Intelligent RSS News Aggregation Based on Semantic Contexts

Weihong Huang

Centre for Internet Computing
The University of Hull, Scarborough Campus
Scarborough, YO11 3AZ, United Kingdom
W.Huang@hull.ac.uk

David Webster

Centre for Internet Computing
The University of Hull, Scarborough Campus
Scarborough, YO11 3AZ, United Kingdom
icr03dew@cic.hull.ac.uk

1

- Context is an increasingly common notion in information retrieval.
- Intelligent applications (including agents)which are context-aware will provide additional information retrieval functions in addition to simple searching.
  - Examples include bridging information between heterogeneous information sources.

2
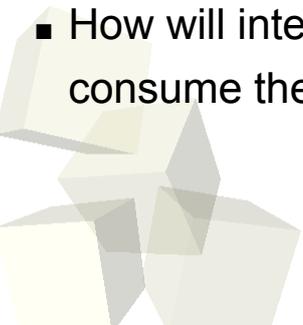
- In the future a number of heterogeneous contextual information formats will emerge.
- Examples of these include Semantic Web based languages such as those based upon XML and RDF.
  - Such languages include RSS and OPML.
- How will intelligent applications be able to consume these heterogeneous resources?
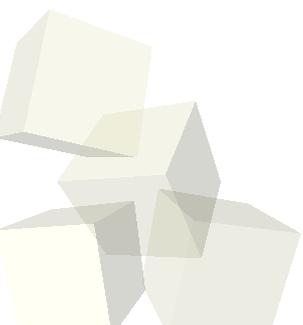
- *How can we standardize this heterogeneity?*
- *Should be flexible enough to be able to cope with any kind of document model.*
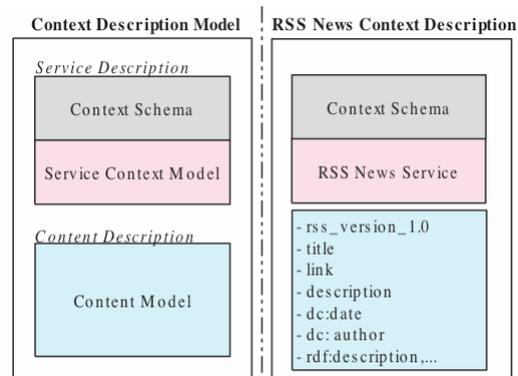- *A method of providing context of information sources to and from intelligent applications.*

- A potential solution to the heterogeneity problem
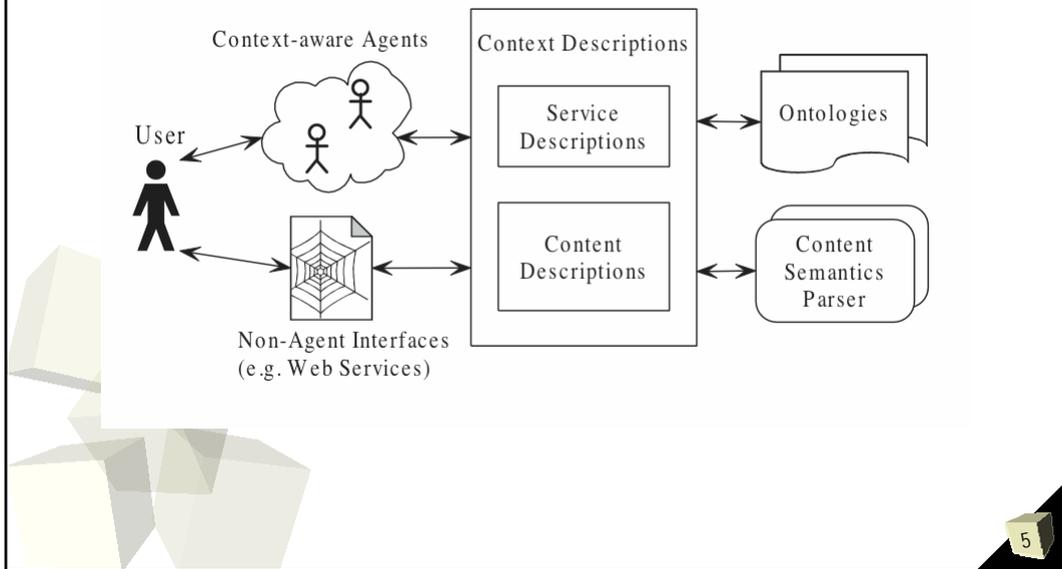
- RSS is used for syndicating simple news headlines on the current web.
- Three diverse formats of RSS

  RSS 0.9 – XML based

  RSS 1.0 – RDF based

  RSS 2.0 – XML based

- What if we want our intelligent news aggregation application to source from any of these(or more) sources.

