A Controlled Natural Language Interface for Semantic Media Wiki Using the Rabbit Language

Jie Bao, Paul Smart, Dave Braines, Nigel Shadbolt

International Technology Alliance

Web 2.0

- Advent of Web 2.0 supports greater user participation in the creation of Web content
- Good way to generate lots of online content
  - e.g. Wikipedia
  - ~3 million (English) articles
- Can we enable better exploitation of user-generated content?
  - retrieval, filtering, reasoning

The Semantic Web

- The conventional web is intended for human consumption
  - content consists largely of natural language text, images, video, etc.
- Semantic Web seeks to make data more amenable to automated forms of information processing
  - standard data model + explicit semantics
- Resource Description Framework (RDF)
  - core data model + some semantics
- Web Ontology Language (OWL)
  - more advanced semantics
  - OWL typically used to create ontologies that describe the conceptual structure of a specific domain of interest

Semantic Web – Benefits

- Combination of standard data models and explicit semantics supports:
  - information exchange and interoperability
  - data integration
  - improved search and retrieval
  - reasoning and inference

Analysts

Intelligent
Agents/Assistants

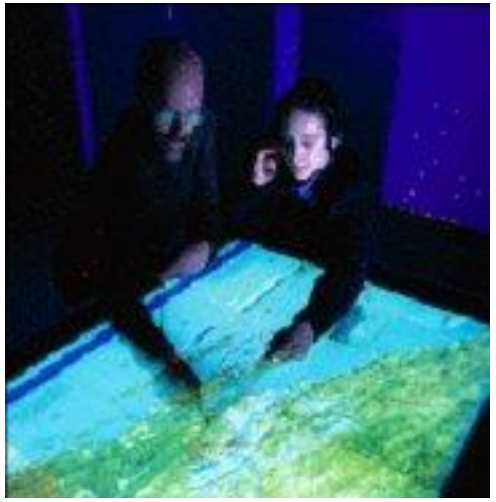
Patrols/Field Reports

Unmanned Vehicles

Shared
Repository

Remote
Sensors

Multinational Planning
Teams



Military Platforms

Non-Military
Organizations

Semantic Web Problems

- But…
  - limited amounts of high-quality, semantically-enriched data available
  - grounding in formal logic presents a usability barrier to many individuals and organizations
  - establishing consensus during ontology development is often difficult – extensive collaboration is required
- So…
  - can we learn from Web 2.0
    - greater user participation
    - delivers lots of content
    - easy to use
    - emphasis on collaborative or, at least, collective efforts

