Position Paper:

Ontology Construction from Online Ontologies

Harith Alani

Ontologies and the Semantic Web

- Ontologies have become the backbone of the Semantic Web
  - They model knowledge to enable machines to share and understand it
  - More and better ontologies are therefore necessary for a wider Semantics Web spread

- The bad news is:
  - Constructing ontologies is not a walk in the park!

Ontology Construction

- Several methodologies have been proposed
  - All emphasise the role of *reuse* to avoid starting from scratch to bring costs down
  - However, there are no tools to facilitate that!

- Several approaches have been researched to extract ontologies automatically from:
  - Databases, text corpora, software systems, etc.
  - Results show a persistent need for background knowledge, not usually explicitly expressed in such knowledge sources

- But how about reusing existing ontologies to construct or assemble new ones?
  - If there are ontologies relevant to you domain of interest ..
  - Background knowledge should no longer be a problem
  - Not starting from scratch
  - Bootstrap the process of ontology building

Ontology Reuse

- Ontology editing tools
  - E.g. Protégé, Swoop, KAON framework
  - Mainly for editing ontologies, but also not much support for reuse

- More ontologies are coming online
  - Several ontology libraries are currently available (eg DAML library, Protégé, Ontolingua)
  - Ontology search engines are now appearing, eg Swoogle

- Such tools and libraries only provide basic search and retrieval services
  - The focus is mainly on search and manual selection
  - They are not designed to support ontology reuse in terms of ontology reconstruction, merging, evaluation, etc.

How can we make use of all those online ontologies to bootstrap ontology construction?

Scenario

- *"Imagine there is a knowledge engineer who is in need of an ontology representing the academic domain. The ontology is to be used for creating a knowledge-base to hold information on staff, projects, conferences, publications, etc."*

- There are many ontologies online that covers various portions of this domain, in a variant level of detail!

- It would be useful if our engineer can quickly and efficiently reuse some of these existing ontologies, to at least bootstrap the ontology construction process

Rank the Ontologies

- Let's assume that the engineer needs to represent the concept "Conference" in the ontology

- Swoogle 2006 offers 115 ontologies with a class that has a label that equals or contains the word 'Conference'

- Now we need to rank them
  - We can't look up every one of these ontologies!
  - Better to have a ranking system that can order the 115 ontologies according to some criteria
  - We can then start analysing, say, the top 5 ontologies
  - We can of course analyse more, or less, ontologies depending of the outcome of our analyses
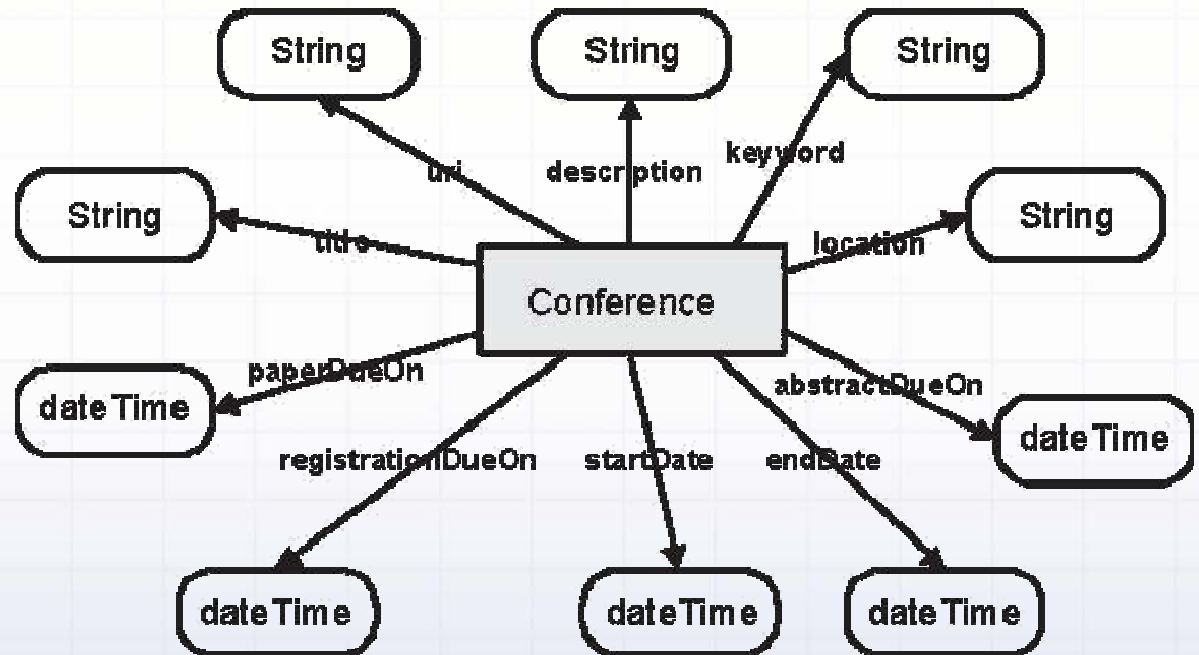
Segment the Ontologies

