News from ISWC 2008

Marco Brandizi

SW Basics Recall

⇨ From the Web of textual unstructured pages...

⇨ ...to the Web of pages + structured, standardised, interconnected knowledge

⇨ Advantages:

  ⇨ Sharing

  ⇨ Collaboration

  ⇨ Discovery (by both humans and machines)

SW Basics Recall

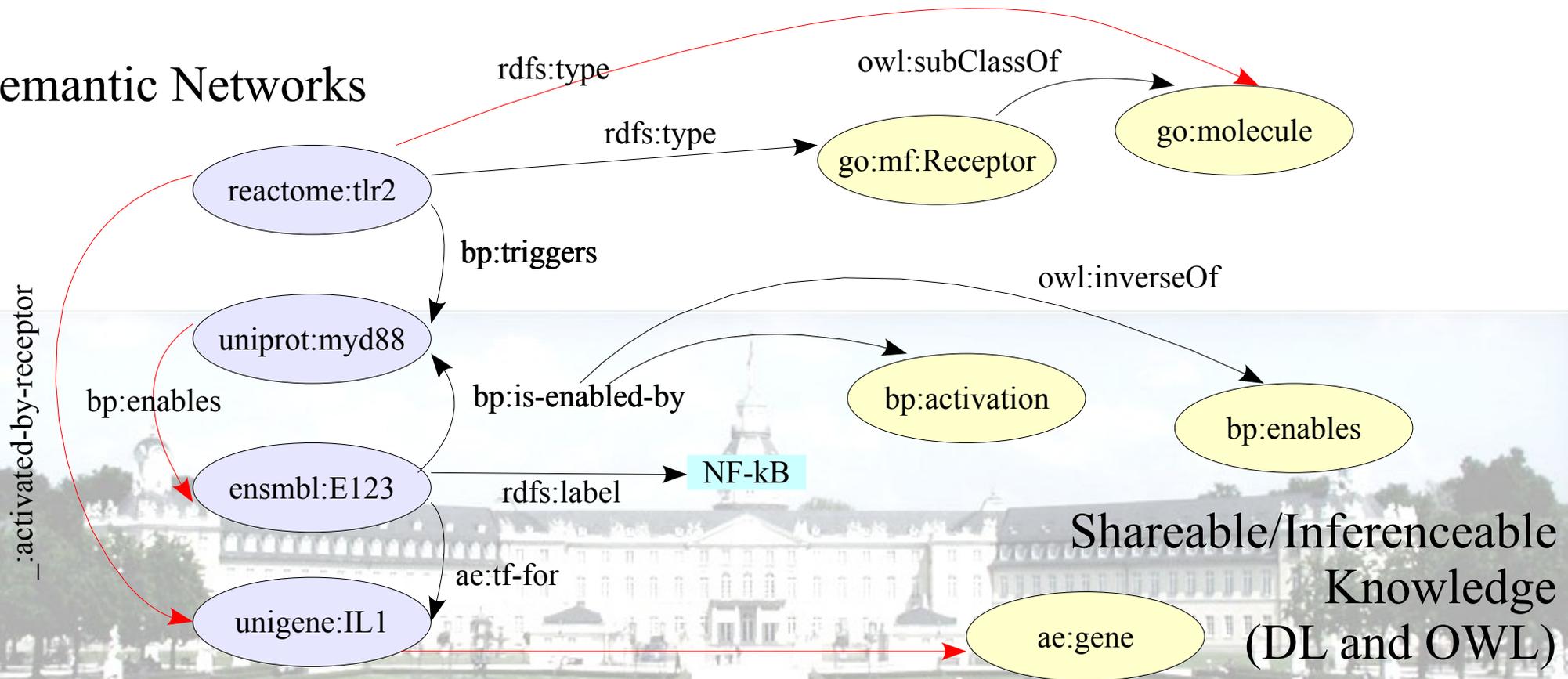## Resources and URIs

reactome:tlr2
uniprot:myd88
http://www.ebi.ac.uk/ensmbl.jsp?seq=E123
lsid:unigene.ncbi.gov:IL1

Statements (RDF)

Subject	Predicate	Object
x	R	y
reactome:tlr2	bp:triggers	uniprot:myd88
ensmbl:E123	bp:is-enabled-by	uniprot:myd88
ensmbl:E123	ae:tf-for	lsdi:unigene.ncbi.gov:IL1

