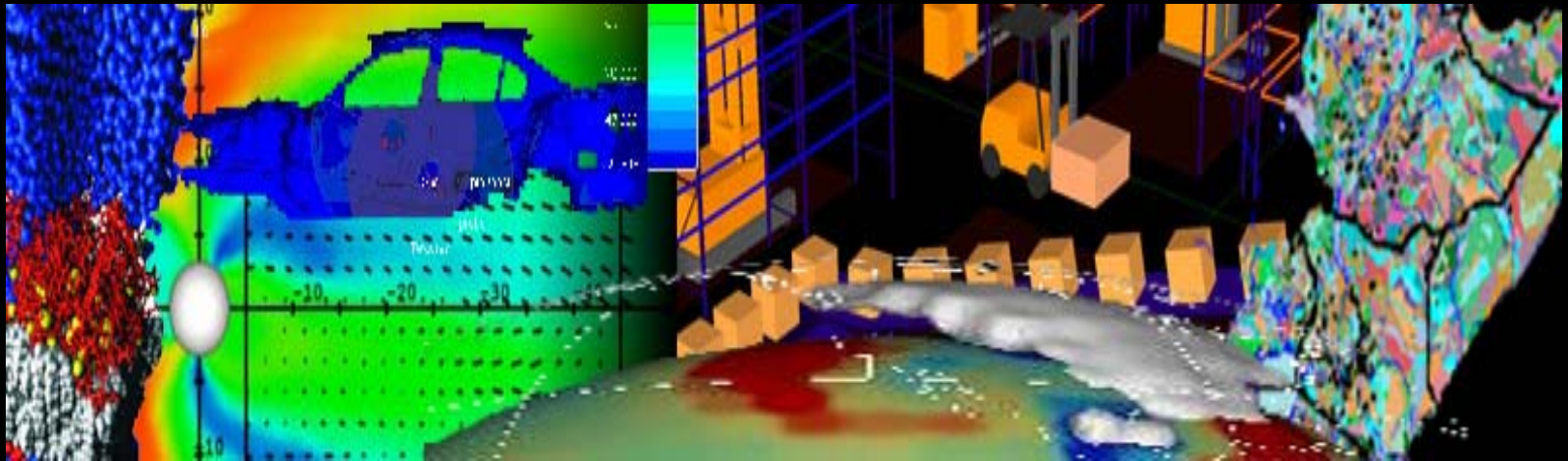


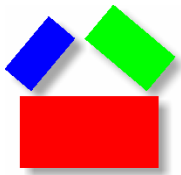
## DATA CENTER

# DATA CENTER

*Make sense of your data*

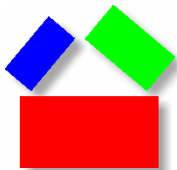


David Brock, Founder and Director  
Data Center  
Massachusetts Institute of Technology



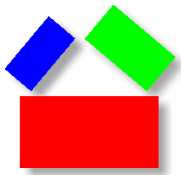
**SOLUTION**

**XML**



## PROBLEM

```
<CallsPerDay>  
    2575  
</CallsPerDay>
```

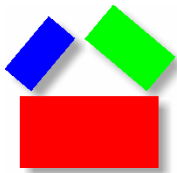


## PROBLEM

`"CallsPerDay"`

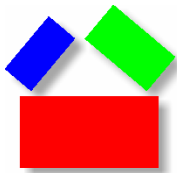
?

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>
```

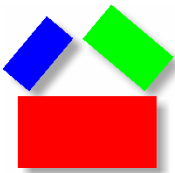


## MORE PROBLEMS