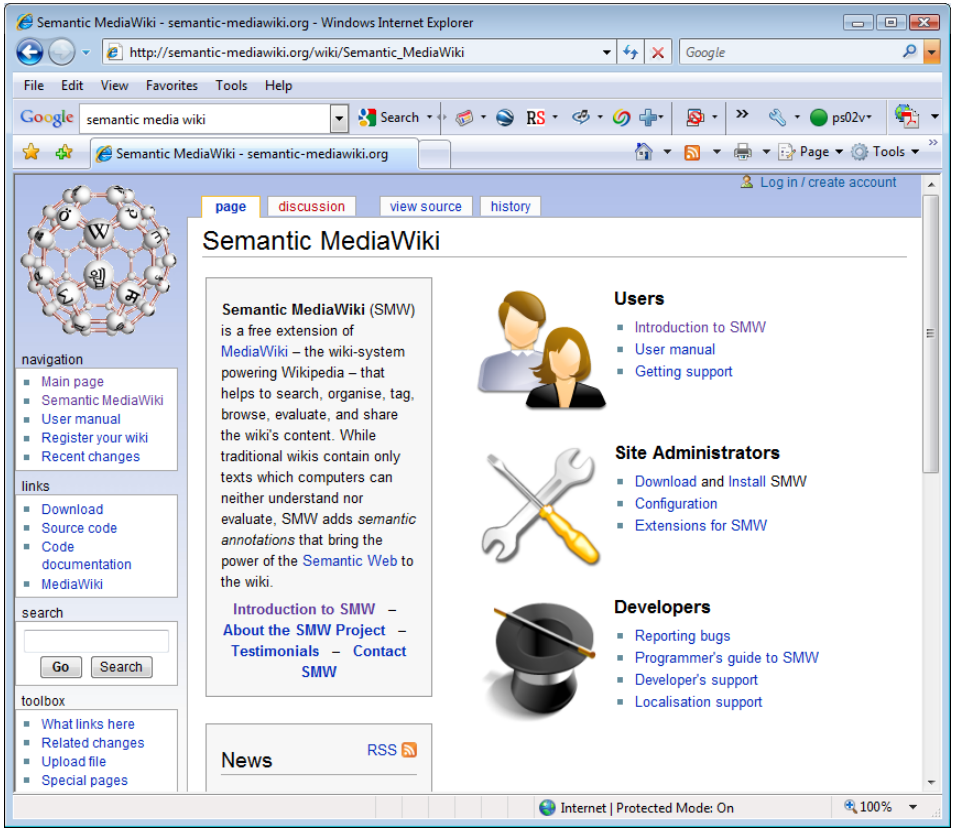
Wiki Systems

- Support multi-user content creation and editing via a Web browser interface
- Encourages large-scale participation
- Easy to use
- Content usually of reasonable quality

- Problems:
  - natural language text
  - difficult for machines to participate in content generation

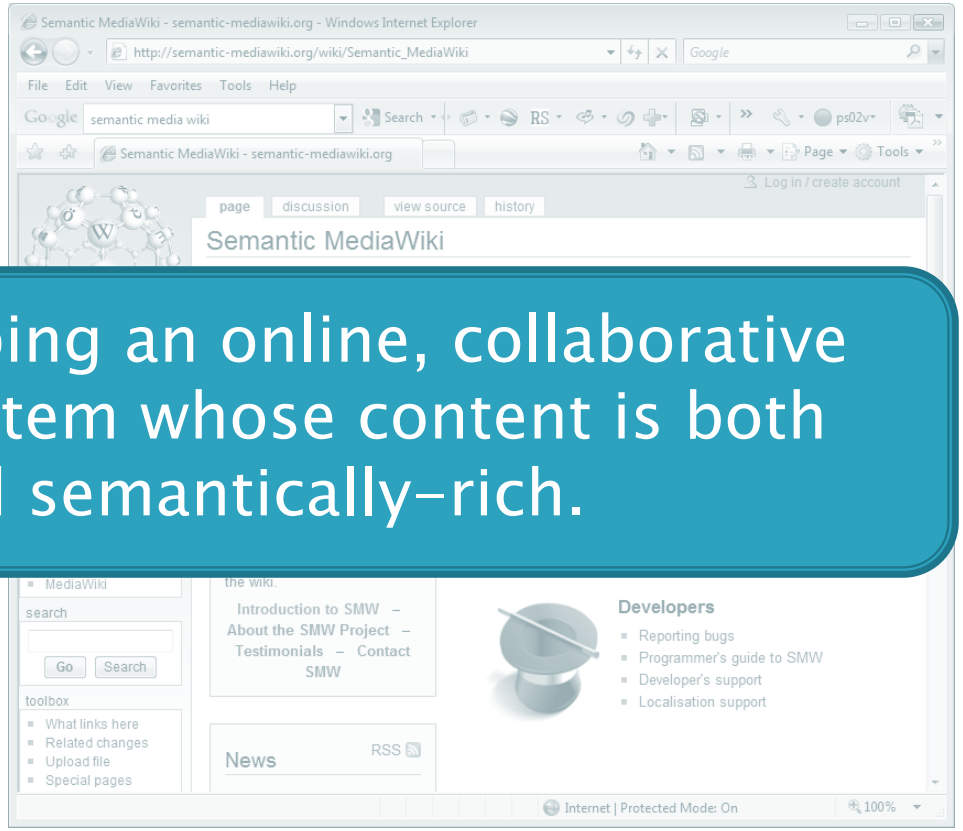
Semantic Media Wiki (SMW)

- Uses Wikipedia engine
- Perhaps the most popular semantic wiki system
- Supports the creation of semantically-enriched content
  - uses semantic annotations
- Combines features of conventional wiki system with semantic technologies



Semantic Media Wiki (SMW)

- Uses Wikipedia engine
- Perhaps the most popular semantic wiki system
- Supports the creation
- Combines features of conventional wiki system with semantic technologies



Good basis for developing an online, collaborative knowledge editing system whose content is both structured and semantically-rich.

