DualRDF: Spreading Activation over wRDF

Retrieval and Selection, Hybrid Reasoning and IR model

Atanas Kiryakov

Apr, 2008

Incomplete Reasoning

- Incomplete reasoning is based on the notion of rationality
- **Rationality** = Usability / Cost
- **Usability** = Relevance in Context
- Modelling relevance and context is necessary
- In IR relevance of documents to query is judged on the basis of relevance of documents to terms
  - Typically context is disregarded or poorly modeled
- **Incomplete Reasoning needs to judge "relevance of facts"**
- Some sort of "weighting" of facts is necessary
- In Semantic Web "reasoning" framework, this means weighting of RDF triples

“Weighted” RDF (wRDF)

- wRDF is inspired by DUAL (Kokinov et Al.)
  - It allows for system which combines connectionist models with symbolic reasoning
  - Cognitive science research over the last years had proven that such models can be very efficient for modelling phenomena like “working memory”, analogy making, “constructive” memory (blending)
- Weights represent **“strength” of a statement** or a relationship
  - As a start we adopt a single real number as “universal” weight
  - The “nuances” can be modelled by means of different predicates
  - There are different ways to mode it (e.g. fuzzy- vs. probabilistic logic)
- **How weights are determined?**
  - One approach: based on co-occurrence (like in LSA)

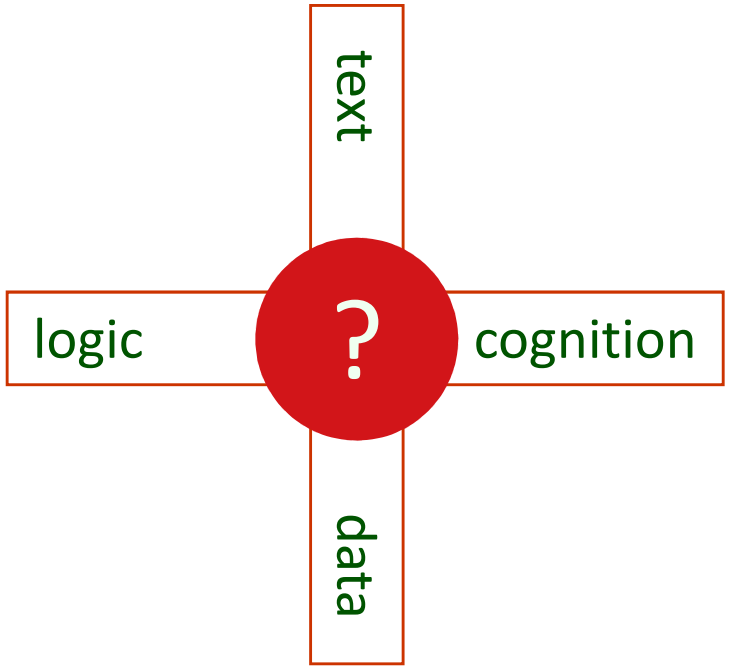
Spreading Activation & Context

- **Context = State + Preferences**
- A symbolic approaches of modeling context proven to be too complex
- Modelling context through concept activation
  - Interests: domains, topics, organizations, regions, people
  - State: location, time-of-day, device
- **RDF Priming:** spread activation to rise the “energy” of the relevant memories
- This can be used for:
  - Modelling working-memory (ignore the low-energy parts)
  - Drawing samples, relevant to the current context
  - Use energy as a “currency” for cost-based reasoning

Reading Context for Indexing

- **Reading is process**
- Hypothesis: people remember those concept which made them strongest impression, while reading
- These are the concepts which got to highest levels of activation in the process of reading
  - Not at the end of the process
  - And not like if the reader “saw” an unordered bag of words
  - Evidence: one can recall what she was thinking of while reading the text even if it was not written there
- Thus **reading context** can be used to determine “key concepts”
- Reading context also allow indexing with more abstract and characteristic features

The Meeting Point



Incomplete Reasoning = IR

- IR^2 = Incomplete Reasoning & Information Retrieval
- Weighted-RDF Allows for Hybrid Querying
- One can **de-compose a document vector** into set of w-triples
  - The document is related to the terms which are referred in it by <document, refers, term> triples, which are weighted with TF.IDF
  - This is only feasible in reduced dimensionality feature space
- Concepts are related to other concepts in the same way as to documents

Implementation Notes

- Spreading of activation (SA) against large RDF graph is a computationally heavy task
- In project RASCALLI, together with out partners from NBU, we experiment with a system named DualRDF, which allows for SA over RDF graphs
  - A “classical” RDF store is combined with
  - SA and Machine Learning component, based on sparse matrix algebra
- This approach allows for easy adoption of existing infrastructure for sparse matrix calculations, including hardware solutions (e.g. FPGA) and libraries with support for massive parallelization

Summary

- We know NN have very limited usability
  - BTW, mathematical logic (purely symbolic logical reasoning) also has very limited practical applications
- Spreading Activation over wRDF is a “cheap” way to:
  - Model context and preferences (and their dynamics)
  - “Guide” cost-based symbolic reasoning
  - Perform relevance ranking on the results
  - Contrary, NN usually do recognition
- “Reading context” can serve indexing and disambiguation
- wRDF provides ground for hybrid reasoning
- If the “cheap” way does not work, we can try quantum computing (FP8)