```
<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
, 2004" >
    <Calls Units="PerDay">
      2575
    </Calls>
  </CorporateRecords>
```

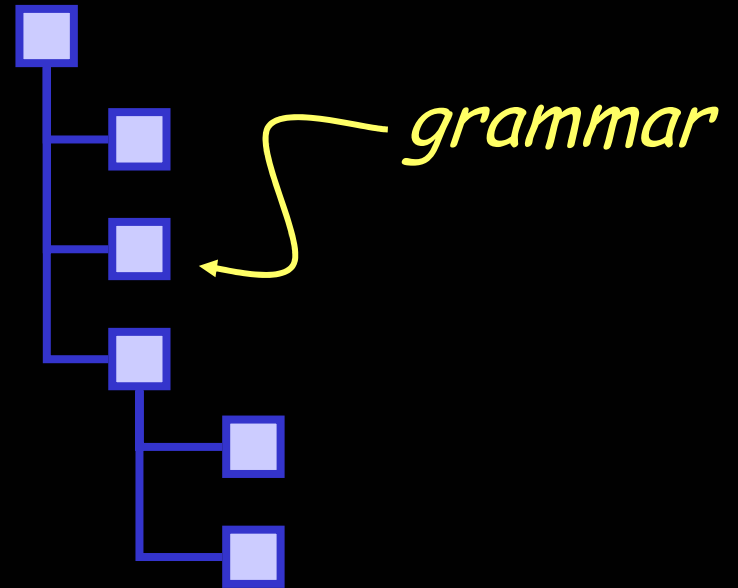


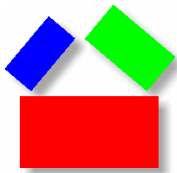
# OBSERVATION

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

"CallsPerDay"

*vocabulary*

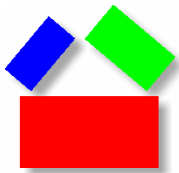




# STANDARDS?

4ML	ARML	BiblioML	CIDX	eBIS-XML	HTTP-DRP	MatML	ODRL	PrintTalk	SHOE	UML	XML F
5AML	ARML	BCXML	xCIL	ECML	HumanML	MathML	OeBPS	ProductionML	SIF	UBL	XML Key
6AML	ASML	BEEP	CLT	eCo	HyTime	MBAM	OFX	PSL	SMML	UCLP	XMLife
7AML	ASML	BGML	CNRP	EcoKnow	IML	MISML	OIL	PSI	SMBXML	UDDI	XML MP
8AML	ASTM	BHTML	ComicsML	edaXML	ICML	MCF	OIM	QML	SMDL	UDEF	XML News
9AML	ATML	BIBLIOML	Covad xLink	EMSA	IDE	MDDL	OLife	QAML	SDML	UIML	XML RPC
0AML	ATML	BIOML	CPL	eosML	IDML	MDSI-XML	OML	QuickData	SMIL	ULF	XML Schema
1ABML	ATML	BIPS	CP eXchange	ESML	IDWG	Metarule	ONIX DTD	RBAC	SOAP	UMLS	XML Sign
2ABML	ATML	BizCodes	CSS	ETD-ML	IEEE DTD	MFDX	OOPML	RDDI	SODL	UPnP	XML Query
3ACML	AWML	BLM XML	CVML	FieldML	IFX	MIX	OPML	RDF	SOX	URI/URL	XML P7C
4ACML	AXML	BPML	CWMI	FINML	IMPP	MMLL	OpenMath	RDL	SPML	UXF	XML TP
5ACAP	AXML	BRML	CycML	FITS	IMS Global	MML	Office XML	RecipeML	SpeechML	VML	XMLVoc
6ACS X12	AXML	BSML	DML	FIXML	InTML	MML	OPML	RELAX	SSML	vCalendar	XML XCI
7ADML	AXML	CML	DAML	FLBC	IOTP	MML	OPX	RELAX NG	STML	vCard	XAML
8AECM	BML	xCML	DaliML	FLOWML	IRML	MoDL	OSD	REXML	STEP	VCML	XACML
9AFML	BML	CaXML	DaqXML	FPML	IXML	MOS	OTA	REPML	STEPML	VHG	XBL
0AGML	BML	CaseXML	DAS	FSML	IXRetail	MPML	PML	ResumeXML	SVG	VIML	XSBEL
