indicated Rita spawned tornadoes near Hollywood, Florida, and a water spout or tornado near Islamorada, in the upper Keys.
      </description>
      <publication_date>Tue, 20 Sep 2005 16:13:49 EDT</publication_date>
    </item>
  </channel4>
</M>
'M' BROWSER
‘M’ DATA FEEDS

Data Explorer

THE DATA CENTER
Making sense of your data
‘M’ DATA FEEDS

NOAA NDBC

Raw Data Feed

YYYY MM DD hh mm WD WSPD GST WVHT DPD APD MWD BARO ATMP WTMP DEWP VIS PTDY TIDE
2005 07 11 17 50 MM MM MM 1.2 5 MM MM 1011.8 16.2 13.8 13.6 MM -0.7 MM

‘M’ Data Feed

<timestamp.1>
  2005-07-11T17:50
</timestamp.1>
<wave.5_height.2>
  1.2
  <unit.5>foot.11</unit.5>
</wave.5_height.2>
‘M’ MODEL EXPLORER