

# Molto Translation Tools

MOLTO  
Translation Prototype

Review

Lauri Carlson  
University of Helsinki

# Cast

- Lauri Carlson
  - TermFactory, translation tools
- Seppo Nyrkkö
  - Team manager, OntoR term harvester
- Lauri Alanko
  - GF c runtime, Haskell
- Inari Listenmaa
  - GF grammars, FactForge
- Chunxiang Li
  - Term editing

# MOLTO use scenarios

- Addition for professional translation workflow
  - To handle constrained sublanguage
- Multilingual content creation scenario
  - Multilingual semantic Wiki

# MOLTO CAT bottlenecks

- Too small language coverage for translators
- Too laborious to build a domain model
- Translation tools cannot solve bad coverage
- To go beyond fridge magnets need robust parser
- Conclusion
  - Concentrate on ontology and terminology acquisition tools
  - Until more robust parsing available
- Consequence
  - TT integration is behind schedule

# Networked interlingua

**“Not a monolith but a boulder field”**

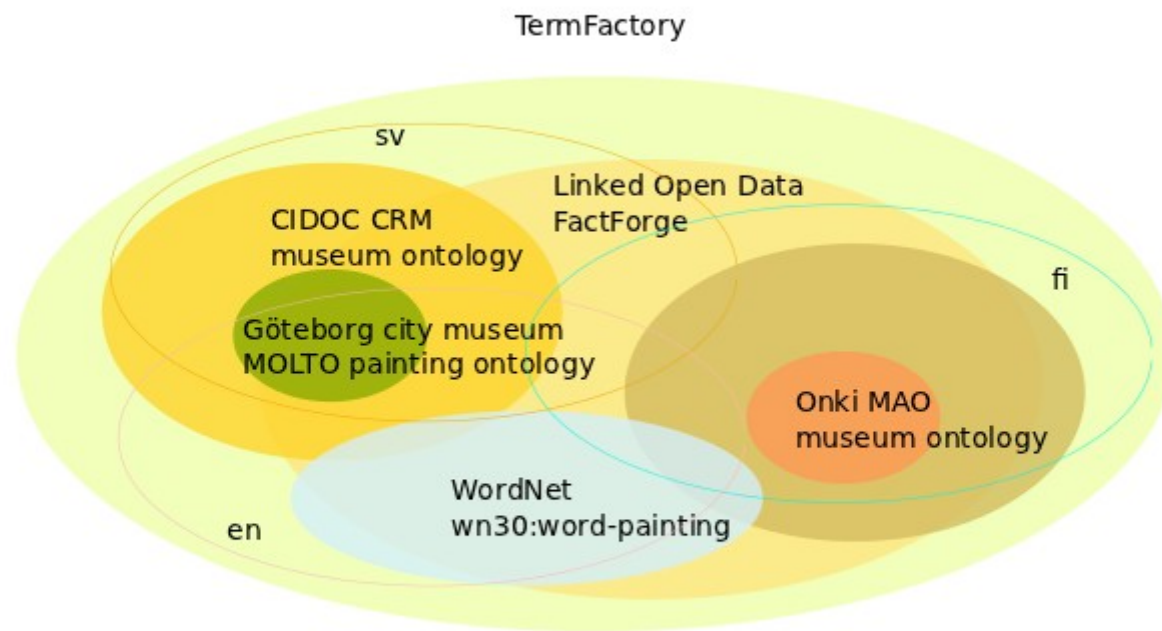
But boulders do not communicate

We want a network of small models that communicate.

Web ontologies are supposed to provide that -

But: **ontology hell**

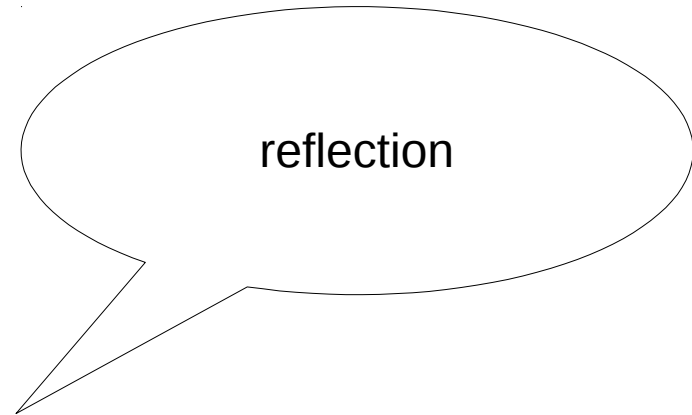
- Nobody bothers to look around before doing own
- Nobody bothers to bridge to others
- Homonymous local names => Synonymous global names



Ontological sources for painting  
one ring to rule them all...