1AHML	BML	xCBL	DASL	GML	JabberXML	MPXML	PML	RETML	SWAP	VISA XML	XBN
2AIML	BML	CBML	DCMI	GML	JDF	MRML	PML	RFML	SWMS	VMML	XBRL
3AIML	BML	CDA	DOI	GML	JDox	MSAML	PML	RightsLang	SyncML	VocML	XCFF
4AIF	BannerML	CDF	DeltaV	GXML	JECMM	MTML	PML	RIXML	TML	VoiceXML	XCES
5AL3	BCXML	CDISC	DIG35	GAME	JLife	MTML	PML	RoadmOPS	TML	VRML	Xchart
6ANML	BEEP	CELLML	DLML	GBXML	JSML	MusicXML	PML	RosettaNet PIP	TML	WAP	Xdelta
7ANNOTE	ABGML	ChessGML	DMML	GDML	JSML	NAML	PML	RSS	TalkML	WDDX	XDF
8ANATML	BHTML	ChordML	DocBook	GEML	JScoreML	xNAL	P3P	RuleML	TaxML	WebML	XForms
9APML	BIBLIOML	ChordQL	DocScope	GEDML	KBML	NAA Ads	PDML	SML	TDL	WebDAV	XGF
0APPML	BIOML	CIM	DoD XML	GEN	LACITO	Navy DTD	PDX	SML	TDML	WellML	XGL
1AQL	BIPS	CIML	DPRL	GeoLang	LandXML	NewsML	PEF XML	SML	TEI	WeldingXML	MGML
2APPEL	BizCodes	CIDS	DRI	GIML	LEDES	NML	PetroML	SML	ThML	Wf-XML	XHTML
3ARML	BLM XML	CIDX	DSML	GXD	LegalXML	NISO DTB	PGML	SAML	TIM	WIDL	XIOP
4ARML	BPML	xCIL	DSD	GXL	Life Data	NITF	PhysicsML	SABLE	TIM	WITSML	XLF
5ASML	BRML	CLT	DXS	Hy XM	LitML	NLMXML	PICS	SAE J2008	TMML	WorldOS	XLIFF
6ASML	BSML	CNRP	EML	HITIS	LMML	NVML	PMML	SBML	TMX	WSML	XLink
7ASTM	BCXML	ComicsML	EML	HR-XML	LogML	OAGIS	PNML	Schemtron	TP	WSIA	XMI
8ARML	BEEP	CIM	DLML	HRMML	LogML	OBI	PNML	SDML	TPAML	XML	XMSG
9ARML	BGML	CIML	EAD	HTML	LTSC XML	OCF	PNG	SearchDM-XML	TREX	XML Court	XMTP
0ASML	BHTML	CIDS	ebXML	HTTPL	MAML	ODF	PrintML	SGML	TxLife	XML EDI	XNS



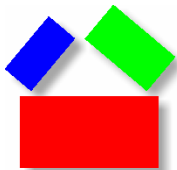


# PROPOSAL

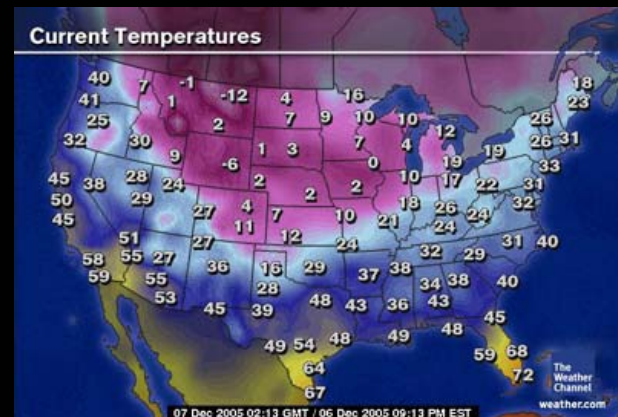
Vocabulary

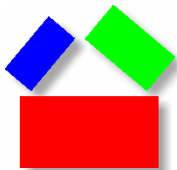
+

Words without structure **Grammar** Structure without words