Semantic Networks

rdfs:type

owl:subClassOf

rdfs:type

go:molecule

go:mf:Receptor

reactome:tlr2

bp:triggers

owl:inverseOf

uniprot:myd88

bp:is-enabled-by

bp:activation

bp:enables

bp:enables

ensmbl:E123

NF-kB

rdfs:label

_:activated-by-receptor

unigene:IL1

ae:tf-for

ae:gene

Shareable/Inferenceable
Knowledge
(DL and OWL)

SW Basics Recall

▷ And much other stuff:
  ▷ Extract RDF from existing DBs and Web sites
  ▷ Linking URIs to computer resources (i.e.: resolving)
  ▷ Quering, Publishing, Integrating RDF knowledge bases
    ▷ Provenance and Trust
  ▷ Schemas or Formal Ontologies or Folksonomies
  ▷ Reasoning, consistency
  ▷ Development frameworks and tools

▷ Possible Application Areas (present at the conference):
  ▷ Social Networking
  ▷ Semantic Desktop (i.e.: sharing personal knowledge)
  ▷ Life Sciences
  ▷ Entertainment, Manufacturing, News...

Realising Semantic Web Applications (RSWA)

⇨ **http://swa.cefriel.it/Teaching/RSWA2008**

⇨ Basic introduction to what SW can be in practice

⇨ Frames SW development into Software Engineering & Software Lifecycle

⇨ Covers:

  ⇨ Modeling with RDF and ontologies

  ⇨ Integration of different data sources in several formats

  ⇨ Information extraction from Web pages (e.g.: GRDDL)

  ⇨ Web Rendering of RDF data (Simile/Exhibit)

⇨ A realistic (not real) application is developed and available on line

SW for Health Care and Life Sciences

⇨ **http://esw.w3.org/topic/HCLS/ISWC2008/Tutorial**
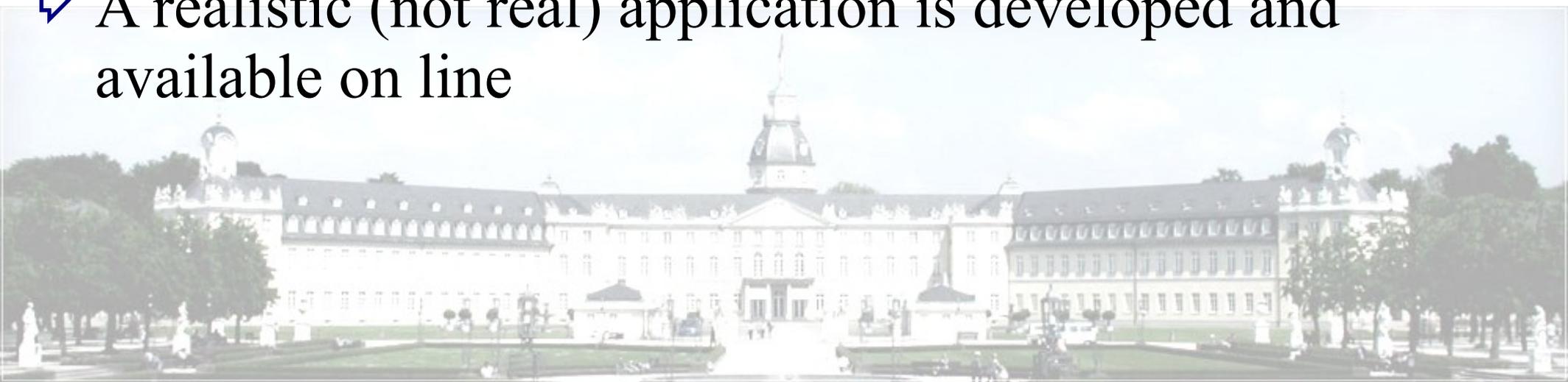
⇨ Tutorial + Workshop

⇨ Several People from the HCLS-SIG (from W3C)

  ⇨ KB for Life Sciences
    http://www.w3.org/TR/hcls-kb/

  ⇨ Converting Relational DBs to RDF
    http://www.w3.org/TR/2008/NOTE-hcls-senselab-20080604/