# TermFactory provides

- RDF/XHTML
- Small ontologies
- Term ontologies
- Localization of URLs and URIs
- Missing: TF to GF mapping!



# TermFactory provides

**Globalisation: interlingual translation between locales**

- of names and addresses
- of concepts and terms
- of canned texts

LANG is just one parameter of locale – concepts too are local

Paradox: We want short names that are unambiguous.

Solution: Localization ontologies in which short names are unique



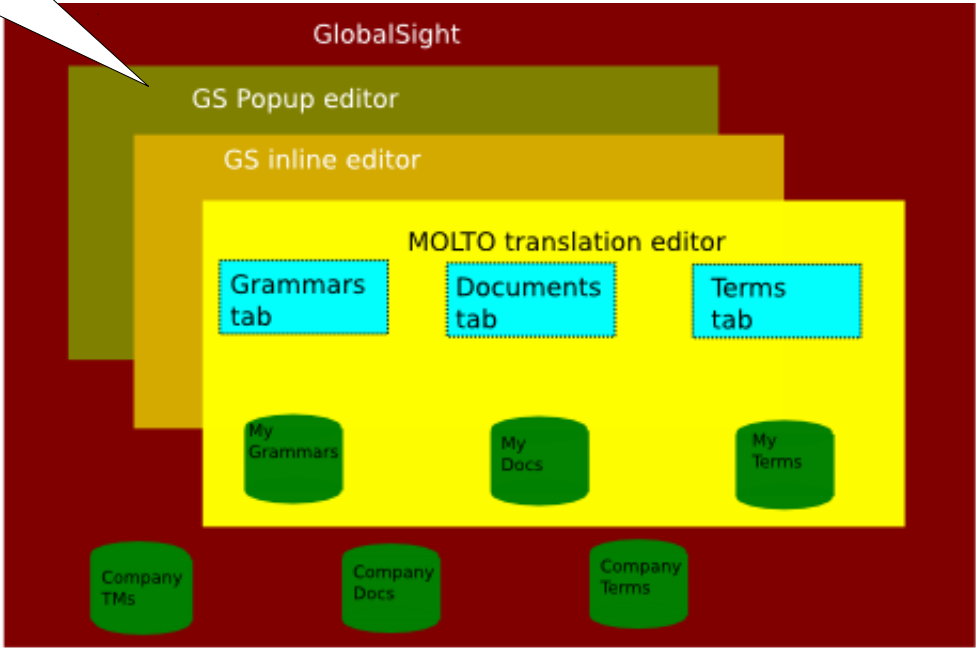
# MOLTO Translation tools

- independent pieces of software
- pluggable parts
- standards compliant
- open
  
- integration not completed

# Current parts

- MOLTO translation editor
  - Krasimir Angelov's GWT editor
  - Further developed at UHEL
- Editor term tab
  - MOLTO equivalents editor embedded in it
- MOLTO GlobalSight translation environment
  - MOLTO translation editor embedded in it
- TermFactory
  - TF term ontology editor and query form
  - TF Mediawiki
- FactForge

workflows



ontologies

Wikipedia Wiktionary

FactForge

OntoR

MOLTO SMT

OntoText MOLTO ontologies

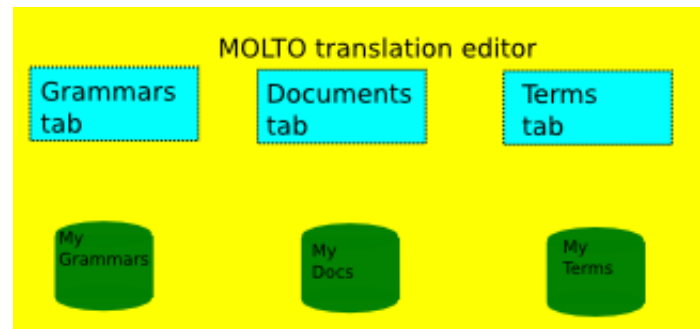
Grammar Editing

 TermFactory Wiki  
TF web services

grammar extension

term acquisition

MOLTO translation editor  
Controlled language translation editor  
Standalone, one-user, web-based

