Wikimania 2009

WikiWord

Multilingual image search and more

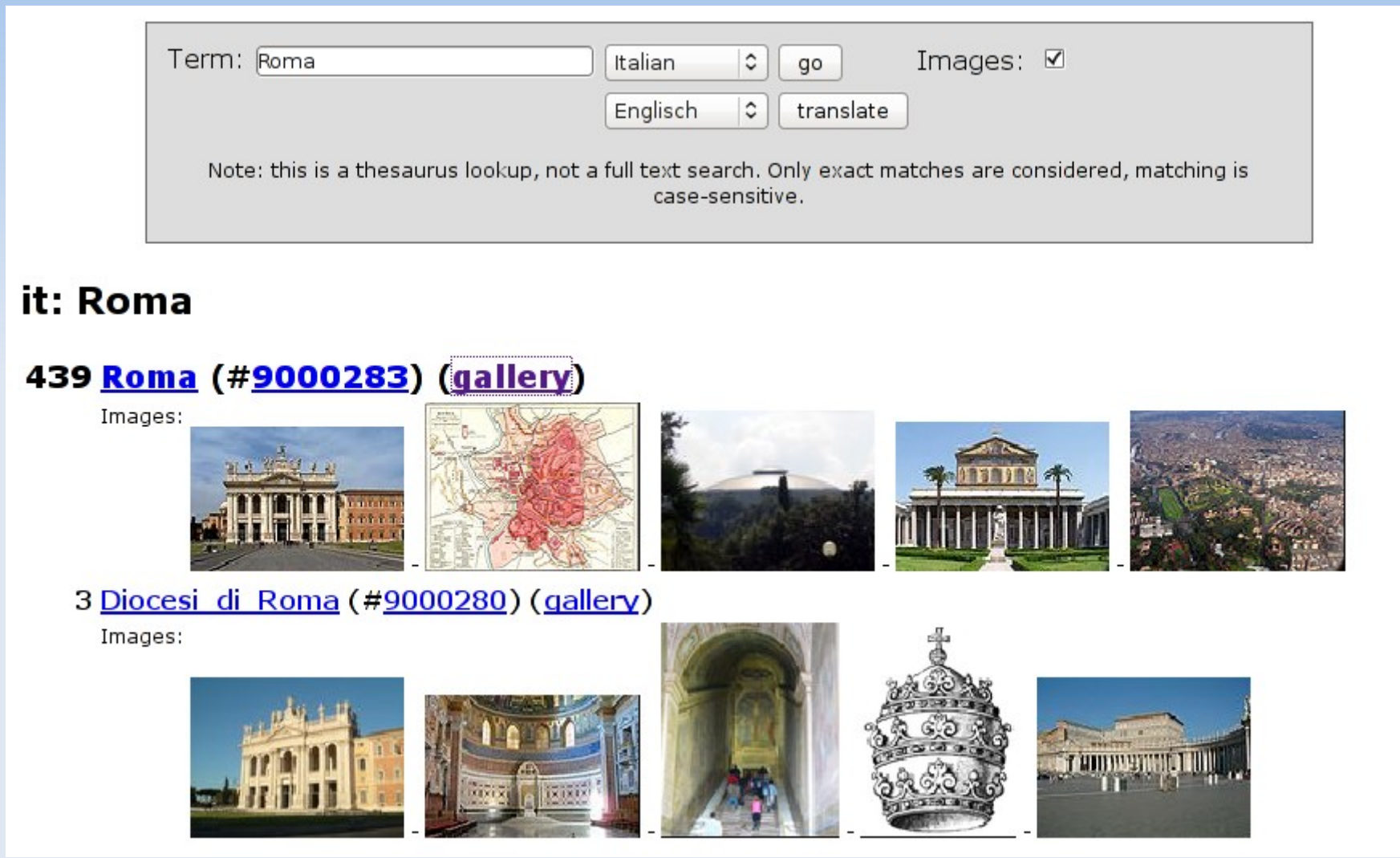
Daniel Kinzler

WIKIMEDIA DEUTSCHLAND

Image Search

1. Specify language, enter a word (term)
2. Find meanings for term (topics/concepts/pages)
3. Find galleries and categories from Commons
4. List concepts, show images for each

Image Search



http://toolserver.org/~daniel/wikiword-trunk/wikiword.php

Wikipedia

- **Goal:** ***find and navigate topics***

- Wikipedia describes topics (concepts)
- Links define relations
- Links and titles define terms for concepts
- Interwiki-links provide cross-language concepts

