



WP11 — AceWiki in MOLTO

Kaarel Kaljurand, Uni Zurich

MOLTO-EEU Kickoff,

Göteborg

2012-01-13

WP11: Multilingual Semantic Wiki

WP11 as proposed

Objectives The main goal of the proposed work-package is to build an engine for a multilingual semantic wiki, where the involved languages are precisely defined (controlled) subsets of the 15 languages that are studied in the MOLTO project.

Description of work The wiki engine would allow the input and presentation of the wiki content in all the languages, and perform formal logic based reasoning on the content in order to enable e.g. natural language based question answering. The users of the wiki can contribute to the wiki in any of the supported languages by adding statements to the wiki, as well as extending its concept lexicon. The wiki would integrate a "predictive editor" that helps the user cope with the restricted syntax of the input languages, so that explicit learning of the syntactic restrictions is not required. Ideally, the wiki would also integrate semantics-support, e.g. a paraphraser and a consistency-checker that could be used to enhance the quality of the wiki articles. The wiki engine is going to be implemented by combining the resources and technologies developed in the MOLTO project (GF grammar library, tools for translation and smart text input) with the resources and technologies developed in the Attempto project (Attempto Controlled English, AceWiki). The task of WP11 will be to combine the technologies developed in the MOLTO project with ACE and AceWiki, concretely: **(1)** porting the ACE grammar from English to the 15 MOLTO languages. The work in this task will be supported by the other MOLTO work-packages who are involved in developing GF-based grammars; **(2)** extending AceWiki to allow input in multiple different languages, i.e. develop AceWiki into a multilingual controlled language wiki. This task includes work on modularizing AceWiki and integrating existing GF tools for translation and smart text input; **(3)** using existing ACE application domains and test cases to evaluate the new multilingual wiki-system.

Multilingual Semantic Wiki

- *multilingual*
 - multilingual interface for editing and querying
 - multilingual content (synchronized!)
- *semantic*
 - notions of consistency, entailment, automatic Q&A
- *wiki*
 - collaborative, fast and accessible
 - user-friendly

Background technologies

- Grammatical Framework and its Resource Grammar Library
- Attempto Controlled English (FOL language with English syntax)
- AceWiki

AceWiki

Goal

User-friendly collaborative creation of expressive knowledge bases, where

- knowledge base is
 - computer processable, i.e. not plain natural language like Wikipedia
 - expressive beyond simple *subj-pred-obj* triples, unlike most *semantic wikis*
- collaborative and user-friendly
 - multiple users with
 - no background in formal methods

History

- Started in ~2006 in the Attempto project (part of the REVERSE project)
- Mostly work by Tobias Kuhn
 - see: *Tobias Kuhn. Controlled English for Knowledge Representation. Doctoral thesis, Faculty of Economics, Business Administration and Information Technology of the University of Zurich, 2010*

Main features

- Wiki content written in a subset of ACE
 - declarative and interrogative sentences
 - subset defined by a *Codeco* grammar
 - the grammar targets the OWL-compatible fragment of ACE
- Collaborative editing (web-based)
- Predictive (look-ahead) editing
 - simplifies entry of content
 - excludes syntactically impossible content
- Reasoning (based on standard OWL reasoning)
 - ACE -> OWL translator
 - every input sentence is tagged as *blue* or *red*
 - automatically generated views (e.g. sub/super concept hierarchy)
 - every question is automatically answered

Article (screenshot)



Navigation:

- [Main Page](#)
- [Index](#)
- [Random Article](#)
- [Search:](#)

Actions:

- [New Word...](#)
- [Export...](#)


planet

- ▶ *We use here the definition of "planet" according to the International Astronomical Union (see <http://www.iau.org>) without the restriction to solar planets.*
- ▶ Every planet is a celestial-body.
- ▶ No planet is a star.
- ▶ No planet is a dwarf-planet.
- ▶ *The distinction between planet and dwarf-planet has been introduced by the International Astronomical Union in 2006.*
- ▶ No planet is a moon.
- ▶ Every planet orbits a star.
- ▶ Every planet that does not orbit the Sun is an extrasolar planet.
- ▶ Which planets orbit the Sun?
 - Earth
 - Jupiter
 - Mars
 - Mercury
 - Neptune
 - Saturn
 - Uranus
 - Venus
- ▶ Every planet is a terrestrial planet or is a gas giant.
- ▶

Word creator (screenshot)

Word Creator ✕

Proper Name Noun Of-Construct **Verb** Transitive Adjective

 Every verb represents a certain relation between things. For example, the verb "owns" relates persons to their possessions.

third singular *
examples: owns, applies to, touches

bare infinitive *
examples: own, apply to, touch

past participle
examples: owned, applied to, touched

* required field

OK Cancel

Articles and words

Every wiki article has a 1-1 correspondence to a content word

Word

- word classes: noun, transitive verb, proper name, *of*-noun, transitive adjective
- corresponding logical notions: individual, concept, (binary) relation
- morph. forms: singular, plural, past participle

Article

- sequence of sentences, questions and informal comments
- has corresponding views (automatically generated and uneditable), e.g. *references, hierarchy*

Predictive editor (screenshot)

ACE Text Editor



Every country that has a capital controls ...

< Delete

text

function word

a
an
at least
at most
every
everybody
everything
exactly
less than
more than
no
nobody
nothing
somebody
something
what

proper name

new...
Austria
Berlin
Bill
Canada
France
Germany
Italy
Jane
John
Lara
Mary
Mike
Paris
Rome
Russia

new variable

X
Y
Z
X1
Y1
Z1
X2
Y2
Z2
X3
Y3
Z3
X4
Y4
Z4
X5

reference

itself
it
the capital
the country

OK

Cancel

Sentences, questions, comments

Declarative sentence

- checked by the reasoner when added/edited
- "red" sentence: not used in reasoning
 - would introduce inconsistency, or
 - expresses something that cannot be expressed in OWL (or its fragment)

Question

- reasoner supplies the answer and updates it if needed
 - answer = set of words

Comment

- ignored by the reasoner
- plain wiki text (with `[[hyperlinks]]`)

Reasoning (screenshot)

The screenshot shows the AceWiki interface. On the left is a sidebar with the AceWiki logo and navigation links. The main content area displays a list of logical rules under the heading 'birth-place'. A dialog box titled 'Conflict' is overlaid on the right, indicating that a sentence is in conflict with the current knowledge base.

AceWiki
AceWiki test

Navigation:
- [Main Page](#)
- [Index](#)
- [Search](#)
- [About](#)
- [Random Article](#)

Actions:
- [New Word...](#)
- [Export...](#)

← → ↻ Kaarel ✕

[Article](#) | [References](#) | [Individuals](#) | [Hierarchy](#)

birth-place

- ▶ No person has more than 1 birth-place.
- ▶ **A person has at least 2 birth-places.**
- ▶ Every person has a birth-place.
- ▶ John is a person.
- ▶ Bill is a fireman.
- ▶ Who has a birth-place?
 - Bill
 - John
- ▶

Conflict ✕

The sentence is in conflict with the current knowledge. For that reason, it cannot be added to the knowledge base.

OK

Reasoning via translation to OWL

Every country that does not border a sea is a landlocked-country.

```
SubClassOf(  
  ObjectIntersectionOf(  
    :country  
    ObjectComplementOf(  
      ObjectSomeValuesFrom(  
        :border  
        :sea  
      )  
    )  
  )  
)  
:landlocked-country
```

Which country is a landlocked-country?

```
ObjectIntersectionOf(  
  :country  
  :landlocked-country  
)
```

Missing features

... compared to other wiki systems:

- history management (diff, revert)
- user management (authorship, badges)
- *talk* (meta) pages
- API (edit, view, query)
 - enables third-party applications
 - enables alternative UIs