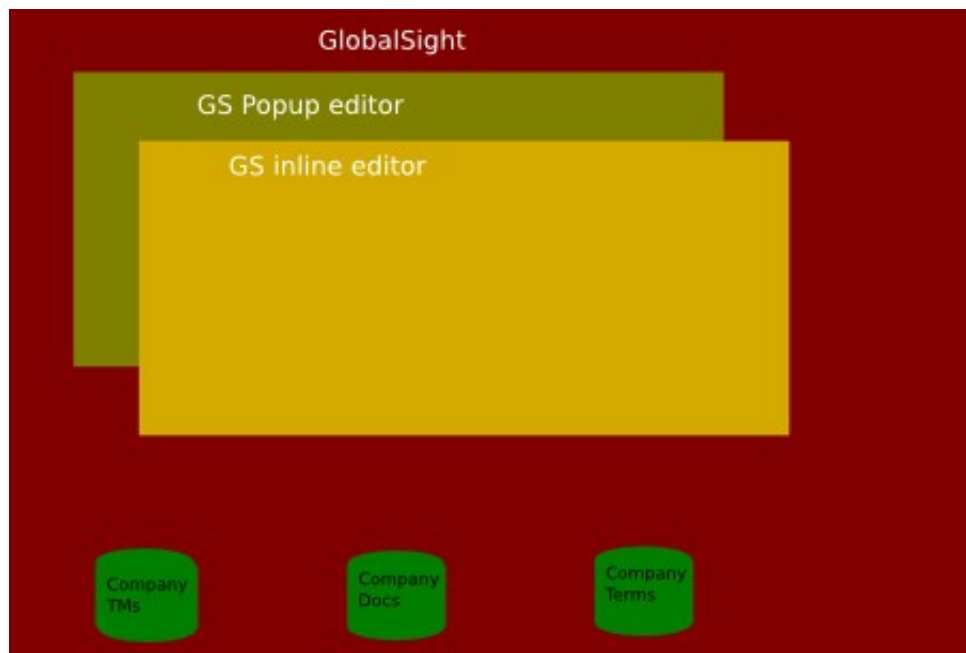


Not really usable yet  
Needs a more robust parser  
Should be made easier to combine with  
more traditional translation editing styles

# Translation editor

- A Google Web Toolkit application (K. Angelov)
- Simple database back end for managing
  - Documents
  - Grammars
  - Terms
- Translates with GF
  - autocompletion from grammar
  - fridge magnets
  - cut and paste text ...