Thesaurus

- Concepts + Terms + Relations = ***Thesaurus***
- Uses:
  - list meanings
  - navigate topics
  - relatedness (semantic proximity)
  - disambiguate in context

WikiWord: Relations

- Pages linked both ways: *related* topics (mutual relevance)
- Language-links overlap: *similar* topics
- Categories: *broader*/*narrower* topics
- Category main article: *same* topic
- Language-links both ways: *same* topic

WikiWord: Terms

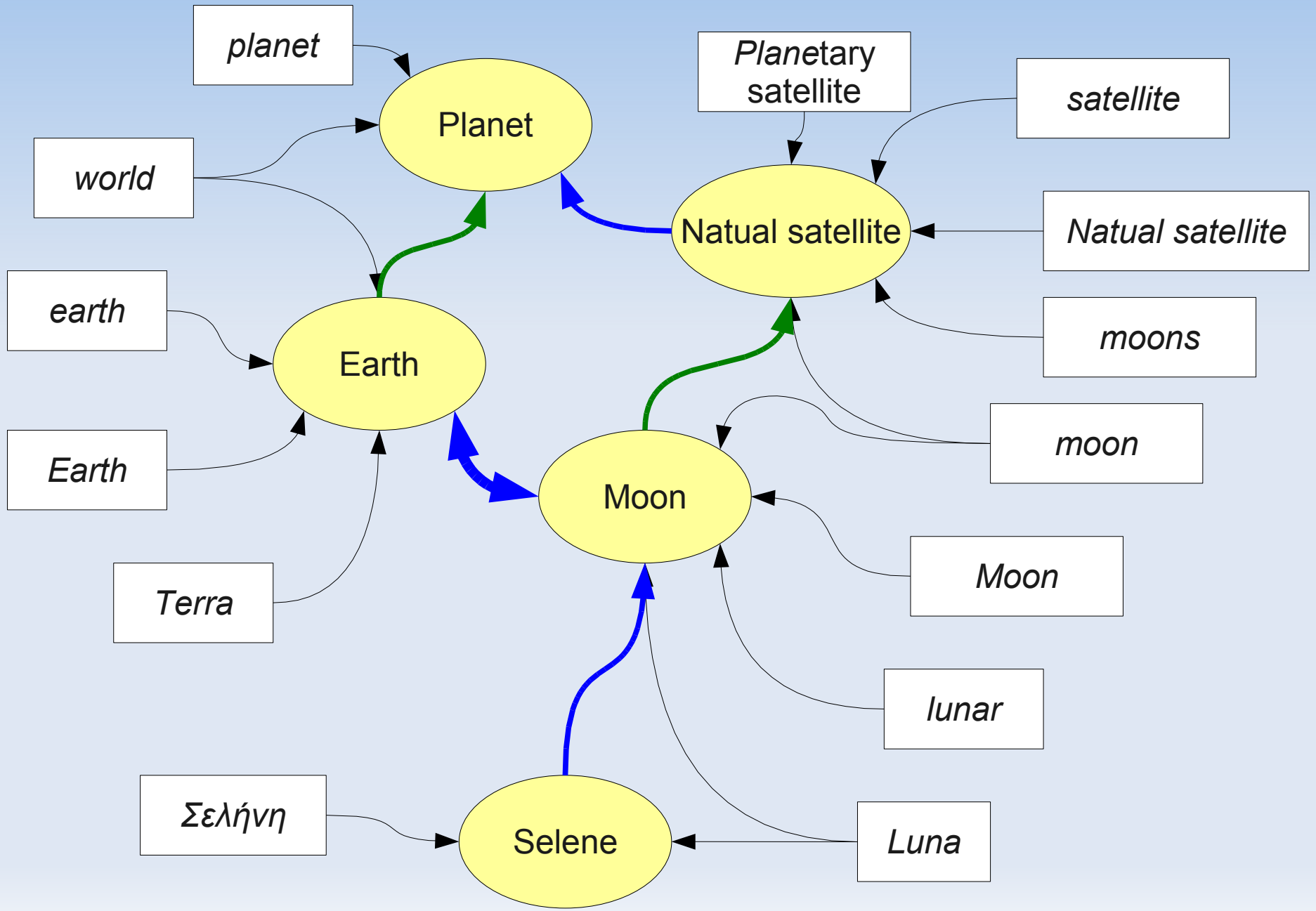
- Page titles
- Redirects
- Disambiguations
- **Link text!**
  - `[[Wikimedia Foundation|WMF]]`
  - *Fequency* of use
  - *Context* of use (not currently used)