⇨ Neuroscientific data and other LS resources integrated to allow for answering of valuable Biological questions:

  ⇨ *"What genes are involved in signal transduction that are related to pyramidal neurons?". The scientific question can be answered with the following query, which searches for gene names and processes from four data sources within the knowledge base. The data sources include: MeSH (Pyramidal Neurons), PubMed (Journal Articles), Entrez Gene (Genes), Gene Ontology (Signal Transduction).*

  ⇨ I.e.: MESH:Pyramidal Neur. → PUBMESH:articles → ENTREZ:gene-records → ENTREZ:products → ENTREZ:annotations → GO:BP:Signal Transduction

SW for Health Care and Life Sciences

▷ Other Sources that were integrated:
  ▷ Drug Data (considers clinical trials, CHEBI)
  ▷ Clinical Observations (integrates SDTM and HL7)
  ▷ Hypotheses/discussions/etc. coming from the SWAN project
    http://swan.mindinformatics.org/

SW for Health Care and Life Sciences

⇨ How it can be useful @EBI:

⇨ EBI-wide: several resources that can be integrated

⇨ More powerful queries and better results (e.g.: Gene expression + Proteomics + Pathways)

⇨ Driven by Specific studies

⇨ Within MA:

⇨ Better representation of biological material and protocols

⇨ Samples rendered in OBI + other ontologies
⇨ Better validation, better searches

⇨ Improvements to the Atlas

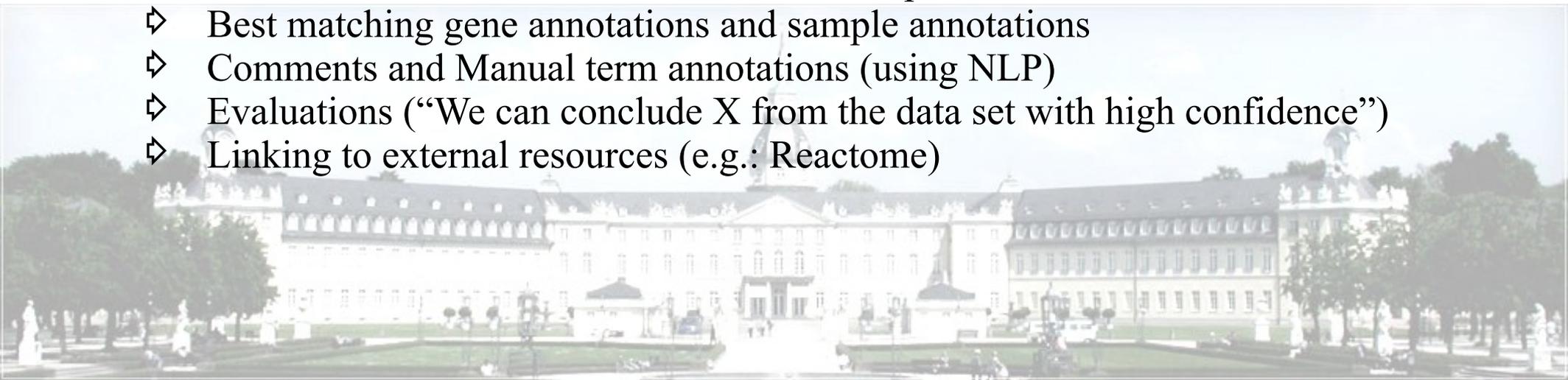⇨ Similar experimental conditions (not only the Factors) or gene sets
⇨ Allow the users to save their own data sets, help them to annotate it:

⇨ Best matching gene annotations and sample annotations
⇨ Comments and Manual term annotations (using NLP)
⇨ Evaluations ("We can conclude X from the data set with high confidence")
⇨ Linking to external resources (e.g.: Reactome)

Key Notes and other general talks

Stefan Decker: How can the Semantic Web Community be more convincing?

- ⇨ http://videolectures.net/iswc08_decker_mib/
- ⇨ One of the most active European SW VIPs
- ⇨ (My personal thought) Being convincing for the SW is important, cause many projects make sense only if they are on large scale => consensus and public funding
- ⇨ He compared Physics (LHS), AI, SW
  - ⇨ SW has a credible feasible agenda, but not so much an appealing message
- ⇨ Let's propose the SW as networked knowledge that helps in collaborative/distributed problem solving.
- ⇨ Examples: SIOC, Semantic Desktop

Stefan Decker: How can the Semantic Web Community be more convincing?

