Enhancing Information Awareness through Directed Qualification of Semantic Relevancy Scoring Operations

Jason Bryant, Greg Hasseler, Matthew Paulini, and Tim Lebo

Air Force Research Laboratory
Information Directorate/RISA
Rome, NY USA
Outline

• Motivation
• Background
• Semantic Mission Domain
• Design
• Applied Use Case
• Results
• Conclusion
Motivations

• Representation: A common representation for analytic scoring and queries.

• Trust: Mission domain solutions for semantic models of analytics provenance and results.

• Features: Enable best-fit technology tiers of analytics to evaluate document relevancy and order query results.
Background

• Information environment complexity
• Relevancy solutions for big data
• Multi-structured analytics
• Interleaved information events, mission assets, and knowledge
Web Domain

Web Pages

Content

Job Sameness

workField

Reasoning

knowsFriend

Inferences

Person Sameness

Place Sameness

Ontologies
Inferencing & Reasoning

Named Graphs

Information A

Information B

Information C

Information D

Sorted Results (Mission)

Information A

Information B

Information C

Information D
Mission Domain Directed Qualification

Named Graphs

- Information A
- Information B
- Information C
- Information D

Sorted Results (Relevancy)

- Information A
- Information C
- Information D
- Information B

Graphs and relationships:
- Information A to Information B: Publisher Geospatial Nearness
- Information A to Information C: Mission Asset Relevancy
- Information A to Information D: VSM Target Similarity
- Information B to Information C: Asset Geospatial Nearness
- Information B to Information D: HITS Relevancy
- Information C to Information D: HITS Relevancy
Enable trend-based analytics queries = Analytics + PROV-O W3C + Relevancy Ontology
Ordered Information Results

- Semantic Metadata Extraction
- Publication Information
- Query Submission
- Semantics
- Semantic Query Results
- Relevancy Scoring
- Shared Referential Key/URI
- Document Store
- Analytics Graph
- Quad Store
- Submission
General Use Case

Analytic Score \(\xrightarrow{prov:\text{wasGeneratedBy}}\) Analytic \(\xrightarrow{prov:\text{used}}\) Publication/Doc

Scoring Attributes

PROV-O & RELEVANCY DOMAIN

ANALYTICS DOMAIN

Analytic Qualifications
Multiple Source Use Case

**Dictionary**

<table>
<thead>
<tr>
<th></th>
<th>V(d1)</th>
<th>V(d2)</th>
<th>V(d3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mission127</td>
<td>0.995</td>
<td>0.992</td>
<td>0.848</td>
</tr>
<tr>
<td>Target12</td>
<td>0.086</td>
<td>0.120</td>
<td>0.465</td>
</tr>
<tr>
<td>JTAC4</td>
<td>0.017</td>
<td>0</td>
<td>0.250</td>
</tr>
</tbody>
</table>

**Results:**
- \( \text{Sim}(d1,d2) = 0.998 \)
- \( \text{Sim}(d1,d3) = 0.888 \)
Results: Result Set Relevancy

Named Graphs

- Information A
- Information B
- Information C
- Information D

Relevancy Scores

Information A
- PageRank: 2.44358
- HITS: 3.67681
- VSM: 0.819324
- Geospatial Nearness: 314.16
- Role Prioritization: 8

Information B
- PageRank: 1.11258
- HITS: 2.67681
- VSM: 0.219324
- Geospatial Nearness: 716.16
- Role Prioritization: 7

Information C
- PageRank: 3.35812
- HITS: 4.67681
- VSM: 0.889112
- Geospatial Nearness: 312.4
- Role Prioritization: 9

Information D
- PageRank: 4.36622
- HITS: 4.12770
- VSM: 0.311923
- Geospatial Nearness: 288.91
- Role Prioritization: 9

Sorted Results (VSM)

1. Information C
   - PageRank: 3.35812
   - HITS: 4.67681
   - VSM: 0.889112
   - Geospatial Nearness: 312.4
   - Role Prioritization: 9

2. Information A
   - PageRank: 2.44358
   - HITS: 3.67681
   - VSM: 0.819324
   - Geospatial Nearness: 314.16
   - Role Prioritization: 8

3. Information D
   - PageRank: 4.36622
   - HITS: 4.12770
   - VSM: 0.311923
   - Geospatial Nearness: 288.91
   - Role Prioritization: 9

4. Information B
   - PageRank: 1.11258
   - HITS: 2.67681
   - VSM: 0.219324
   - Geospatial Nearness: 716.16
   - Role Prioritization: 7
Conclusion

Stepping stone to:

• Trend-based analytic queries
• Analytics inferencing and rule engine
• Enable a common representation for content, graph, and semantic-based analytic results.
• Interlinking domain knowledge, raw data, knowledge graphs, and analytic results.
• Query Intelligence