- Depending on the size and scope of the ranked ontologies, the system can:
  - Take an ontology as a whole
  - Or only take the section that describes "Conference"

- Segmentation enables the system to cut out only the parts of interest from an ontology

conference.owl

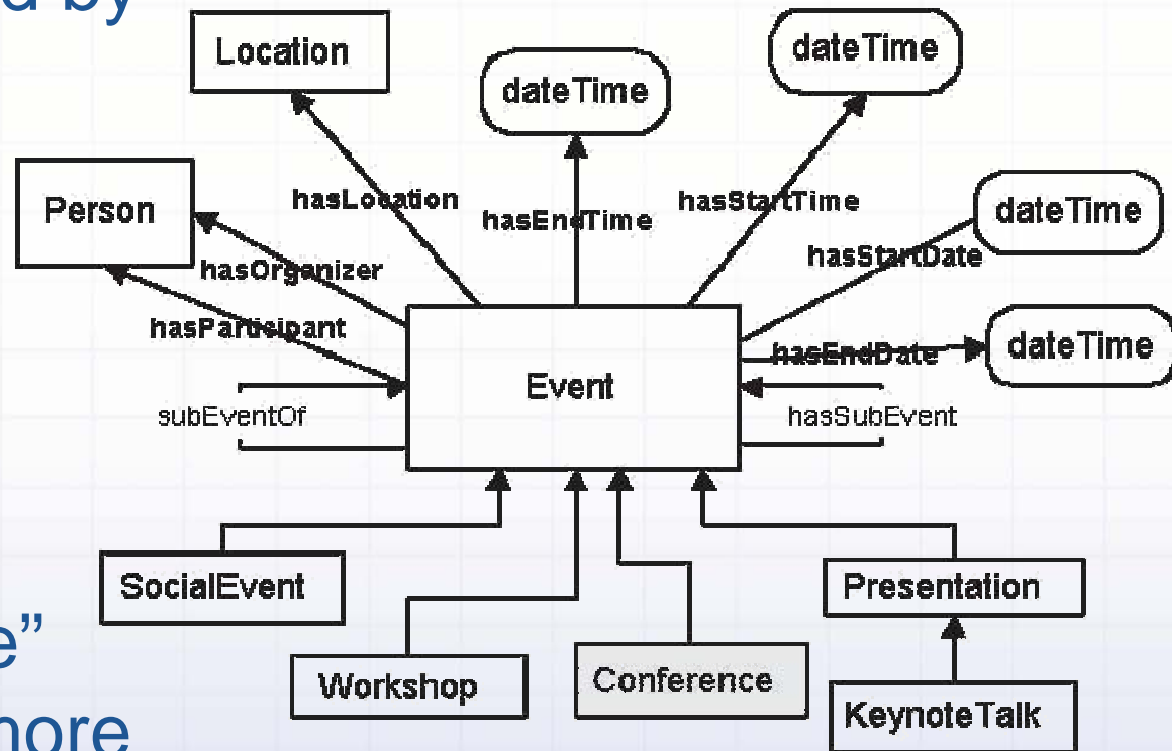- 1st hit in Swoogle 2005, 7th in Swoogle 2006

- Comprises of:
  - 1 Class
  - 10 Attributes

We Need More!

