The Open World Assumption

or

Sometimes it's nice to know what we don't know

Nick Drummond, Rob Shearer
Overview

‣ What is the Open World Assumption?
‣ OWA in relation to the Semantic Web
‣ Choosing between OWA and CWA
‣ Some examples
‣ Closure of the Open World
‣ Discussion
This talk is not...

- OWL vs Databases
- Databases deal with HOW data is stored - you can store OWL in databases
- OWL is about representing knowledge with machine understandable semantics
In the beginning...

- Closed World Systems require a place to put everything
- You can’t say anything until there’s somewhere to say it
  - Slot on a frame, field on an OO class, column in a DB
- We state what is possible
In the beginning...

- When we have an empty OWL ontology, everything is possible
- We then constrain an ontology iteratively, making it more restrictive as we go
- We state what is not possible

Pig → Animal and (hasLimbs only Leg)
Negation as Failure (NaF)

- Can pigs fly?
- In CWA, because the table doesn’t contain this fact, we assume false
- In the OWA, unless we have a statement (or we can infer) “pigs can/cannot fly” we return “don’t know”
  - NaF - only false if “not(pigs can fly)”

<table>
<thead>
<tr>
<th>Animal</th>
<th>Can Fly?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penguin</td>
<td>No</td>
</tr>
<tr>
<td>Shark</td>
<td>No</td>
</tr>
<tr>
<td>Hummingbird</td>
<td>Yes</td>
</tr>
</tbody>
</table>
What is the Semantic Web?

- A vision of a computer-understandable web
- Distributed knowledge and data in **reusable** form
Semantic Web Languages

- On the Semantic Web, we expect people to **extend** our models

- But we don’t want to worry in advance how...
Incomplete Information

- The OWA assumes incomplete information by default
- We can intentionally underspecify and allow others to reuse and extend
  - eg All sharks liveInHabitat some WaterHabitat
  - Are there fresh/seawater sharks?
  - Do we care? Someone might
- It can be useful to reuse
Reuse is good

- Be more specific when the application demands it
- In OWL, we extend an ontology by adding statements. i.e., we cannot take any away
- By only committing to an answer if there is a statement to back it up, OWL remains monotonic
  - if we extend an ontology, all existing true statements remain true
Unique Name Assumption (UNA)

- If 2 things have different names (IDs) they are, by default, different
- But...

![Illustration of teeth with text labels in different languages: Gnashers, Teeth, Dents, Pearly Whites, Dientes, Zähne, Denti]
Unique Name Assumption

- CWA typically makes the UNA
  - Useful for counting
- OWA doesn’t make the UNA
  - To allow later assertion that two things are the same or different (or this may be inferred)
  - note: negation is required for distinctness
    - RDF cannot make assertions about things being different
    - OWL and other logics can
Terminology note - “Constraints”

- Much confusion
- Can mean...
  - Integrity constraints (CWA)
    - prevent “incorrect” values from being asserted in a model
    - used for validation/parsing/data input
    - single model that contains only the facts asserted
  - logical axioms (OWA)
    - eg restrictions, property domain/range
    - everything can be true unless proven otherwise
    - multiple possible models can satisfy the axioms
    - this may cause some unintuitive inferences
How do we choose?

- Not always clear cut
- Many problem domains have aspects of both

<table>
<thead>
<tr>
<th>Open World Problem</th>
<th>Closed World Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does Nick Drummond know Chris Date?</td>
<td>Is there a train from Manchester to Edinburgh today? (only x trains and y destinations)</td>
</tr>
<tr>
<td>What are the potential side-effects of drug X? (may be others than stated in a given resource)</td>
<td>Find me drugs that are not licensed for X? (would need closure for each)</td>
</tr>
<tr>
<td>Others??</td>
<td>Others??</td>
</tr>
</tbody>
</table>
Why the Open World?

- Underspecification
  - abstract, nested and unnamed entities
- Easily reusable (and extendable)
- Good at knowledge level (Ontology)
- Good at “schema”-“schema” mapping
  - eg asserting/inferring equivalents
- They naturally deal with incomplete information
  - eg Domain knowledge (eg science) - where we don’t know all of the answers yet
- Inference???
Why not(Open World)?

- Paradigm shift
  - Involves technology/experience catch up
- Some problems are inherently closed world (often those that we ask “which are not...” or have a finite number of elements)
  - but is possible to close the open world (later)
- Dealing with defaults/exceptions
- CWA good at dealing with schema-data mapping
  - integrity constraints, validation (parsing, form generation)
  - Data structures are typically closed
- Meta-query
  - What do we know???
Interpreting Knowledge

- Is there a speaker at tea/coffee?
- Are there going to be biscuits at this meeting?

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>09:00</td>
<td>Welcome</td>
<td>Jessie Kennedy</td>
</tr>
<tr>
<td>9:10</td>
<td>Data webs: new visions for research</td>
<td>David Shotton</td>
</tr>
<tr>
<td>9:40</td>
<td>Closed World Assumption</td>
<td>Chris Date</td>
</tr>
<tr>
<td>10:25</td>
<td>Open World Assumption</td>
<td>Nick Drummond</td>
</tr>
<tr>
<td>10:40-11:00</td>
<td>Tea/Coffee</td>
<td></td>
</tr>
<tr>
<td>11:00</td>
<td>The Semantic Gap between Databases and Ontologies</td>
<td>Catherine Dolbear</td>
</tr>
<tr>
<td>11:30</td>
<td>Nullogy</td>
<td>Chris Date</td>
</tr>
</tbody>
</table>

- Database says “No”
- OWA says “Don’t know” unless a blank is interpreted as “Activity and not(hasSpeaker)”
Interpreting Knowledge

- I want to treat my patient with a painkiller that is not an anticoagulant.

<table>
<thead>
<tr>
<th>Drug</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin</td>
<td>Painkiller</td>
</tr>
<tr>
<td>Wharfarin</td>
<td>Anticoagulant</td>
</tr>
<tr>
<td>Paracetemol</td>
<td>Painkiller</td>
</tr>
</tbody>
</table>

- Database says “Aspirin”, “Paracetemol”
- OWA can’t say this unless we make explicit “Paracetemol is not an anticoagulant”
Other Examples?

- XMLS vs OWL DL?
- Bio Example 1?
- Bio Example 2?
- Policy?
Closure of Open World

- Common or garden closure
  - disjoints, universals, covering & closure axioms
- Domain, Concept, Role closure
- k operator, query set subtraction etc
- where to handle non-monotonicity
  - (in query/app level, once only transform?)
Other Issues

- Over/under constraining
- SPARQL, RDQL
- Single vs Multi model?
Conclusion

- OWA and CWA both have a place in many applications
- OWA is good for describing knowledge in a way that is extensible
- CWA is good for constraining and validating data
Thankyou

thanks to Bijan Parsia, Alan Rector, Robert Stevens and BHIG at Manchester