➪ A question was:
  ➪ "Networked Knowledge" is a better definition, but maybe still not enough (one of the questions). Probably we need appealing message + concrete applications
  ➪ Physicians came at the LHS result after long time, but they knew exactly what they wanted.
  ➪ They have appealing message: a bounce of questions about what we are and we come from. "Networked Knowledge" sounds better than previous defs. But still not alike exciting
➪ Mr. Decker reply:
  ➪ I agree
  ➪ We have some positive experience, but still not the ability to do things systematically
  ➪ Probably IT people are not the best one for that
➪ My personal thoughts
  ➪ Probably we need appealing message + more concrete results
  ➪ Life Sciences requires more formalism than SIOC
  ➪ Semantic Desktop doesn't seem to be much relevant to LS at the moment, but:
    ➪ More connectivity and user trasparent information collection?
    ➪ Spreading of widgets and other alike tools?

Data Intelligence
Evelyne Viegas (Microsoft Research)

⇨ http://videolectures.net/iswc08_viegas_di/
⇨ Data confidentiality and Research:
  ⇨ People have right to privacy, researches need data in order to do research
  ⇨ Current tools are not flexible enough:
    ⇨ On site access, Non disclosure licences
    ⇨ Examples about anonymity and cross-data integration: we must be concerned of "Little Brother" too, not only Big Brother
⇨ Possible solution: a framework where privacy violation risks are evaluated by analysing the semantics of queries
  ⇨ "How many searches about Britney Spears" OK
  ⇨ "What Britney Spears searched" KO
  ⇨ A single result with an address is at risk
  ⇨ See also: Differential Privacy [in the presentation]
⇨ My notes
  ⇨ Hard to explain to key decisors / lawmakers
  ⇨ Do "100% non reserved results" become an asymptote in this things too?
  ⇨ Anyway they are interesting ideas for Life Sciences

OWL Panel

⇨ A discussion about OWL, OWL2, strengths and weaknesses of OWL

⇨ Main points

⇨ OWL has become a KR of choice, despite it was not designed for that (T. Finin)

⇨ But probably won't become a widespread standard, because of its unsuitablility in several applications (Dumontier):

⇨ Space-time; Uncertain knowledge and probabilistic reasoning, Trust/Provenance and disagreement

⇨ Performance is still a problem

⇨ Nonetheless it is useful in many applications, used in KBs with thousands or millions triples (I. Horrocks)

⇨ Industry people vs Researchers dichotomy is a problem:

⇨ Busy Father def= a Father and hasChildren >= 2

⇨ Missing children for someone working on real world

⇨ Skolem constant 1 and Sk2 for a researcher...

OWL Panel

⇨ I. Horrocks mentioned OWL2, other references about:

⇨ OWL Datatypes: Design and Implementation (talk by B. Motik)

⇨ OWL 2: The next step for OWL, http://dx.doi.org/10.1016/j.websem.2008.05.001

⇨ Main OWL 2 features:

⇨ **Expressivity**

⇨ Data Types Improved

⇨ OWL Datatypes: Design and Implementation (talk by B. Motik)

⇨ Backing DL improved ($\mathcal{SHOIR}$)

⇨ **Formalization**

⇨ MOF as per OMG/MDA specifications

⇨ Better serialization

⇨ Declarations avoids ambiguities (e.g. implicit classes)

⇨ **Modularization**

⇨ Profiles: OWL 2 flavours are now inspired to application scenarios, rather than expressivity/computability

⇨ EL (classification), QL (constructive query answering), RL (rule-based reasoning

⇨ Imports are more flexible

⇨ Redirection: load location can be != URI

⇨ Versioning

⇨ OWL specie declaration (needed to be interpreted and didn't work)

Other Key Notes

⇨ Freebase: An Open, Writable Database of the World's Information (J. Giannandrea)

  ⇨ A large public, user-contributed database of general knowlegde. Seems "Semanticised Wikipedia" (so DBpedia is relevant too)

  ⇨ Emphasis on simplicity: no OWL, but user-contributed schemas, knowledge domains, simple query language (by example)

  ⇨ Simplicity/large user base vs. Formalization/niche user base (Life Sciences)

⇨ Multimedia Semantic Web (R. Jain)

  ⇨ Multimedia objects annotation and events reconstruction

  ⇨ Indexing multimedia Dbs by means of event-related annotations

  ⇨ Can improve searches (e.g. in photo collections)

  ⇨ The approach can be interesting in bio-repos where time is relevant (e.g. pathways, time series in microarrays) ?