- The conference.owl ontology is not enough for what we need!

- System can reuse additional ontologies to enrich this ontology with more detail

- This is the 2nd ontology returned by Swoogle (05&06)

- The "Conference" class here has more detail than in previous ontology

Comparison and Merging

- System now needs to:
  - Compare the two ontologies (or ontology segments)
  - Find and merge additional representations into the first ontology
  - Iterate this cycle with more top-ranked ontologies
  - Present the result to the user to verify, modify and change as required

Proposed Architecture
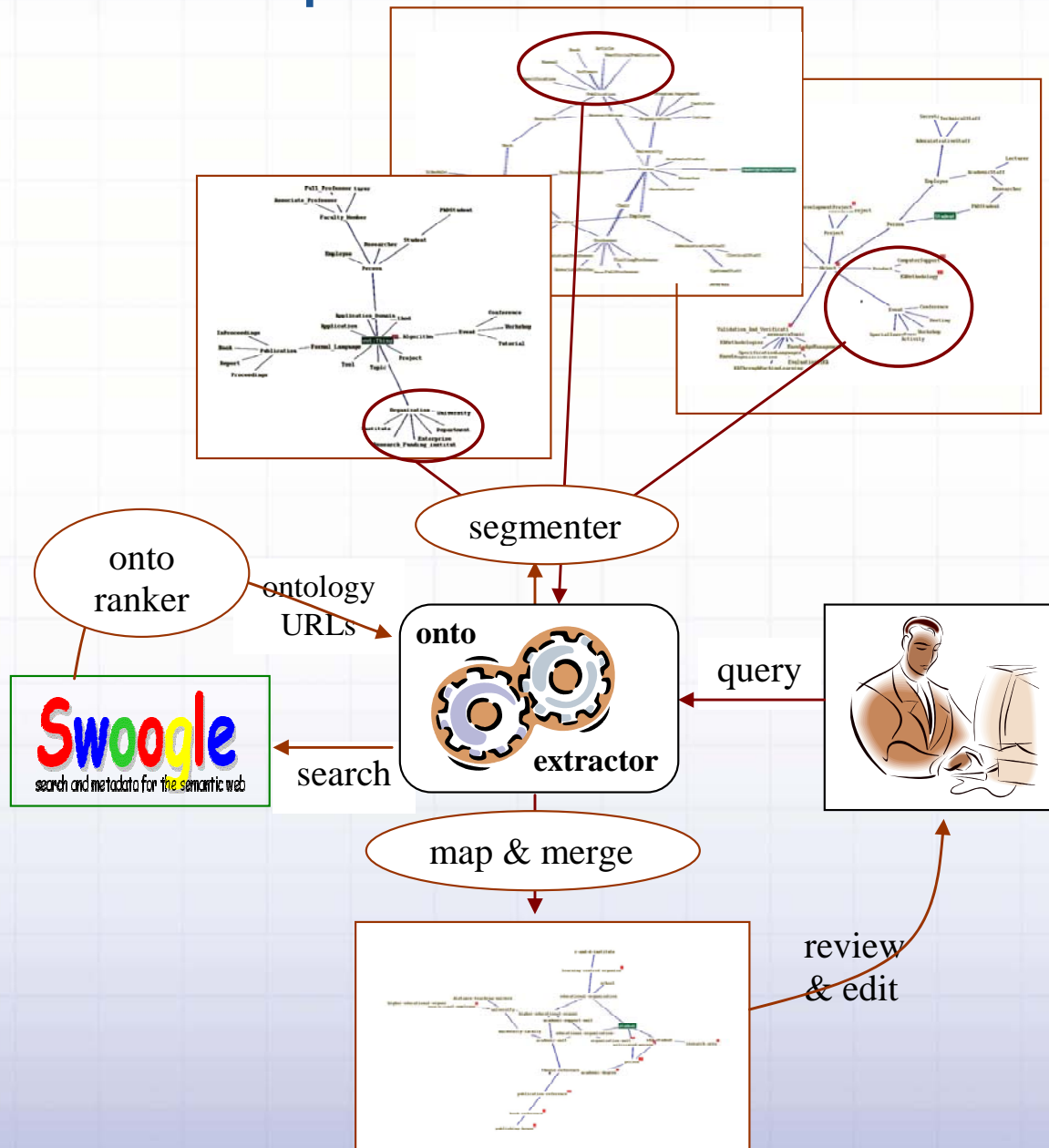
System Processes

- Search for relevant ontologies

- Rank the returned list of ontologies

- Segment ontologies if required

- Map and merge acquired segments

- Evaluate the results

- Present to the user and repeat cycle as required

Search for Ontologies

- First step is to find a list of relevant ontologies to analyse

- Searching for:
  - Specific keywords (e.g. Swoogle)
  - Metadata search (e.g. Maedche et al 03)
  - Structure-based queries
  - Query expansion

Ontology Ranking

- Rank the list of identified ontologies

- Ontology ranking techniques
  - Structural characteristics (e.g. Alani & Brewster 05)
  - User ratings (e.g. Supekar 05)
  - Content coverage (e.g. Jones & Alani 06)

Ontology Segmentation

- May need to extract parts of the ontology, depending on size and desired cope is too big

- Users can control how generous the segmentation should be

- Several segmentation approaches have been investigated based on:
  - Simple graph length (e.g. Noy et al 2003)
  - Structure (e.g. Bhatt et al 2004, Seidenberg & Rector 2006)
  - Clustering algorithms (e.g. Stuckenschmidt & Klein 2004)
  - Specific views (e.g. Magkanaraki et al 2003, Volz et al 2003)
  - Application queries (e.g. Alani et al 2006)

Onto Mapping & Merging

- System needs to compare and merge ontology segments

- A lot of work has been done in this area
  - Prompt suite (Noy & Musen 2003)
  - Chimeara (MsGuinness et al 2000)
  - Ontolingua (Farquhar et al 1996)
  - Crosi (Kalfoglou & Hu 2005)

Ontology Evaluation

- Some quality checks to the assembled ontology may help to