Some Problems

- Usability
  - semantic content (esp. ontologies) difficult to create
  - departure from Web 2.0 emphasis on ease-of-use
  - even experienced knowledge engineers can find it difficult to create/edit ontologies
- Automatic content integration
  - sometimes content needs to be automatically imported without user intervention
  - recall the case of sensor feeds
- Expressivity constraints
  - semantic wikis (including SMW) do not always support the full range of OWL modelling formalisms and axioms
- Inference constraints
  - limited support for rule representation and inference

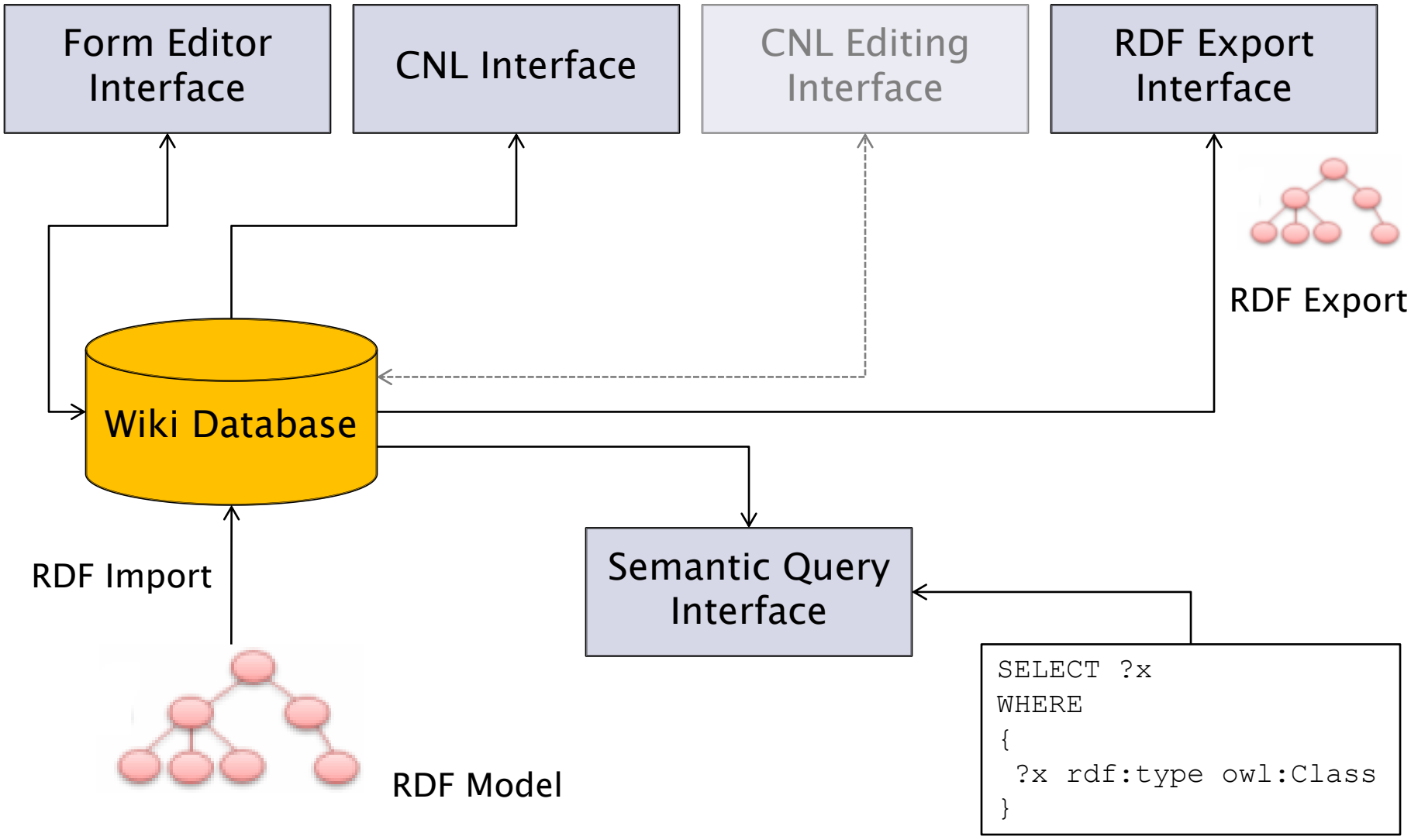
Some Solutions

- Usability
  - use CNLs
  - potential production and comprehension benefits
  - multiple OWL-compliant CNLs are available:
    - e.g. Rabbit, Sydney OWL Syntax, ACE-OWL
- Automatic content integration
  - develop an RDF import mechanism for SMW
  - support the automatic creation of wiki pages and page content from external RDF/OWL models
- Expressivity constraints
  - extend SMW with an OWL meta-model
- Inference constraints
  - implement rule representation and inference capabilities for SMW (reported elsewhere)

Research Aims

- Extend expressivity of SMW to provide full support for OWL
- Support the creation of ontologies and ontology content within SMW
- Explore ways to serialize SMW contents as (multiple) CNLs
- Investigate mechanisms to support wiki content creation using (multiple) CNLs
- Develop CNL editors to support content creation