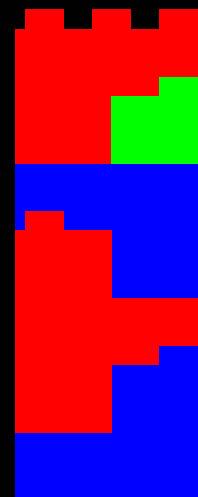
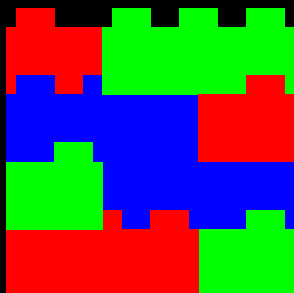
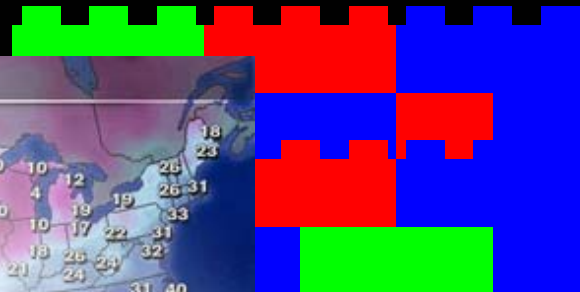
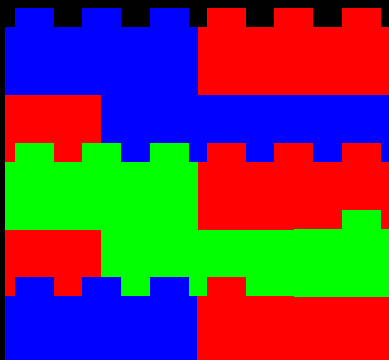


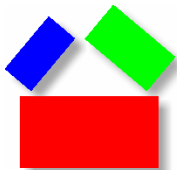
# INTEGRATE





# 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:**

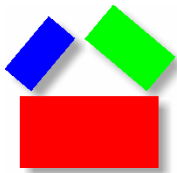
## SCHEMA

**Data:**  $^{+}[0-9]\d{2}-\d{3}-\d{4}\$$

**Attributes:** party.5, duration.1, telephone\_number.1

## LOCALIZATION

**Data:** 电话 , telefoongesprek , 전화 , телефонныйа вызов



# DICTIONARY DEVELOPMENT



Oxford  
English  
Dictionary  
OXFORD UNIVERSITY PRESS

Oxford English Dictionary



Unified  
Medical  
Language  
System

National Library of Medicine  
Unified Medical Language System



United States Department of Defense



Princeton University, WordNet



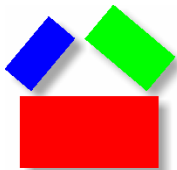
American Chemical Society  
Chemical Abstracts Service



West Law Publishing  
Black's Law Dictionary

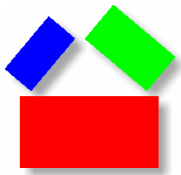


Acronym Finder  
Acronym Dictionary



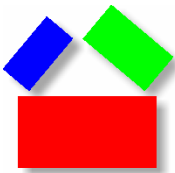
# DICTIONARY DEVELOPMENT

- 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

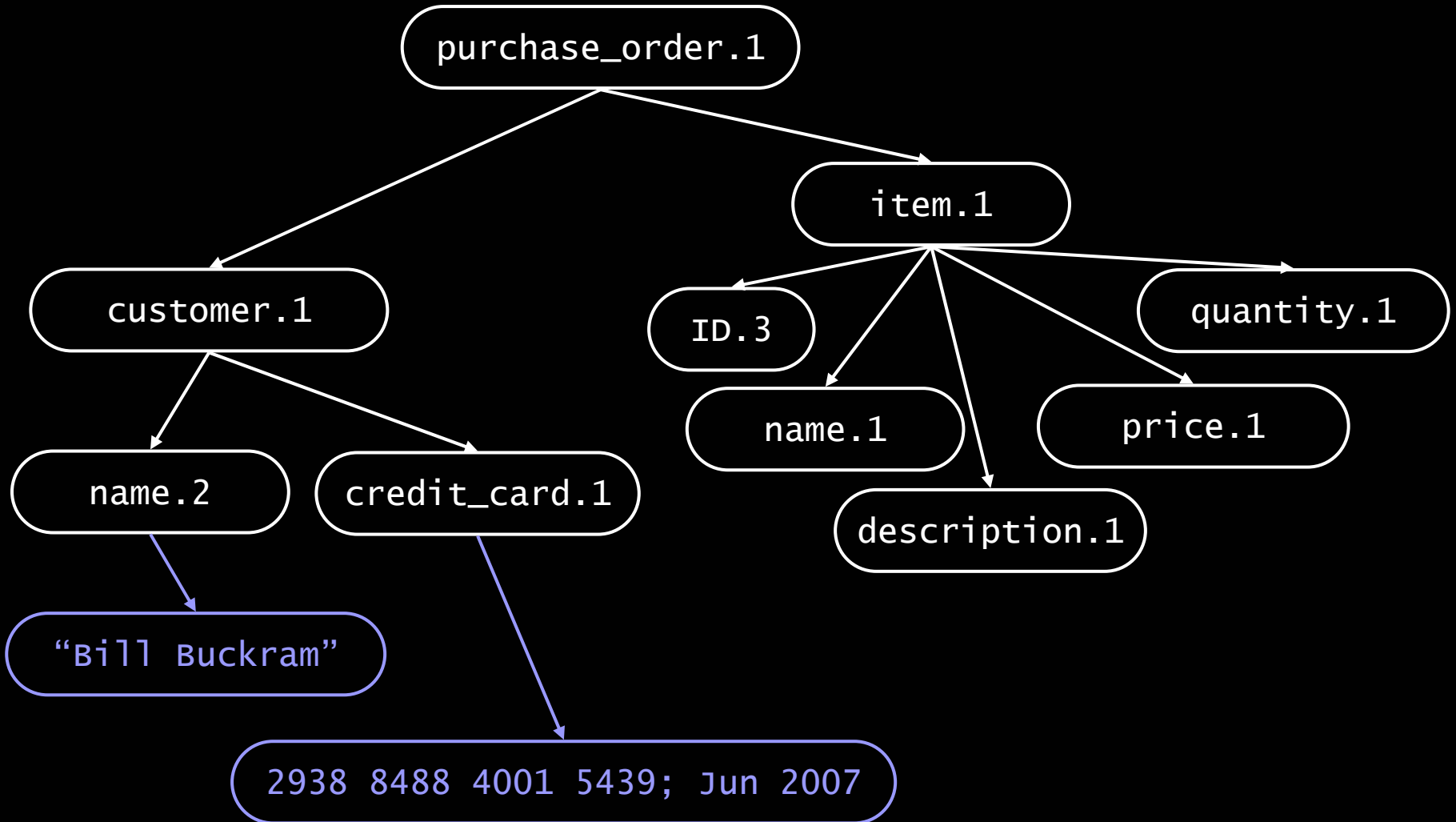


# GRAMMAR

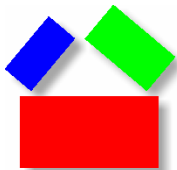
Grammar



# GRAMMAR

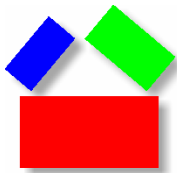






# PATTERNS

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

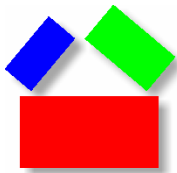
current.1\_account.2\_balance.6

UnitPrice

unit.1\_price.2

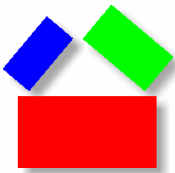
PaymentMethod

payment.2\_method.2



## KEY-VALUE PAIRS