  - Resolve inconsistencies
  - Identify semantic gaps

- Detailed evaluation is best left to the user, but some could be automated:
  - Using reasoners (e.g.Racer, Pellet, Fact++)
  - Automated OntoClean (e.g. Volker et al 2005)
  - EON workshop on Monday!

User Feedback

- User then assesses the ontology the system produces

- User can ask system to
  - Search for additional concepts
  - Repeat process with different thresholds
    - Change the ranking technique
    - Analyse more ontologies
    - Use larger segments
    - etc

Challenges

- A challenging system no doubt!

- The required technologies are rather new and far from perfect

- Integrating those technologies into a single production line will be a good testbed

- There are additional challenges that the system will need to deal with, apart from those specific to each process ..

Additional Challenges

- Availability of relevant ontologies
    - Can't reuse what doesn't exit yet!
    - Need for good number and variety of ontologies to make reuse worthwhile!
    - Many ontologies never leave their labs
    - But more ontologies will become available, given time and encouragement to share!

- Danger of producing a Frankensteined ontology
    - The produced ontology might be too large and messy!
    - Can happen if many large ontologies are used
    - Users might struggle to clean or modify the resulting ontology
    - System cut-off thresholds can help avoiding this fate
        - More interaction with users, Gradual augmentation, Constant size checking
        - User can pause, stop, or rewind system to fiddle with settings as required

- Quality control
    - May need to restrict reuse to only *quality* ontologies or trusted ones
    - Good ranking and evaluation processes may help reduce this problem

Conclusions

- More ontologies are coming online

- Many people sweated over those ontologies!

- Time to start planning for proper reuse!

- Several semantic web technologies have been researched and studied, usually in isolation!

- Bringing them together can give a great push to reuse

- Users will remain the main drivers
  - Reuse is meant to simply bootstrap ontology development
  - Users are expected to modify, delete, add, etc