System Architecture

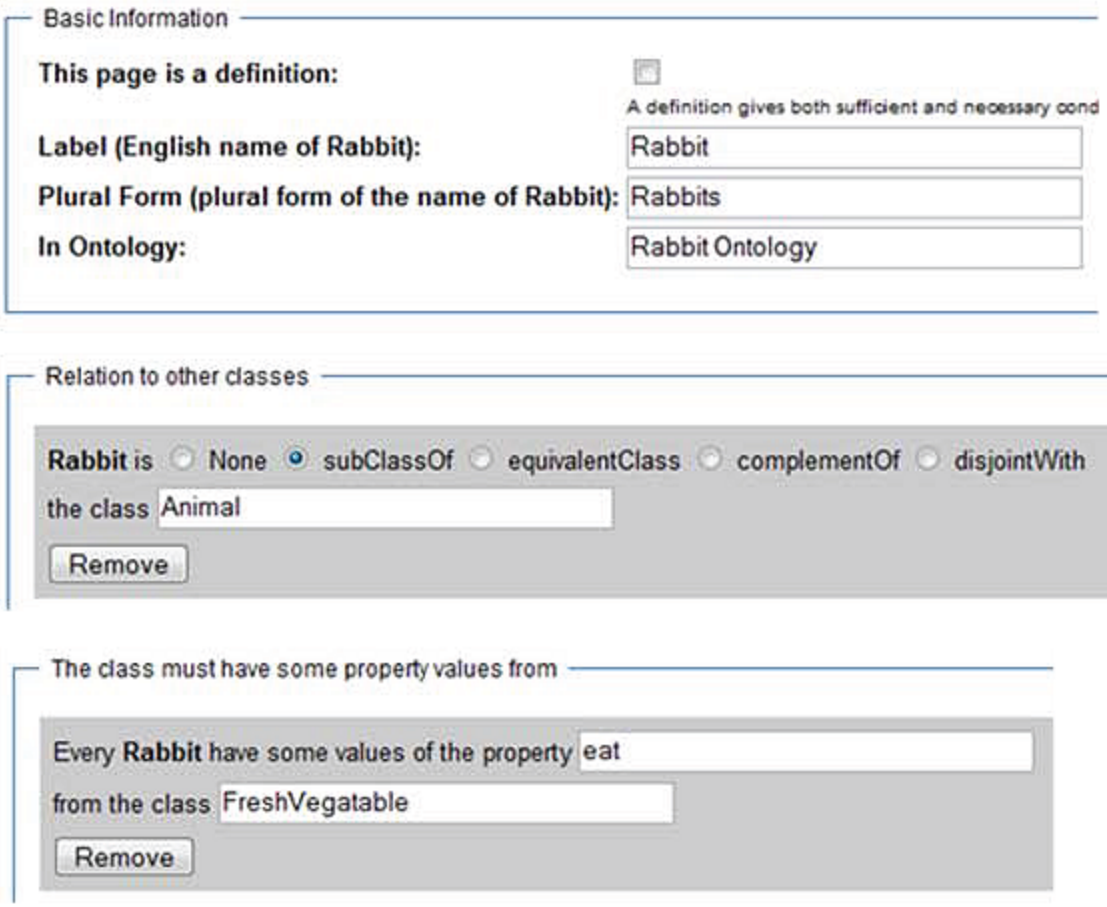


SMW-OWL Meta-Model

- Required for CNLs, RDF import, ontology development
- Use wiki templates to create OWL meta-model extensions for SMW
- Each wiki template is created using the wiki scripting language
- OWL elements (e.g. classes, subClassOf axioms) are represented using individual wiki templates
- Instances of the templates encode information about the classes, properties and individuals in a specific ontology

OWL Abstract Syntax:

Class(Rabbit partial intersectionOf (Animal
restriction(eat someValuesFrom(FreshVegatable))))

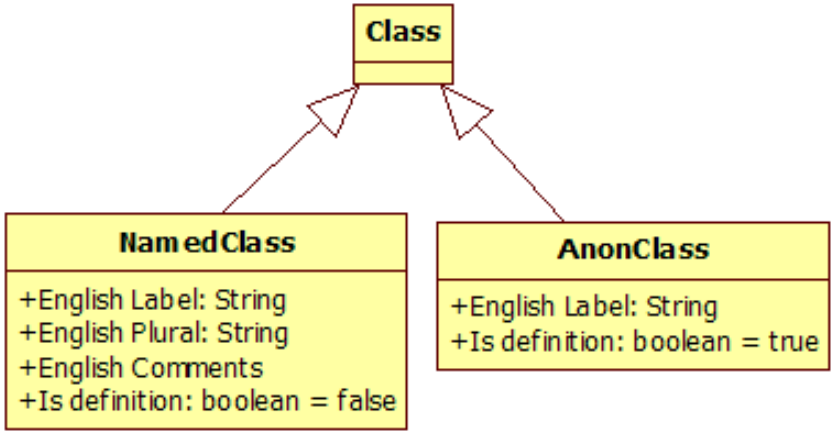



```
{{NamedClass
| is definition=No
| label=Rabbit
| plural=Rabbits
}}
```

```
{{NamedClassRelation
| type=subClassOf
| class=Animal
}}
```

```
{{someValuesFrom
| on property=eat
| on class=FreshVegatable
}}
```

OWL Class Templates



NamedClassRelation