## GlobalSight TMS (translation management system)



Provides typical CAT tools  
Supports commercial translation workflow

Extensive webservice API  
Somewhat dated

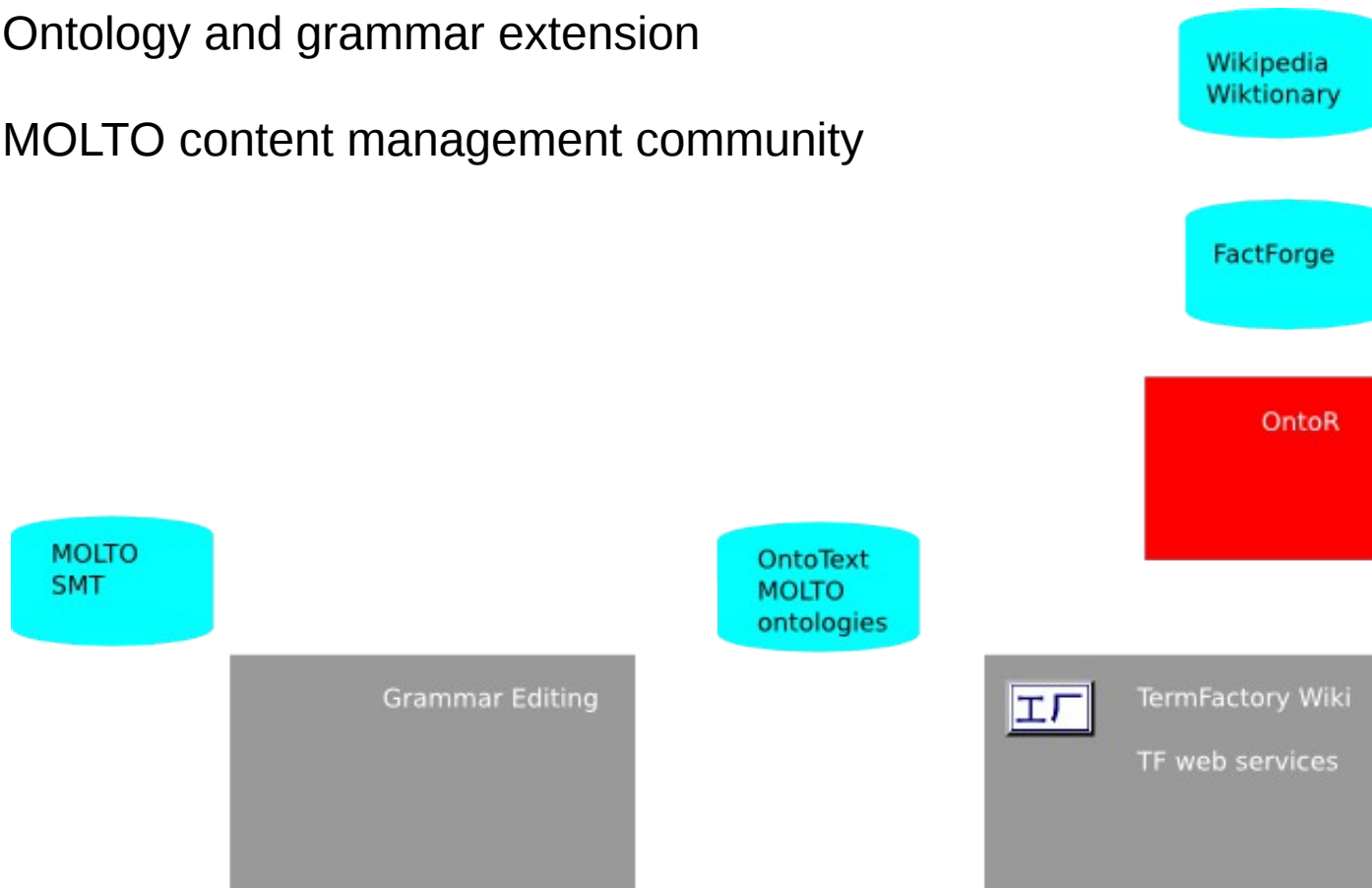
A testbed for trying out alternative MOLTO workflows

# GlobalSight TMS

- Translation management system
- Oriented toward translation company scenario
- Workflow management
- Translator pool management
- TM, TB, web CAT editors
- Email, authentication, roles ...
- open

## Ontology and grammar extension

MOLTO content management community





# TermFactory

- Ontology and term management
- Sharing of concepts and terms
  
- Links translators, grammarians and ontologists
- Manages the boulder field (well, semantic web)
  
- Access to the Web of Data
  - OntoText KRI
  - FactForge
  - OntoR

# TermFactory

## **Provides**

- Globalisation of concepts
  - Ontology management
- Globalisation of names
  - Term ontology management
- Globalisation of addresses
  - Location mapping
- Crowdsourcing
  - Wiki and comment system