- An example application of the Context Description Model.
- The header describes the type of service information for the agent to expect.

Context Description Model	RSS News Context Description
Service Description	
Context Schema	Context Schema
Service Context Model	RSS News Service
Content Description	- rss_version_1.0 - title - link - description - dc:date - dc: author - rdf:description,...
Content Model	

---

aggregation

- Utilizing context-aware agent services to locate relevant news items from RSS sources.
- Uses JADE agent framework.
- Procedure:
  - The context header is parsed followed by the content.
  - The agent now knows to expect an RSS payload.
  - The format field of the content is checked and appropriate parser chosen: JENA for 1.0 and Informa for 0.9 and 2.0
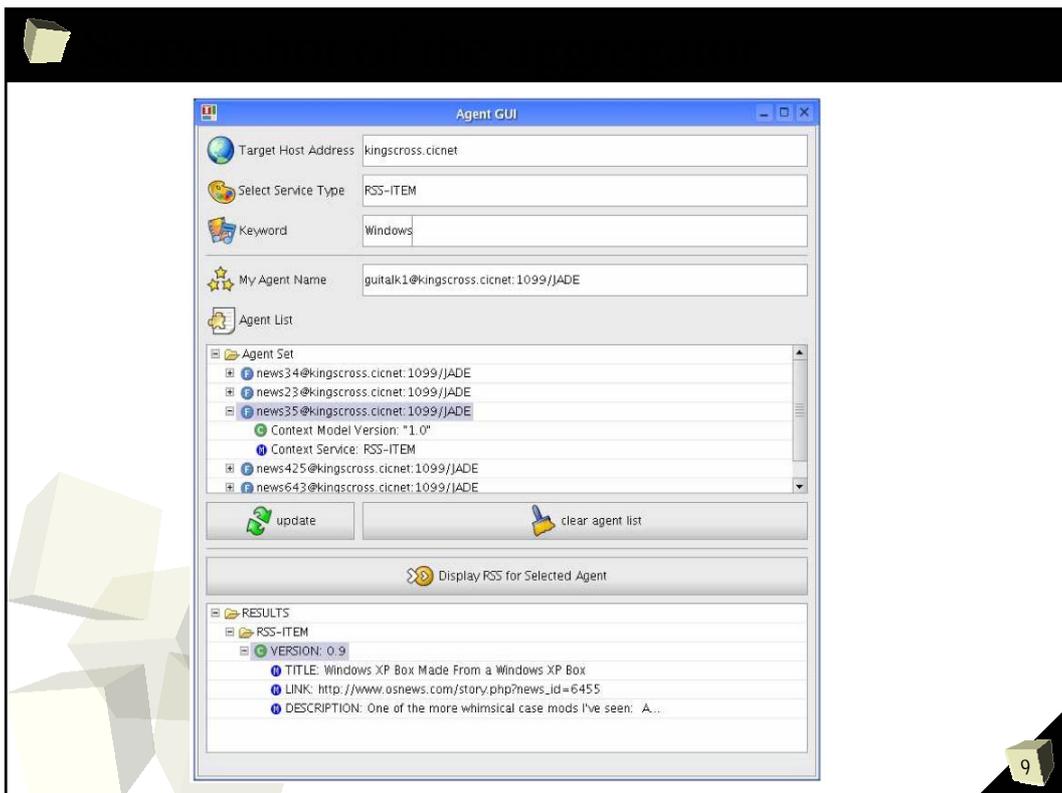
- Agent can dynamically retrieve and aggregate news related to the user's interests.

- Taxonomy/Ontology-based searches.

- RSS 1.0 supports taxonomy namespace, but is in relatively small use.

  Can we extract additional taxonomic information from the source website of the RSS feed?