WikiWord: Properties

- Infoboxes
- Biography templates
- Categorie names

- Rudementary
- Use DBpedia

Structure

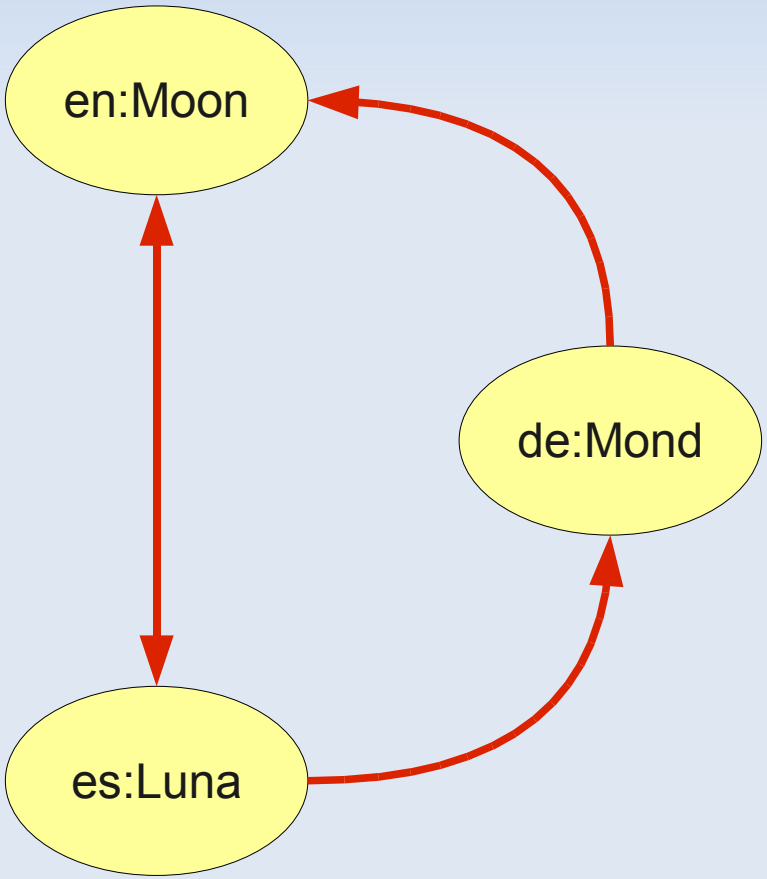

planet
Planet
world
earth
Earth
Earth
Terra
Planetary satellite
satellite
Natual satellite
Natual satellite
moons
moon
Moon
Moon
lunar
Luna
Σελήνη
Selene

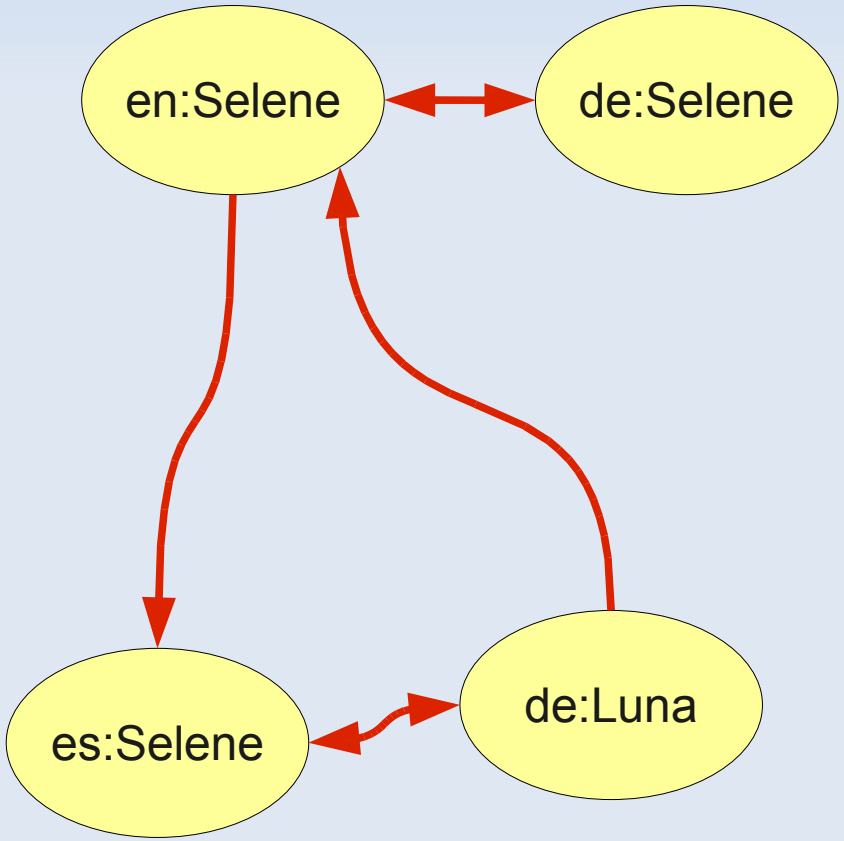
WikiWord: Merging

- Put concepts from all languages into one thesaurus
- Remember which language each concept came from
- If two concepts reference each other with language links, merge them
- Remember which languages are covered by a merged concept
- Each concept can cover each language only once