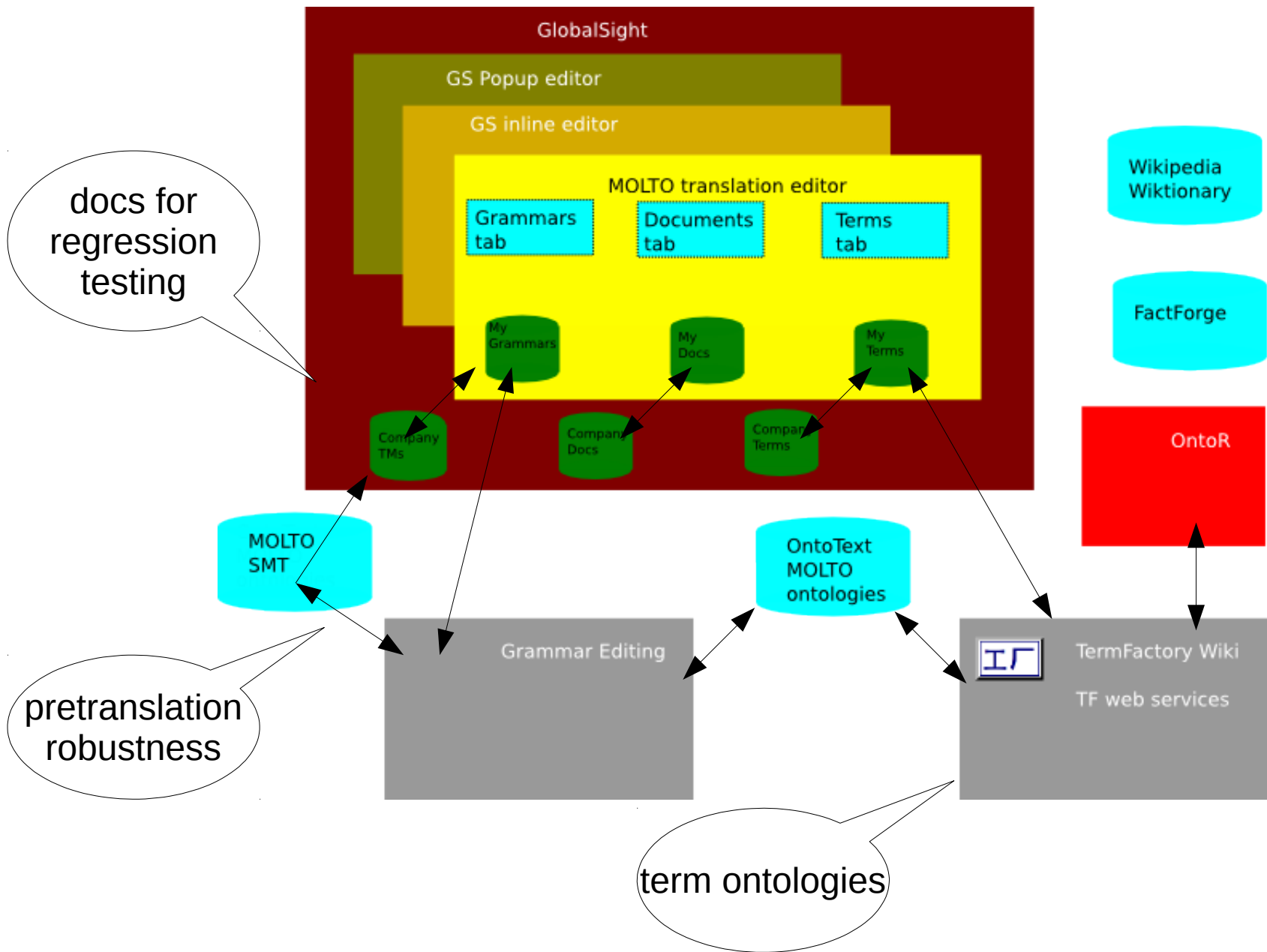
# TermFactory

## **Main facilities**

- Ontology query
  - Mix and match small and large ontologies
- Ontology editor
  - Edit ontologies as html
- Ontology rewriting
  - Manage conversions and refactor names
- Location mapping
  - Use local addresses for global resources
- Location indexing
  - Manage versions in different locations

# Putting it together

- integrate
  - Translation editor with TMS (Globalsight)
    - accept/submit jobs, share resources
  - Translation editor with TF
    - get terms and submit term proposals to community
  - Translation editor with GF
    - translate, browse grammar/s, add terms
  - TF with GF
    - provide terms with grammar information
    - Associate words with concepts
  - TF with OntoText



# Editor + TMS

- Translation editor to GlobalSight
  - (or other TMS, eg. TranslateWiki:)
- Up/download n-texts between editor and TMS
- Formatting and layout management
  - XLIFF (n=2?)
- Save/retrieve translation units between editor and TM
  - TMX
- Communicate terms between TB and TF
  - TBX

# Editor to GF

- GF web API
  - [http://localhost/my\\_grammar.pgf?command=cmd](http://localhost/my_grammar.pgf?command=cmd)
    - grammar
      - A description of the grammar
    - parse (cat input from)
      - parse input as cat from lang
    - linearize (tree to)
      - Generate abstract syntax tree to lang
    - translate (cat input tree to)
    - complete (cat input from limit)

# Editor to GF

- GF web API cont.
  - [http://localhost/my\\_grammar.pgf?command=cmd](http://localhost/my_grammar.pgf?command=cmd)
    - random
    - abstrtree
    - parsetree
    - Alignment
  - Wanted:
    - Robust parsing
    - Add words to grammar
    - Larger test grammars



# Editor to TF

- Editor term tab
  - Equivalents editor
    - Specify sources
      - Merge results from different sources with source indication
    - Access terms in translation
      - Extract from translation
      - Insert into translation
    - Search terms from TF
      - Submit a query to TF
      - Get a table of equivalent
    - Submit term proposals to TF
      - Convert equivalents table to a TF ontology
      - Upload it to TF or edit a TF ontology with it