■ OSNews topic extraction
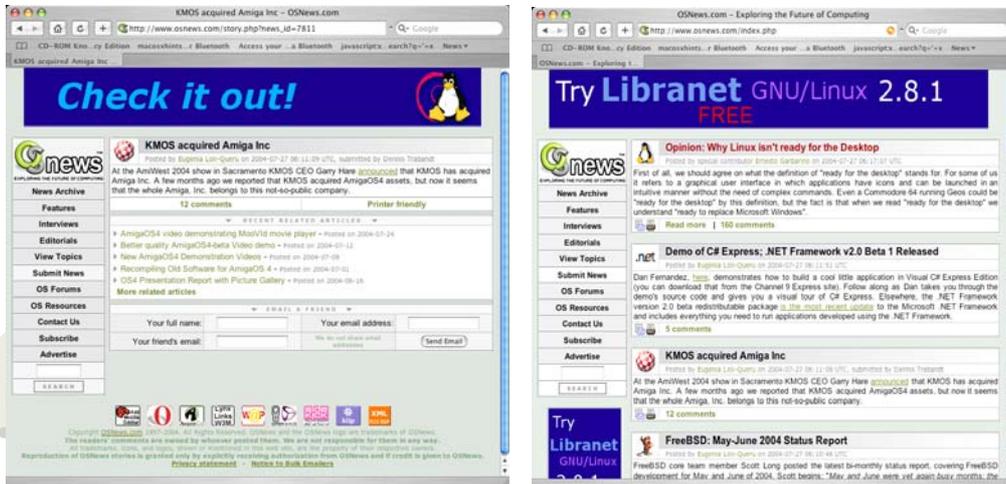
■ Can then re-publish the news item with
appropriate content version.

11

12
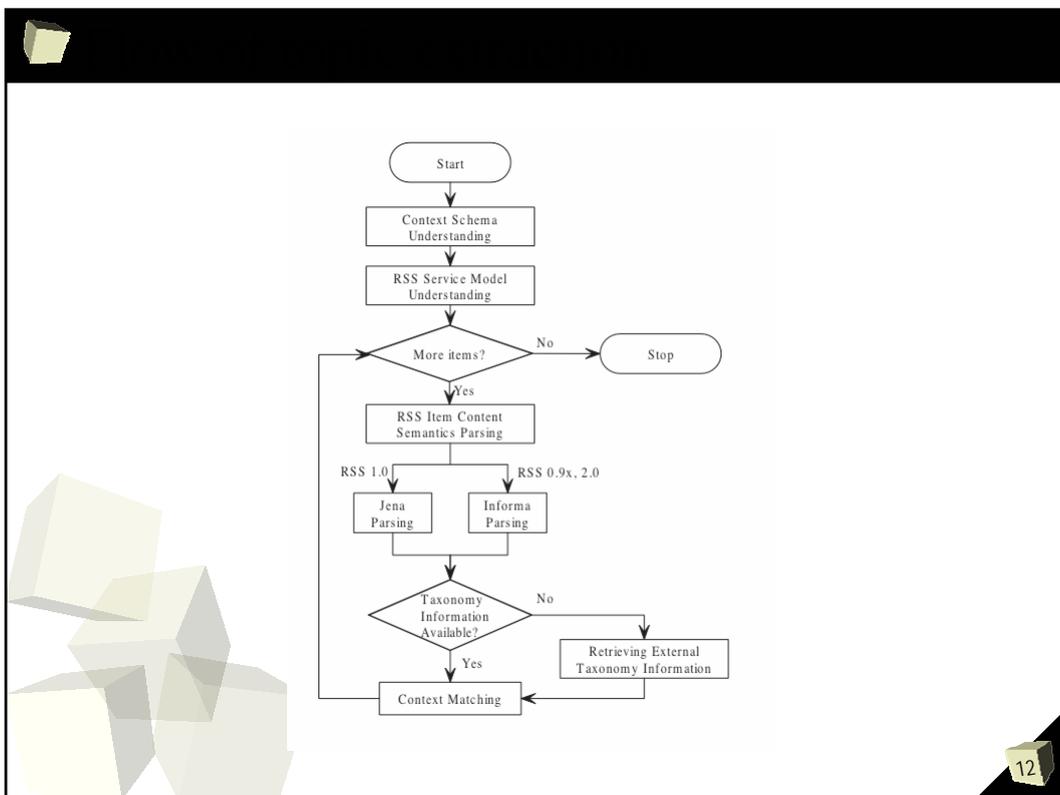
- Context-based service architecture for sharing of heterogeneous information sources between intelligent applications.
- Agents can be used to automatically retrieve semantically matched information through context aware services and reduce the workload of manual filtering.

13

- *Thank you*

- *Demonstration*
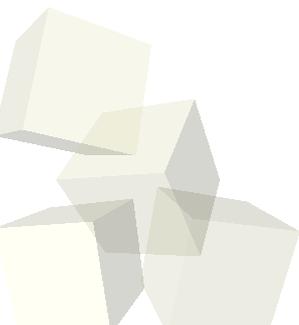
14

- Intelligent aggregation For example searching for news sources not already subscribed to.
- Agents provide the ability to propagate search requests in a distributed peer to peer manner.
- Agents provide the 'intelligence' to perform meaningful operations on this information.

15