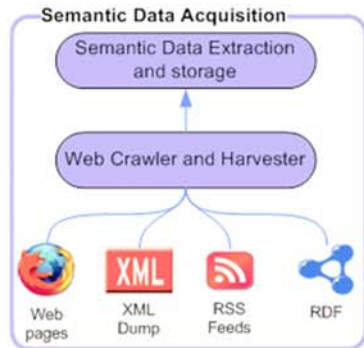


## Formalisation of Semantic Search Framework

### Introduction

- Conventional information retrieval techniques are document-centric search paradigm. Semantic search is an object, entity or knowledge-centric approach.
- Semantic web is not only a web of interlinked documents, but a web of relations through which users are able to explore and search knowledge in large information space.

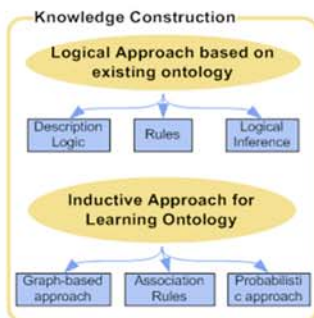
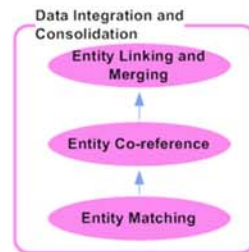


### Semantic Data Acquisition

- The current web is a tremendous repository of documents linked by hyperlinks.
- Large amount of data is stored in an unstructured format, structured data is mostly provided by researchers in the semantic web community.
- To bootstrap the semantic web, the unstructured data needs to be harvested, extracted, cleaned and converted into a machine-processible format.
- Data sources: web pages, XML dump (e.g. DBLP dataset), RSS feeds, and RDF data.

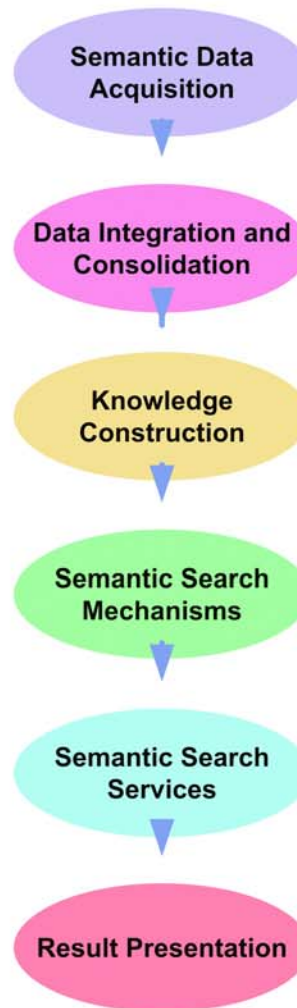
### Data Integration and Consolidation

- Semantic web proposes reuse and interoperability. Republishing data is not encouraged, instead, by providing links to the URIs of the published data.
- Different sources may publish information about the same entity. Entity matching is the first step to integrate and consolidate data.
- By referencing to URIs of existing resources, the need for republishing information is avoided.
- Information objects are linked, merged to form a web of interlinked objects.



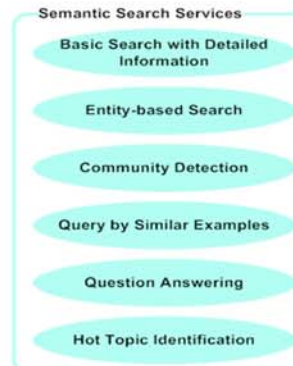
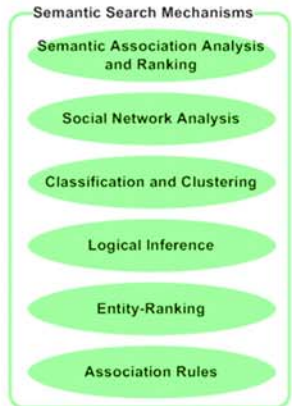
### Knowledge Construction

- On the semantic web, knowledge is encoded in ontologies using knowledge representation languages such as RDF/S or OWL.
- In logical based approaches knowledge is deduced through logical assertions and rules.
- The logical approach is limited in scope, coverage and handling uncertain knowledge.
- Knowledge can also be obtained using inductive approaches, in this way, we can have learning ontologies. The links between entities are weighted based on the result obtained using association rules, probabilistic approaches for instance.



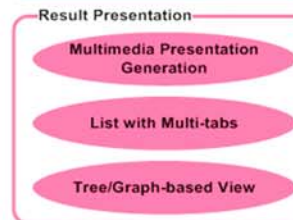
### Semantic Search Mechanisms

- The basic search approach in the semantic web is using logical inference.
- Different entities have different characteristics and need different processing mechanisms before injected into search engines.
- Semantic association utilises graph-based analysis approach to speculate relations between entities. Limitation of most existing work is that it is done in the closed world.
- Social network analysis primarily concentrates on relations between people. However, in existing work analysis in finer granularity is not investigated.
- The future research will investigate semantic search using ontologies populated with uncertain knowledge under the open world assumption on the semantic web.



### Semantic Search Services

- Based on the semantic search framework and the knowledge bases build, interesting search serviced can be provided.
- Entity-based search allows users to explore and search in the large information space using links between them.
- Community detection is also enhanced due to the fact that different parties can reuse and share machine-processible data.
- The surge of interest in Web2.0 also enables search applications to provide more related results by exploiting collaborative knowledge.



### Visualisation and Presentation

- Semantic search results can be presented in several possible ways.
- For example, multimedia presentation generation (e.g. MANA system), traditions list presentation with multi-tabs, faceted browsing and tree or graph-based view.