+type: {subClassOf, equivalentClass, complementOf, and disjointWith}
+class: Class

AnonClassRelation

+type: {subClassOf, complementOf}
+class: Class

ClassListRelation

+type: {intersectionOf, unionOf}
+classes: List<Class>

someValuesFrom

+on property: Property
+on class: Class

allValuesFrom

+on property: Property
+on class: Class

hasValue

+on property: Property
+has value: Individual

oneOf

+individuals: List<Individual>

Cardinality

+on property: Property
+cardinality: integer

maxCardinality

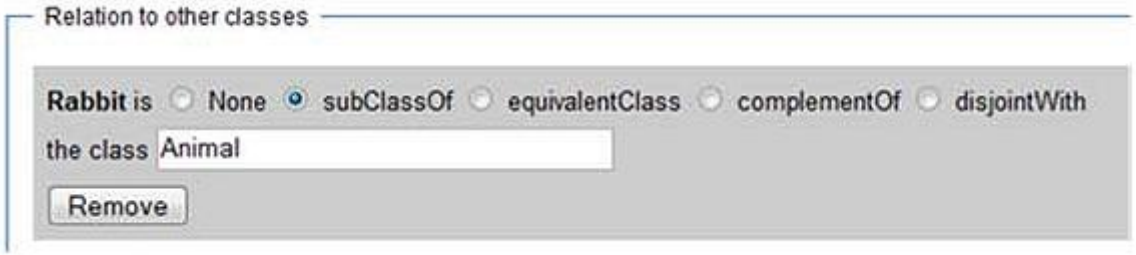
+on property: Property
+max cardinality: integer

minCardinality

+on property: Property
+min cardinality: integer

Form-Based Editing Interface

- Each wiki template is associated with UI components that support the editing of data associated with instances of the template




```
{{NamedClassRelation
| type=subClassOf
| class=Animal
}}
```