Applications

Peter Mika from Yahoo

⇨ Search Monkey: an "eco system" where it is possible:

  ⇨ to extract/publish RDF data (dataRSS = Atom extended with RDFa)

  ⇨ Search information and build search services

    ⇨ Render search results (exploiting returned RDF data)

  ⇨ Interesting ideas for data import/export

  ⇨ Related stuff: Yahoo Pipes, Dapper (www.dapper.net)

⇨ A similar project (Finding Answers on the Semantic Web, http://www.talis.com/nodalities/pdf/nodalities_issue4.pdf):

  ⇨ Watson Semantic Gateway, to define RDF export

  ⇨ PowerAqua: to allow Natural Language queries

Other Applications

⇨ An architecture for semantic navigation and reasoning with patient data - experiences of the Health-e-Child project (T. Hauer)

  ⇨ An example of relational data converted to OWL, integrated with ontologies, used to serve queries (via inference)

⇨ Nepomuk: a set of tools for the Semantic Desktop (poster, try it at http://nepomuk.semanticdesktop.org/)

⇨ Creating and Using Organisational Semantic Webs in Large Networked Organisations

  ⇨ Simple forms-based approach to capture metadata and frame it into ontologies.

Development Tools

Dev Tools

⇨ RDF123: a tool for converting spreadsheet to RDF
   ⇨ Simple
      ⇨ row = object, column = property
      ⇨ Template graph and conversion expression in the spreadsheet
   ⇨ Yet, interesting idea
⇨ Supporting Collaborative Ontology Development in Protégé (Tudorache T et al)
   ⇨ Provides: Changes, Annotations, Versioning, Controversies management
   ⇨ Supported by a Changes & Annotation ontology
      ⇨ Main elements: ontology components (classes, properties, etc.), changes, annotations, users and roles
   ⇨ Has a collaborative environment that keeps Protegé users connected
   ⇨ Has a Protegé Web client (limited features for the moment)

SW Engineering, Theory, Research

SW Engineering, Theory, Research

⇨ Integrating Object-Oriented and Ontological Representations: A Case Study in Java and OWL (Puleston C, et al)
  ⇨ A clear description of how to mix Object-oriended models with RDF-oriented models, pro/cons of both models, advantages of mixing them
⇨ An Interface-Based Ontology Modularization Framework for Knowledge Encapsulation (Ensan F, Du W)
  ⇨ A concept can be refined by an external module (interface)
  ⇨ Only those DL definitions in the module that are needed to "implement" interfaces will be imported by the client module
  ⇨ A sample implementation for the SWOOP editor is provided
  ⇨ Approach is interesting possible problems to be seen (community support, performance)
⇨ Using Semantic Distances for Reasoning with Inconsistent Ontologies (Huang Z, Van Harmelen F)
  ⇨ Essentially: When P and ~P, ontologies is "reduced" by eliminating one of the two
  ⇨ Which one is computed by choosing the one semantically closer to the query (works for subsumption queries)
  ⇨ For what concerns the Bio domain, the general idea of reducing is possibly interesting

Finally...

My 2¢...

⇨ Future Web won't be simply W3C's SW
  ⇨ It is not only about machines, e.g.: social networking
  ⇨ It won't bee much "formal" (micro-formats, schemas and lightweight ontologies)
  ⇨ The SW layer cake by W3C will probably become a reference architecture, like ISO/OSI or CORBA, hardly we will see applications with all the layers implemented
⇨ Life Sciences domain is a niche with its peculiarities
  ⇨ Formal ontologies are much more needed
  ⇨ Sophisticated reasoning is needed
    ⇨ And efficient/scalable reasoners too
  ⇨ Some aspects are much more relevant than other domains (e.g. fine control of privacy/restricted access)
⇨ There are many unexplored opportunities in Bio-SW domain

http://semtext.org/2004-02/slides/img4.html