More details

- Version 0.5.2 (released: 2012-01-05)
- Written in Java using the Echo Web Framework
- Parsing
 - Codeco
 - Attempto Parsing Engine (APE)
- Reasoning
 - API: OWL API
 - Reasoners: Hermit, Pellet, FaCT++ (via OWLlink), ...
- Storage
 - ACE
 - filesystem: one file per article
 - import-export format
- Open source: GitHub, LGPL

AceWiki in MOLTO

Required extensions

- Multiple languages
 - English
 - Swedish
 - ...
- Multiple grammars
 - ACE
 - Phrasebook
 - Math
 - ...
- Multiple reasoners / entailment explanation (?)
 - ACE-based (RACE, OWL, ...)
 - Sage
 - ...

Multiple languages

- The content of every article viewable/editable/queryable in every language
- The wiki UI localized to every language

Multiple grammars

- Different grammars on different sections of the wiki
- Interaction between grammars
 - e.g. detecting sentences that can be parsed by multiple grammars
- Integrated grammar editor
 - full GF editor
 - UI for adding/editing words and their forms
- UI for switching between grammars

Status (January 2012)

Some recent work (after GF Summer School 2011)

- Modularity
 - different grammar engines (Codeco, GF)
 - separated backend
- Multiple grammars (via PGF/JPGF)
 - predictive editing (but without word classification)
 - translation
- Multiple languages
 - wiki URL for each language (see the *Foods*-grammar demo)

ACE in GF

ACE in GF

- General research topic: **multi-lingual CNL**
- Benefits for ACE:
 - non English speakers gain access to ACE
 - GF provides new ACE editing tools
- Benefits for GF:
 - access to reasoning

Existing GF-implementations of ACE

Angelov and Ranta

- K. Angelov and A. Ranta. *Implementing Controlled Languages in GF*.
- based on ACE v6.0 (early 2009)
- covers 7 languages (multi-linguality via the GF RGL)

Normunds Grūzītis

- <http://valoda.ailab.lv/cnl/>
- only Latvian
- only OWL-compatible fragment
- ACE as hidden machine-readable format (i.e. no wordform support)

Kaljurand

- based on *Angelov and Ranta* (with very minor updates)
- <https://github.com/Attempto/ACE-in-GF>

ACE-in-GF current coverage

How many ACE sentences can ACE-in-GF parse?

- AceWiki subset of ACE
 - 20k exhaustively generated sentences of length up to 10 tokens
 - 30% are parsed by GF
 - minor (?) problems related to indefinite pronouns and relative clauses
- Full ACE
 - 3k APE regression test sentences
 - 20% are parsed by GF
 - missing words

How much does ACE-in-GF overgenerate? Not tested yet (had some problems with GF generation).

Links

AceWiki development

- <https://github.com/AceWiki/AceWiki>

AceWiki demos and publications

- <http://attempto.ifi.uzh.ch/acewiki/>

Multi-lingual demo (JPGF-powered Foods-grammar)

- *Eng*: <http://attempto.ifi.uzh.ch/webapps/gfeng/>
- *Ger*: <http://attempto.ifi.uzh.ch/webapps/gfger/>
- *Ita*: <http://attempto.ifi.uzh.ch/webapps/gfita/>

ACE in GF

- <https://github.com/Attempto/ACE-in-GF>

Discussion

Open questions

- Interfacing (with GF)
 - future of JPDF? Should we use a GF webservice instead?
 - (REST) API for AceWiki (to provide more access points to the wiki)?
- ACE in GF
 - how (much) will UGOT contribute?
 - coverage: full ACE vs the AceWiki-supported subset?
 - how many languages?
- Reasoning
 - what type of "reasoning" will be part of AceWiki?
 - how do MOLTO case studies interact with ACE?
- Underlying storage format (ACE?, GF abstract tree?)
- ...

Thank You!