- Multiple templates can be associated with a wiki page to create an editing interface for ontology elements (i.e. classes, properties and individuals)

CNL Generation

- Wiki templates are also used to generate CNL
- Each wiki CNL generation template contains embedded semantic queries to retrieve information from the wiki database
- The retrieved information is then structured according to the syntax of the target CNL (e.g. Rabbit) – again using wiki script
- Accommodating new CNLs (e.g. ACE) requires relatively minor changes to the wiki script
  - future work: enable users to create/modify their own CNL generation templates

CNL Generation

http://tw.rpi.edu/proj/cnl/Template:CNL.Rabbit.getConceptRelationAssertions

```
{{#vardefine:label|{{CNL.getLabel|{{{1}}} }} }}
{{#vardefine:super |
{{#ask: [[:{{{1|{{FULLPAGENAME}}}}}]]
|?Category= |mainlabel=-|format=list|link=none
}} }}

{{#if: {{#var:super}}
|{{#arraymap:{{#var:super}}|,|xxx|<li>Every
[[:{{{1}}}{{!}}{{#var:label}}]] is a kind of
[[:xxx|{{CNL.getLabel|xxx}}]] }}|
}}
```

Category:Rabbit

Category:Rabbit [Edit]	
Label:	Rabbit
Plural:	Rabbits
In ontology:	Rabbit Ontology

"Category:Rabbit" in "Rabbit" controlled natural language

- **Rabbit** is a Animal.
- No **Rabbit** is a NonRabbit.
- **Rabbit** and Hare are equivalent.
- **Rabbit** and Wolf are mutually exclusive.
- Every **Rabbit** is exactly one of Bugs Bunny OR Peter Rabbit.
- Every **Rabbit** is a White Rabbit or a Black Rabbit.
- Every **Rabbit** eats FreshVegatable,
- Every **Rabbit** has part Whisker.
- Every **Rabbit** has child(ren) only **Rabbit** or nothing.
- Every **Rabbit** has eye color of Red.
- Every **Rabbit** has leg(s) exactly 4.
- Every **Rabbit** has head at least 1.
- Every **Rabbit** has parent at most 2.
- **Rabbit** is a concept, plural Rabbits.

CNL Editor Requirements

- Interface to support the creation and editing of wiki content using CNLs
  - light-weight integration with SMW environment
  - language agnosticism - support for multiple CNLs
    - Rabbit, ACE-OWL, etc.
    - requires flexible representation of grammar rules
  - constrain user input to grammatically-correct sentences
    - intellisense / autocompletion capabilities
  - display, sort, filter, search asserted CNL sentences
  - view 'related' sentences in other ontologies
  - view inferred CNL sentences
    - display reason why sentence has been inferred
  - provide logical consistency checking, redundancy checking and error diagnosis
  - speech input?
  - enable users to create/modify CNL input grammars?

AceWiki

- Developed by Tobias Kuhn at the University of Zurich
- Wiki system based on a subset of ACE
- Includes predictive editor that constrains user input to ACE-compliant sentences

- Differences:
  - underlying wiki system
  - editing interface
  - light-weight extensions
  - support for multiple CNLs
  - customization of target CNLs

Applications

- Coalition Planning
  - ontology-mediated collaborative planning
  - users
    - brigade staff
- Human Terrain Analysis
  - cultural profiling
  - cultural analysis and training
  - users
    - cultural anthropologists, psychologists, IO/PSYOP teams, indigenous individuals/organizations
- Intelligence Gathering/Analysis
  - social network analysis
  - activity monitoring
  - users
    - intelligence analysts, platoon leaders

Summary

- We have developed an OWL meta-model extension to SMW to support the representation of OWL ontologies
- We have provided a light-weight form-based interface to support ontology editing
- We have provided an RDF import mechanism to support the import of existing ontologies
- We have developed multiple CNL 'verbalizers' to support the serialization of semantic wiki content to CNLs
- Future work:
  - enable users to create/customize CNL output
  - implement wiki-based CNL editing capability

http://tw.rpi.edu/proj/cnl/Main_Page

baojie@cs.rpi.edu