Digital Overload—and What to Do about It

Jim Larson
Co-program Chair
SpeechTEK Conference
May 23-25, 2016, Washington, DC
Digital Overload

• We consume 3x more information as in 1960.
  — University of California, San Diego

• The average computer user checks 40 websites a day
  — New York Times
The Big Picture

Query via speaking or typing

Browse results

Data management

© 2016 Larson Technical Services
How We Access Data

• Search
  – Query
    • User knows what information he/she needs
    • Formulates a request

• Browse
  – User will recognize desired information when he/she sees it
  – Selects an area likely to contain results and examines content

© 2016 Larson Technical Services
Digital Overload—and What to Do about It

• Look at disciplines that solve the digital overload problem
  – Database management systems
  – Library systems
  – Recipe collections
  – Sheet music collections
  – Etc.
How We Access Data

Organize
Query
Evaluate
Browse
How We Access Data

- Create ontology

Organize
Query
Evaluate
Browse
Example Ontology

• Concept names for real-life things
• Relationships among concepts
Another Ontology Format: Entity-Relationship Diagrams

- **Person**
  - Name

- **Employee**
  - Number
  - Name
  - Salary
  - Hire Date

- **Customer**
  - Name
  - Address
Why Use an Ontology?

- Use ontology terms to describe data

Employee

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Salary</th>
<th>Hire Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Able</td>
<td>48,000</td>
<td>6-18-2013</td>
</tr>
<tr>
<td>7</td>
<td>Baker</td>
<td>55,000</td>
<td>8-20-2014</td>
</tr>
<tr>
<td>19</td>
<td>Carson</td>
<td>88,000</td>
<td>5-03-2012</td>
</tr>
</tbody>
</table>

© 2016 Larson Technical Services
Why Use an Ontology?

- Use ontology terms to describe data

```xml
<employee>
  <number> 5 </number>
  <name> Able </name>
  <salary> 48000 </salary>
  <hire date> 6-18-2013 </hire date>
<employee> Able </employee>
```
How We Access Data

Organize
Query
Evaluate
Browse

• Create ontology
• Create metadata
Metadata

• Metadata is data about data
  – Card catalog in library
  – Table of contents and index in book
  – Names of columns in spreadsheet
  – Tags in XML description

• Metadata uses terms from an ontology to describe data
<employee>
  <number>5</number>
  <name>Able</name>
  <salary>48000</salary>
  <hire date>6-18-2013</hire date>
<employee> Able </employee>
How to Create Metadata?

• Manually
  – Tagging data fragments is tedious, time consuming, and expensive

• Clever AI algorithms
  – AI learning systems can be trained to tag data
How We Access Data

- Select data set

Organize
Query
Evaluate
Browse
Identify Source

• Index of indexes
• Cross indexes
• Query router
How We Access Data

- Select data set
- Formulate request using metadata terms

Organize
Query
Evaluate
Browse

© 2016 Larson Technical Services
Query Formats

• Use ontology terms to formulate queries and commands

**SQL**
select name
from employee
where salary > 50,000;

**Natural Language**
Get employees with salary greater than 50,000

---

**Query By Example**

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Salary</th>
<th>Hire Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.</td>
<td></td>
<td>&gt;50000</td>
<td></td>
</tr>
</tbody>
</table>

May be spoken or typed
Review and Modify Related Queries

```markdown
ontology
ontology
ontology **web language**
onology **language**
onology **defined**
onology **and epistemology**
onology **synonym**
onoteny recapitulates phylogeny
ontology **epistemology axiology**
onology **example**
**Ontology** - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Ontology

Google Search  I'm Feeling Lucky
```

© 2016 Larson Technical Services
How We Access Data

- Organize
- Query
- Evaluate
- Browse

• Apply to sample data
• Examine content
## Compare Result to Small Data Set

### Employee

<table>
<thead>
<tr>
<th>Number</th>
<th>Name</th>
<th>Salary</th>
<th>Hire Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Able</td>
<td>48,000</td>
<td>6-18-2013</td>
</tr>
<tr>
<td>7</td>
<td>Baker</td>
<td>55,000</td>
<td>8-20-2014</td>
</tr>
<tr>
<td>19</td>
<td>Carson</td>
<td>88,000</td>
<td>5-03-2012</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Result

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker</td>
</tr>
<tr>
<td>Carson</td>
</tr>
</tbody>
</table>
How We Access Data

Organize
Query
Evaluate
Browse

• Apply to sample data
• Examine content
• Refine request
Incorporate Result into Another Request

AddressBook

<table>
<thead>
<tr>
<th>Name</th>
<th>City</th>
<th>State</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Able</td>
<td>New York City</td>
<td>NY</td>
<td>00503</td>
</tr>
<tr>
<td>Baker</td>
<td>Portland</td>
<td>OR</td>
<td>97006</td>
</tr>
<tr>
<td>Carson</td>
<td>St Paul</td>
<td>MN</td>
<td>23456</td>
</tr>
</tbody>
</table>

Result

<table>
<thead>
<tr>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baker</td>
</tr>
<tr>
<td>Carson</td>
</tr>
</tbody>
</table>

Select city

From AddressBook, Result

Where Result.Name = AddressBook.Name;

May be spoken or typed
How We Access Data

- Extract intent

Organize
Query
Evaluate
Browse
Extract Intent Using VoiceXML Grammar Rules

• Example grammar rule with Script Syntax:

```xml
<rule id="action">
  <one-of>
    <item> small <tag> out.size = "small"; </tag> </item>
    <item> medium <tag> out.size = "medium"; </tag> </item>
    <item> large <tag> out.size = "large"; </tag> </item>
  </one-of>
  <one-of>
    <item> green <tag> out.color = "green"; </tag> </item>
    <item> blue <tag> out.color = "blue"; </tag> </item>
    <item> white <tag> out.color = "white"; </tag> </item>
  </one-of>
</rule>
```

• ECMAScript structure:

```javascript
action: {
  size: "large",
  color: "white"
}
```
Extract Intent Using Wit.ai

• Name the intent
  – Example: temperature_get

• Provide several expressions for how users request the intent
  – What is the temperature?
  – How hot is it?
  – How cold is it?
  – Can you tell me the temperature?

• Map user request to an intent
  – “What’s the temperature” -> temperature_get
How We Access Data

Organize

Query

Evaluate

Browse

• Display
• Page
• Scroll
• Zoom
• Sort and manipulate
• Etc.
Larson’s Manifesto

1. Data suppliers
   – Use standard ontologies
   – Create self-describing data
   – Use standard metadata formats

• List of ontologies
  – [https://www.w3.org/wiki/Lists_of_ontologies](https://www.w3.org/wiki/Lists_of_ontologies)

• List of XML languages

• List of metadata standards
Larson’s Manifesto

2. System developers enable users to:
   – Use aids for formulating queries
   – Review and modify other people’s queries
   – Review result, resolve ambiguities
   – Iteratively refine query
Questions