Question Answering in the InfoSphere: Semantic Interoperability and Lexicon Development

Paul Thompson
and
Steven Lulich

Institute for Security Technology Studies
Dartmouth College
15 May 2002
Outline

• The Joint Battlespace Infosphere
• Fuselets
• DARPA IXO Sensor Networks
• Multiple Query Optimization
• Semantic Interoperability and Lexicon Development
• Conclusions
Basic “Infosphere” concept

• An information broker will offer essential information exchange services

IM Policy Maker

‘Infosphere’ essential services
(A core capability enabling enabling Information Management among systems)

Likely the same set

An existing (or new) System
(Come as you are…almost)

The Global Grid

Another existing (or new) System
(Come as you are…almost)

Info Source

Info Consumer
Information Interoperability using XML

XML-JBI

- Personnel
- BDA
- Orders of Battle
- Weather
- Intentions
- Targets
- Etc....

XML Info Object

- TBMCS
- ABCS
- IBS
- AFATDS
- GCSS

C4ISR Databases

Human Interfaces

Application Interfaces

Data Interfaces

Mobile Interfaces
Dartmouth Physical Sensor (Version 1)

GPS Location Sensor (Motorola)

RF Technologies Wireless Networking

Processor (Intel 8051)

Dallas Semiconductor I-Buttons

DURIP funded 100+ such devices

Other efforts: Crossbow, UC Berkeley, UCLA,....
Challenge Problem

Q1: “How many people are in the room?”

Q2: “How many people are in the room?”

Q3: “How many people are talking?”

Q4: “How many people are listening?”
Ingredients ("Sensor Net OSI Layers?")

| Q1   | \( (q(1,1)&q(1,2)) | (q(1,3)&q(1,4)) \) |
|------|----------------------|
| Q2   | \( (q(2,1)&q(2,2)) | (q(2,3)&q(2,4)) \) |
| Q3   | \( (q(3,1)&q(3,2)) | (q(3,3)&q(3,4)) \) |
| Q4   | \( (q(4,1)&q(4,2)) | (q(4,3)&q(4,4)) \) |

Query-Network-Sensor Mapping and Fusion:
subqueries mapped to network nodes to satisfy all the queries (optimally?)

Sensor Resources:
- s1 \((x,y,z)\) acoustic
- s2 \((x,y,z)\) motion
- s3 \((x,y,z)\) IR

Sensor Routing Tables:
- s1  s5
- s2  s6
- s3  s7
- ....

NLP  Distributed Query Optimization  Resource Discovery  Sensor Markup Language  Ad hoc routing
Ingredients - Score Card

<table>
<thead>
<tr>
<th>Q1</th>
<th>(q(1,1)&amp;q(1,2))</th>
<th>(q(1,3)&amp;q(1,4))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(q(2,1)&amp;q(2,2))</td>
<td>(q(2,3)&amp;q(2,4))</td>
</tr>
<tr>
<td></td>
<td>(q(3,1)&amp;q(3,2))</td>
<td>(q(3,3)&amp;q(3,4))</td>
</tr>
<tr>
<td></td>
<td>(q(4,1)&amp;q(4,2))</td>
<td>(q(4,3)&amp;q(4,4))</td>
</tr>
</tbody>
</table>

Query-Network-Sensor Mapping and Fusion:

Subqueries mapped to network nodes to satisfy all the queries (optimally?)

Sensor Resources:
- s1 (x,y,z) acoustic
- s2 (x,y,z) motion
- s3 (x,y,z) IR

Sensor Routing Tables:
- s1
- s2
- s3
- s5
- s6
- s7

NLP  | Distributed Query Optimization | Resource Discovery | Sensor Markup Language | Ad hoc routing
---   |-------------------------------|--------------------|------------------------|------------------
Hard | NP Hard, approximate?          | Scalability in time and numbers? | Semantics, consistency, dynamics? | Scalability in time and numbers?
Natural Language Processing Level

• “Question and Answer” systems
  – TREC Question Answering Track, eg
  – “Who won the women’s downhill gold medal?”

• “Message Understanding”
  – Bounded domains - military intelligence, logistics, etc
  – message retrieval using information extraction

• Inference capability
Semantic Interoperability and Lexicon Development

• DARPA hand-built Ontologies and KBs
  – DARPA Agent Markup Language (DAML)
  – Rapid Knowledge Formation (RKF)

• Dynamic Ontology Mapping
  – Learning component
  – Connectivistic Lexicon
Distributed Query Optimization Layer

**User 1:** How many fuel trucks are there in region x?

Parsed to a boolean search expressed in terms of sensor outputs and primitive operations.

\[ q(u1, si, opj) \& q(u1, sk, opn) | \ldots \]

**User 2:** Where are the armored personnel carriers in region y? (y intersects x)

\[ q(u2, si2, opj2) \& q(u2, sk2, opn2) | \ldots \]

Implement this in a network of fuselet servers using relationships, redundancies, inclusions, etc to get good performance according to some metric.
Effectiveness and Efficiency of a Mapping

$q(1,1,\text{op1}) \& q(1,2,\text{op2})$  \& means downstream flow is smaller

$q(2,1,\text{op3}) \mid q(2,2,\text{op4})$  \mid means downstream flow is the same

$q(1,2,\text{op2}) < q(2,2,\text{op4})$

$q(\text{client, sensor, operation})$

Fusion nodes of different loads/power

Links of different latencies/bw
Resource Discovery

- JINI, CORBA, Brokers, Matchmakers, etc
- Distributed, coherency, overheads, etc
- Semantics?
- Indexing?
- Scalability of distributed, dynamic implementation
Sensor Markup Language (XML see DARPA DAML) - need “evidential reasoning glue”?

Land-based vehicle tracking system

Logistics/infrastructure monitoring of an adversary

Is there a fuel truck at x?

P(fuel truck at x) = P(A)

P(A&B)

P(A|B)

P(fuel truck delivery needed at y) = P(B)

Fuel truck

Fuel truck

semantics ok

P(A&B)
Ad hoc routing in dynamic environments

- nodes dynamic
- requirements dynamic
- tradeoff between management overhead and usable BW/latency?
- sensor networks: bi-directional vs unidirectional links?
- lots to do still
C4ISR for Tactical Air-Land Combat System Concept
Nested, Closed-loop Control of Highly Automated Forces

- Reachback exploitation, assessment, and planning (semi-automated)
- Long distance information connectivity
- Forward/organic real-time exploitation and targeting (highly automated)
- Highly automated forces

Mission objectives
- Mission objectives
- Refine plan
- Current mission status
- Assess plan
- Entity locations and identities
- Exploit data
- Target sets, constraints, context data
- Results, ambiguities
- Target sets, constraints, context data
- Results, ambiguities
- Entity locations and identities
- Track and identify
- Sensor data
- Engage targets
- Look, move, and fire controls
- Fight
- Refine plan
- Comms
- Comms
- Comms
- Comms
Conclusions (cont.)

• Question answering from JBI and sensor networks
  – Different issues from ARDA / TREC
  – Can build on
    • ARDA / TREC
    • NLP interfaces to RDBMS and Knowledge Bases
    • Data mining from stream data
• Semantic Interoperability is crucial