```
<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>
```

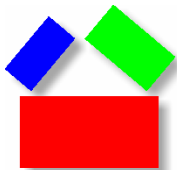


# TABLES

```
<EMS_Agencies_Table>
  <Agency>
    <Name>Dixie County EMS</Name>
    <County>Dixie</County>
    <DateOfLicensure>
      2/1/2002
    </DateOfLicensure>
  </Agency>
</EMS_Agencies_Table>
```

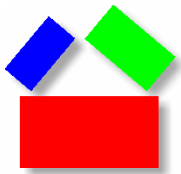
```
<Table>
  <Row>
    <Name>Edward Taylor</name>
    <Country>United Kingdom</country>
    <BirthDate>3/4/1968</BirthDate>
  </Row>
</Table>
```

```
<table>
  <title></title>
  <cols>
    <th>Name</th>
    <th>County</th>
    <th>DateOfLicensure</th>
  </cols>
  <data>
    <tr>
      <td>Edward Taylor</td>
      <td>United Kingdom</td>
      <td>3/5/1968</td>
    </tr>
  </data>
</table>
```



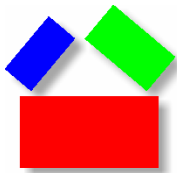
## TABLE PATTERN

```
<table>  
  <tag1>data</tag1>  
  . . .  
  <heading>data</heading>  
  . . .  
  <row>  
    <column>data</column>  
    . . .  
  </row>  
</table>
```

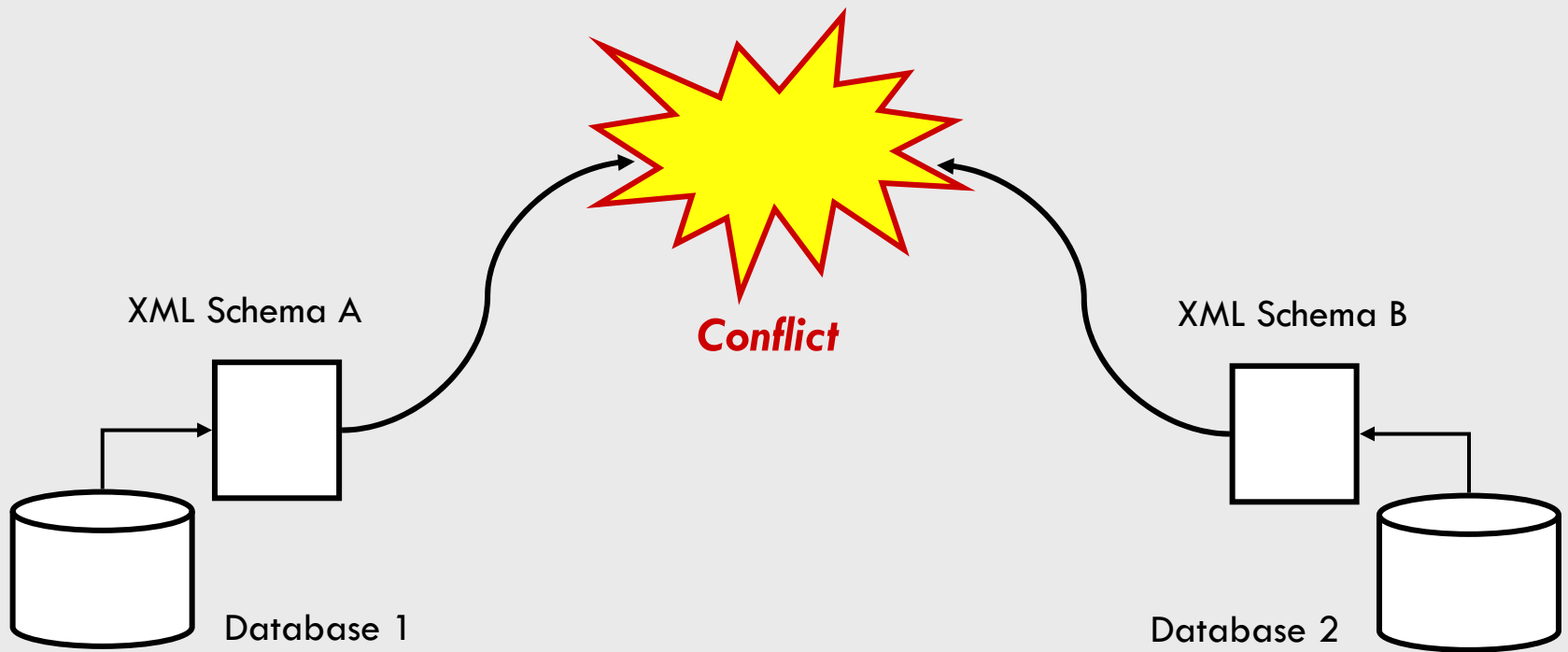


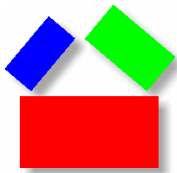