# TF web API

- Query
  - String query;
    - string to be parsed as a SPARQL query string or a list of documents or resources
  - String queryType;
    - URL, URI, ANY
  - String answer;
    - returned document
  - String[] repos;
    - array of repository names
  - String engine;
    - name of query engine to use (null for default)
  - ...
  - String message;
    - message returned from service

# TF web API

- Edit
  - String op;
    - string parsed as edit operation.
  - String rewrite;
    - a rewrite command
  - ...
  - String answer;
    - returned document
  - String message;
    - message returned from service

# TF to GF

- The leading idea:
  - A TF ontology can associate
    - to a TF concept one or more GF abstract syntax terms from some GF abstract grammar
    - To a TF term in some language one or more GF concrete syntax terms from some GF concrete grammar
  - The association can be by enumeration or it can be generated by rule (including GF)
    - enable GF verbalisation of parts of the ontology in NL (or parsing some NL into ontology concepts)
    - mix and match existing GF grammars in new domains (for instance borrow terms from a more general domain)

# TF to GF

- Concepts to abstract syntax
  - How to map TF concepts to GF abstract syntax terms
- Terms to concrete syntax
  - How to associate GF concrete syntax with TF terms
- Museum case
  - Dana and Ramona
- Generalise, plug in TF terms

# TF to Ontotext

- Tight integration
  - Add OWLIM as a TF database backend
    - Already have Jena RDB and TDB

## Loose integration

- SPARQL / Update endpoint
  - There is a SPARQL Axis2 service endpoint for TF
  - Must see about TF editing through SPARUL queries

# More or less understood

- More (WP3)
  - Molto editor vs Globalsight
    - (or TranslationWiki)
  - Molto editor vs TermFactory
  - TermFactory vs Ontotext
- Less (WP2)
  - Grammar editing
  - GF vs SMT
  - GF vs Ontotext

# To be added

- OntoText repository for TermFactory
  - SPARQL query/update endpoint
- Robust parsing
  - How to use in editor
- PGF on-the fly compilation
  - Adding new terms to grammar
- OntoR
  - Term / ontology harvesting tool from text



# To figure out

- GF grammar – ontology interface
  - Ramona, Dana
  - How to interface extension with translation?
- SMT hybrid/robust translation
  - Through translation memory?
  - Through GF translation back end server?

# Alternatives

- Web localisation scenario
- ACEWiki
- TranslateWiki
- Translation and ontology development on one platform
- Creating multilingual content

# WP 9 – evaluation types

- Progress / diagnostic evaluation per site
  - Being done by each site as appropriate
    - Collect the results
- p/d evaluation for integration/s
  - Bug/feature request tracking system/s (common/per subsystem)
- Adequacy evaluation
  - Usability
    - User feedback
    - Tests
  - Translation quality
    - Automatic
    - Manual
- Quantitative measures
  - some in DoW

# WP 9 p/d evaluation/site

- Progress / diagnostic evaluation per site
  - Automatic
    - SMT
    - Others?
  - Manual
    -
  - Quality dimensions, criteria, measures
    -
  - Where to get data for reporting

# WP 9 p/d for integration

- Progress / diagnostic evaluation for integration
  - Bug tracking system/s
    - trac
  - How can we get others involved to evaluate ours
    - easy if they need to integrate with it
    - very hard otherwise
- Can we create more cases of the easy type?
  - John, Tomas ↔ ?
  - Jordi ↔ ACEWiki

# WP 9 usability

- Adequacy evaluation for usability
  - Users / user communities
    - \_ Math formula verbalization, exercise book
    - \_ Generating museum descriptions
    - \_ GF grammar editing
    - \_ Translation of patents
    - \_ Terminology/ontology management
  - What platforms?
    - \_ TF wiki
      - for ontologies/terms
    - \_ ACEWiki
      - for museums and math
    - \_ GlobalSight
      - for patents, for translation treebank revision
    - \_ Trac
      - for integration
  - How to activate users to contribute?
    - \_ How to collect responses
      - Visit counts, 'like' buttons
      - Disqus comment tags
      - Paid job submission

# WP 9 translation quality

- Adequacy evaluation for translation quality
  - SMT
    - automatic measures
    - human evaluation
      - Hire professional patent translators in June
    - compare to Google
  - GF
    - What to evaluate?
      - Museum, Math, Patents, extensions?
    - Fidelity
      - Maarit Koponen (diss.)
    - Fluency
      - BLEU, UPC tools
    - Fitness for purpose
      - TAUS