# APPLICATIONS

Applications

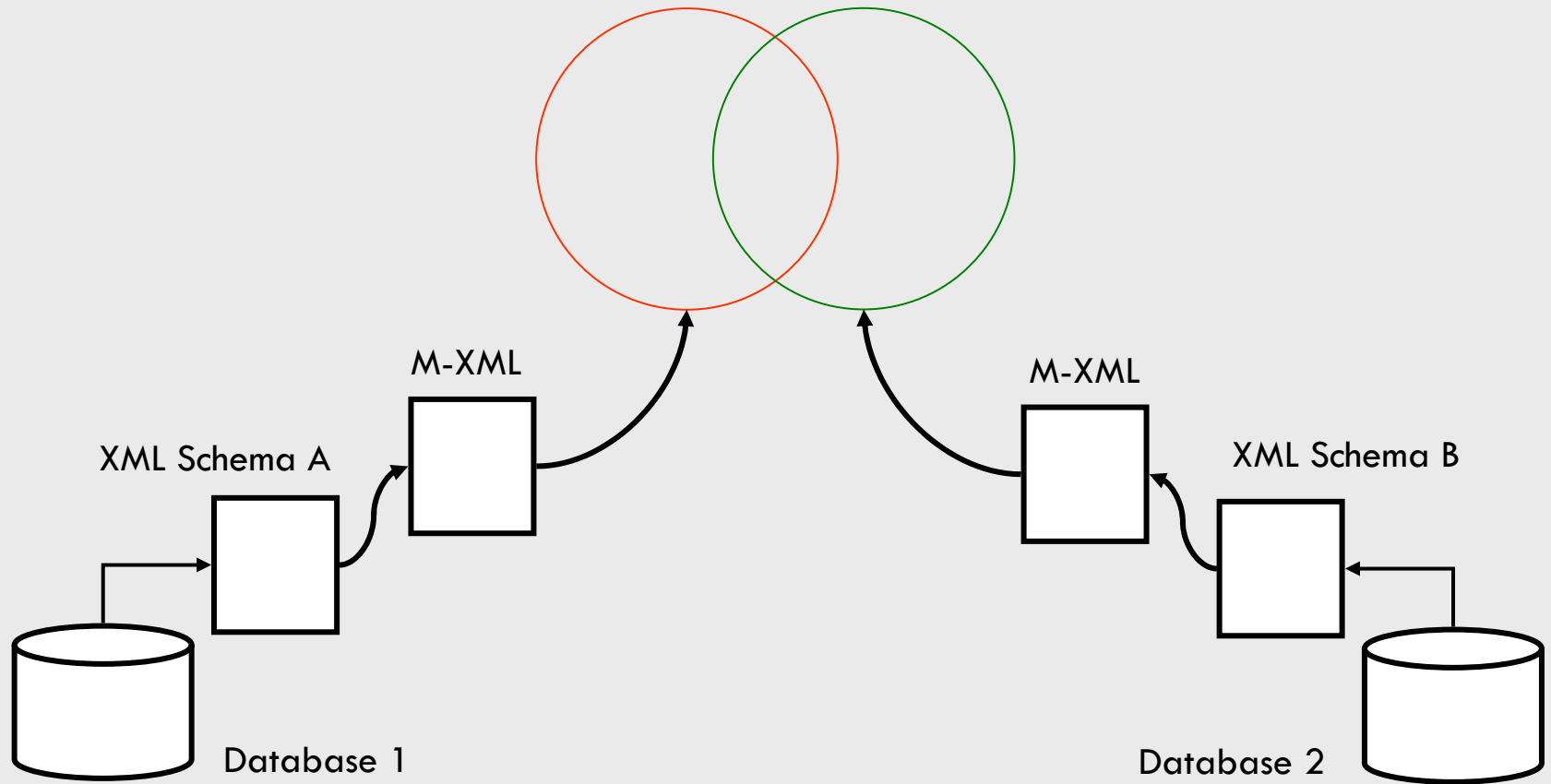


# WEB SERVICE INTEGRATION

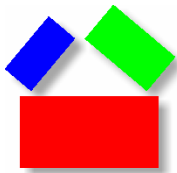




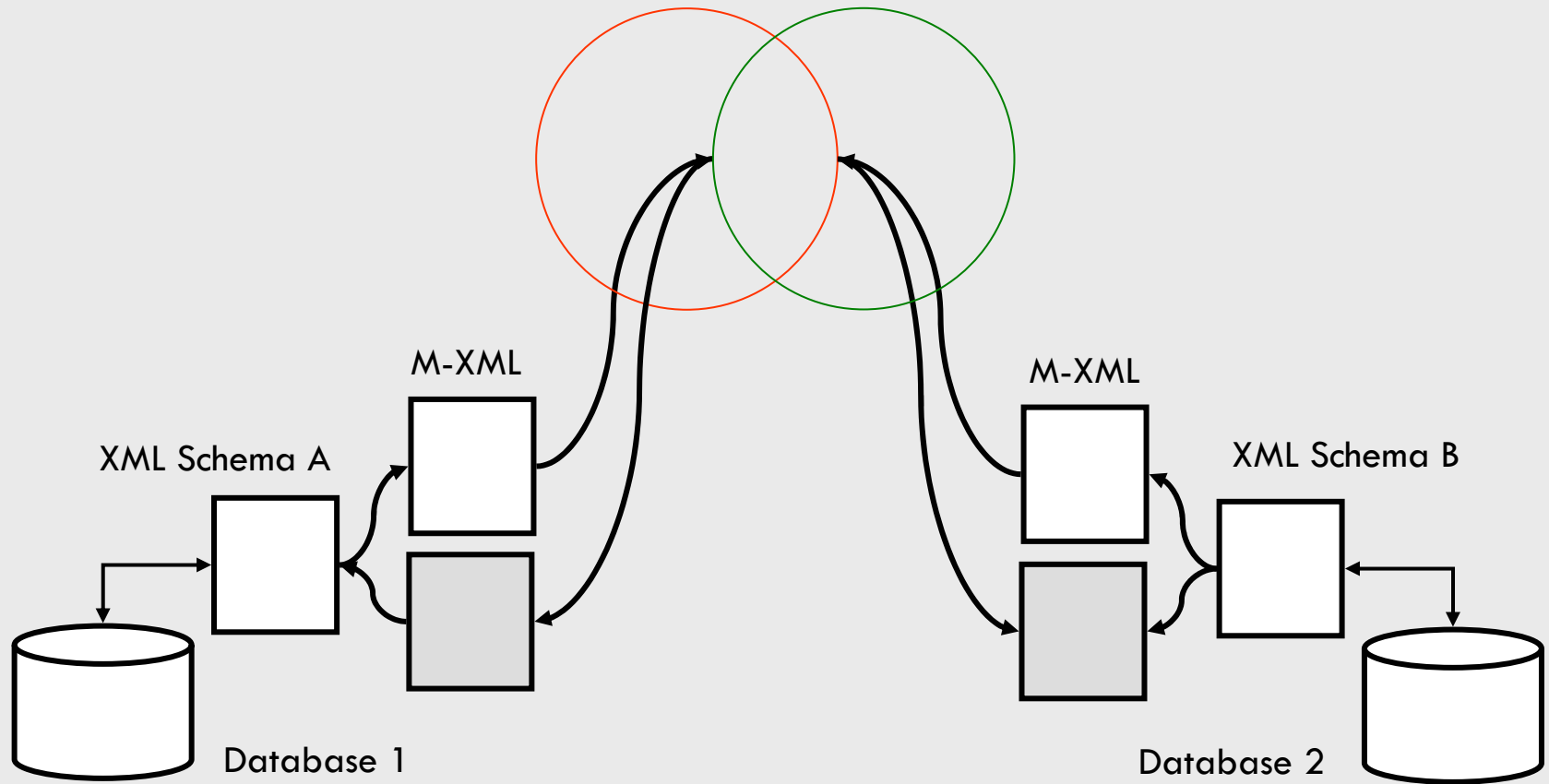
# WEB SERVICE INTEGRATION

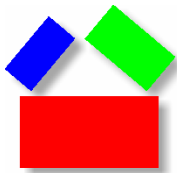






# WEB SERVICE TRANSLATION





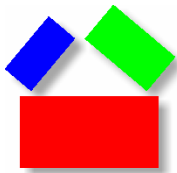
# 'M' NEWS FEEDS



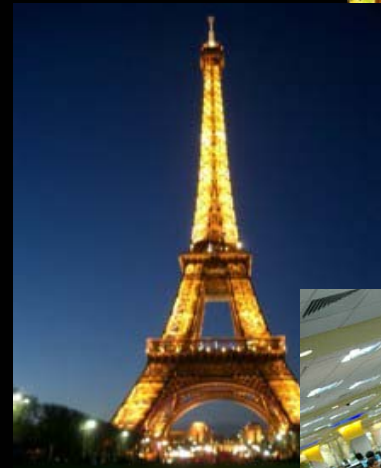
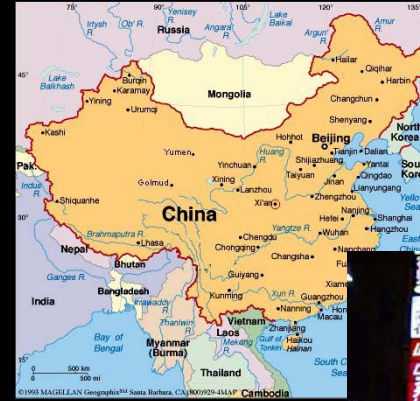
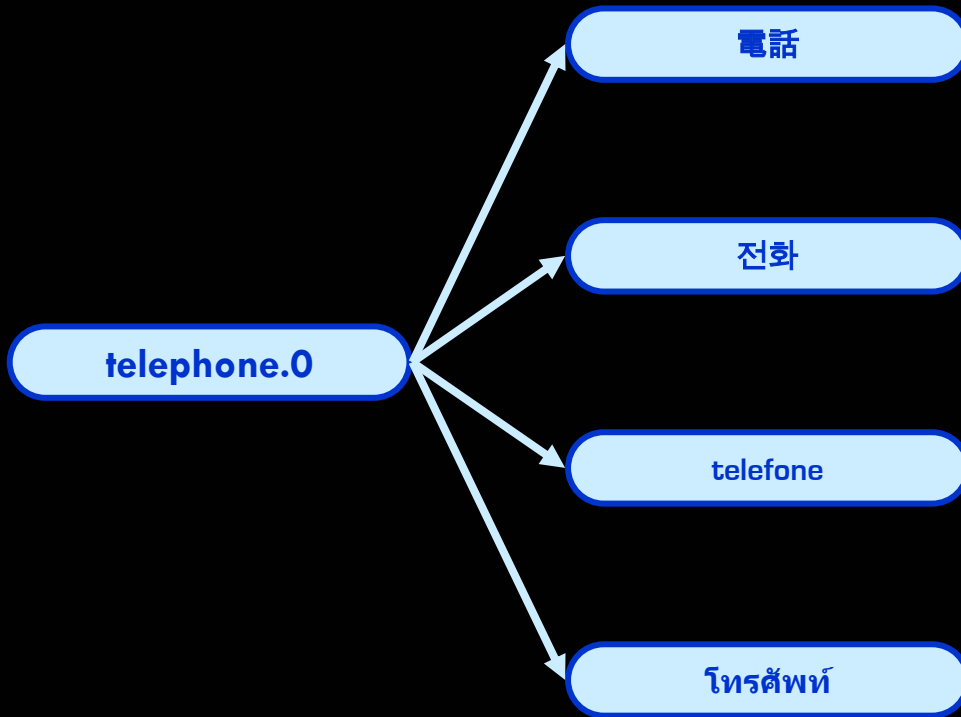
M	1.0
---	-----

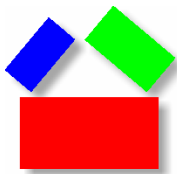


```
<M>
  <version>1.0</version>
  <channel4>
    <title>CNN.com</title>
    <link2>http://www.cnn.com/rssclick/?section=cnn\_topstories</link2>
    <description>
      CNN.com delivers up-to-the-minute news and information on the
      latest top stories, weather, entertainment, politics and
      more.
    </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>
    <link>http://www.cnn.com/rssclick/2005/Weather/09/20/rita/
      Index.html?section=cnn\_topstories
    </link>
    <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' INTERNATIONALIZATION





# 'M' BROWSER

Fusion 0.7

File Edit View Tools Help

← → ↺ ⏸ ↻ ↺ ⚙

Order

Customer

Name Bill Buckram

Credit Card 234 234 234 234

Manifest				
	Identification Number	Title	Quantity	Unit Price
▶	209	Duke: A Biography of the Java Evangelist	1	\$10.75
	208	100% Pure: Making Cross Platform Deployment a Reality	1	\$12.45
	204	Making the Transition from C++ to the Java(tm) Language	1	\$15.95
	202	Web Servers for Fun and Profit	1	\$24.85
	210	I Think Not: Dukes Likeness to the Federation Insignia	1	\$8.85

Receipt

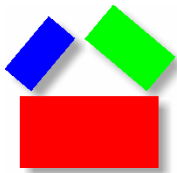
Subtotal \$72.85

Tax \$4.74

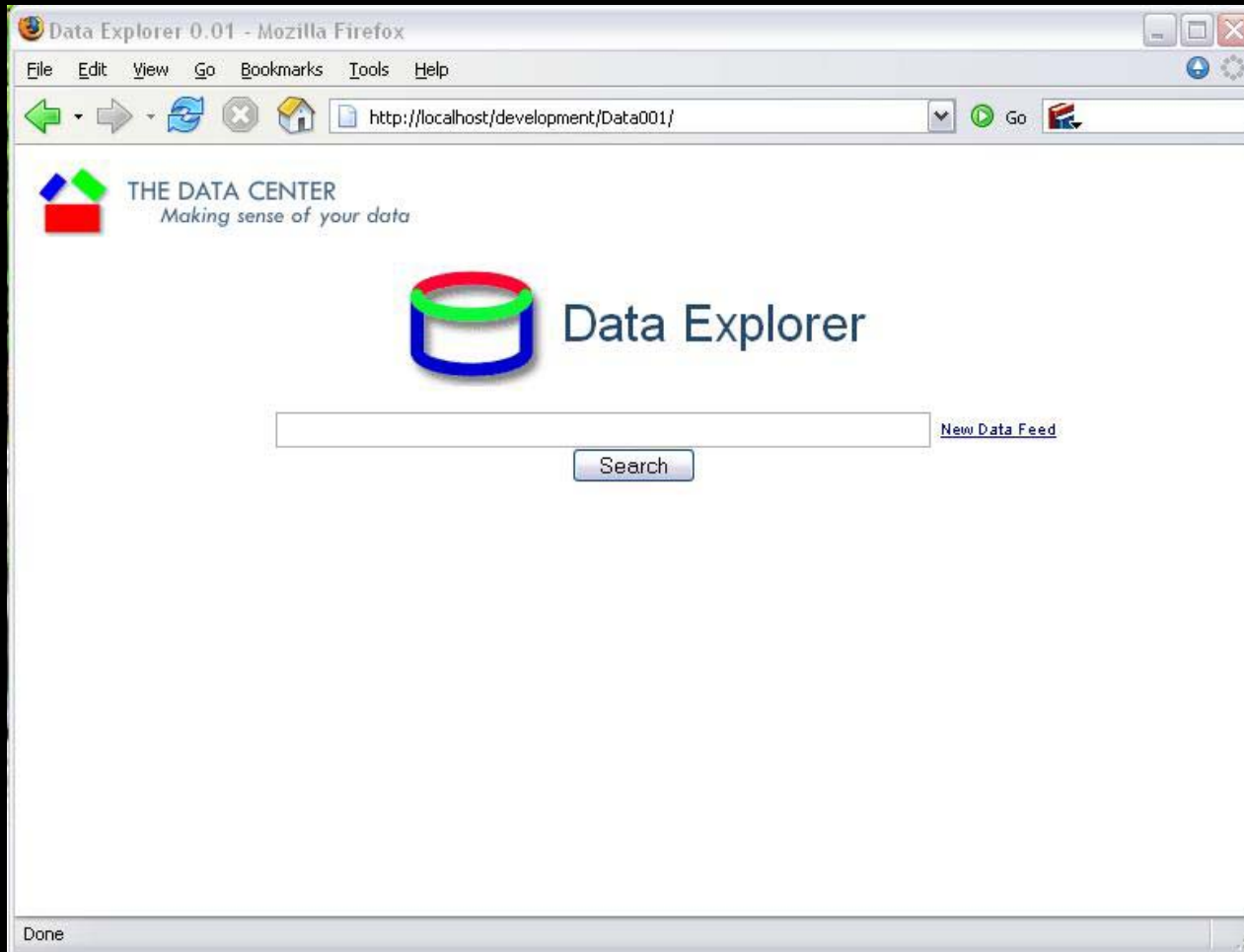
Total \$77.59

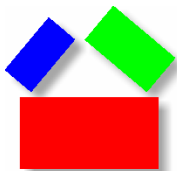
Status Ready...

M 1.0



# 'M' DATA FEEDS





# 'M' DATA FEEDS

## NOAA NDBC

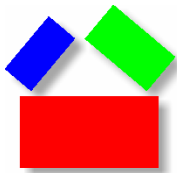
### Raw Data Feed

```
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