Merging

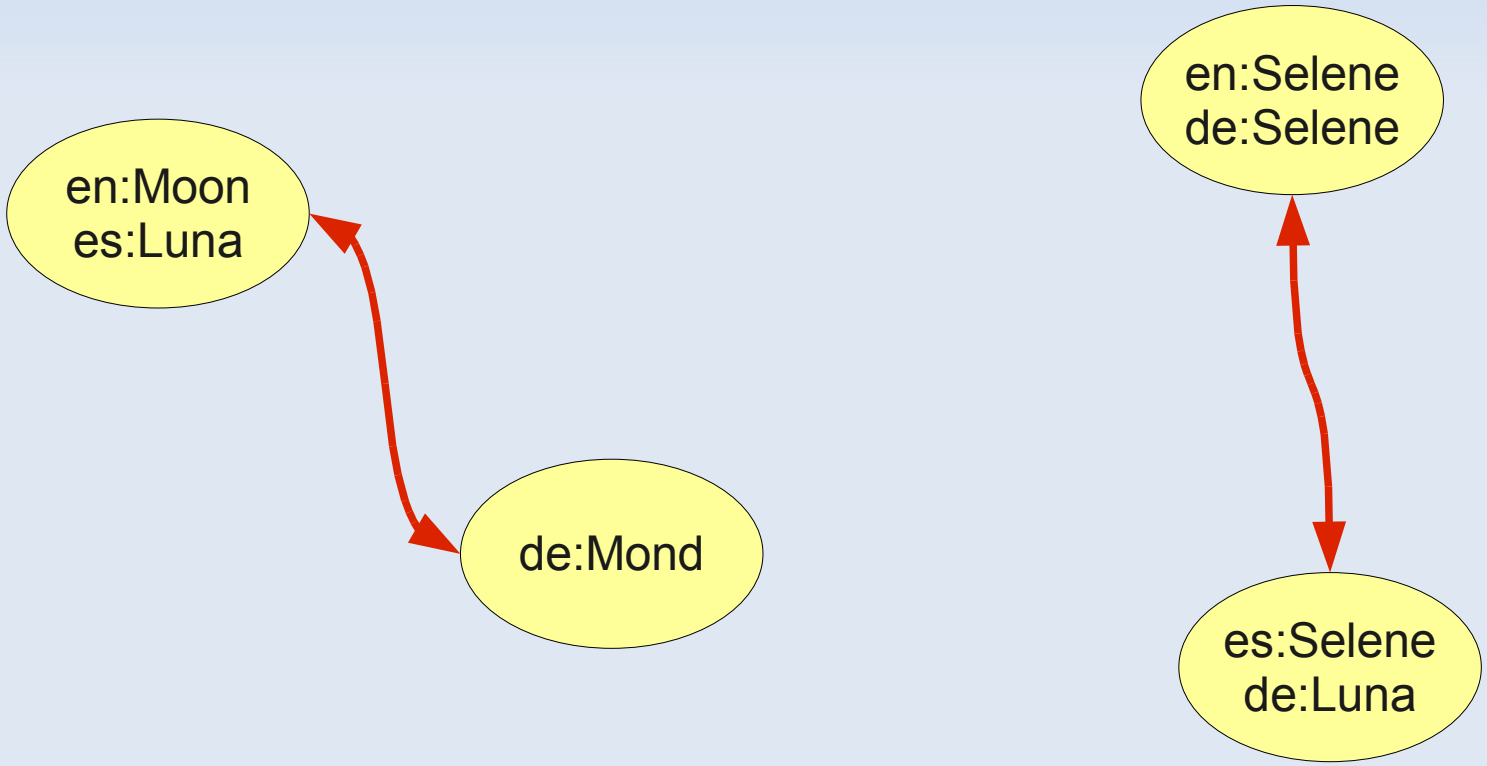
- Put concepts from all languages into one thesaurus
- Remember which language each concept came from





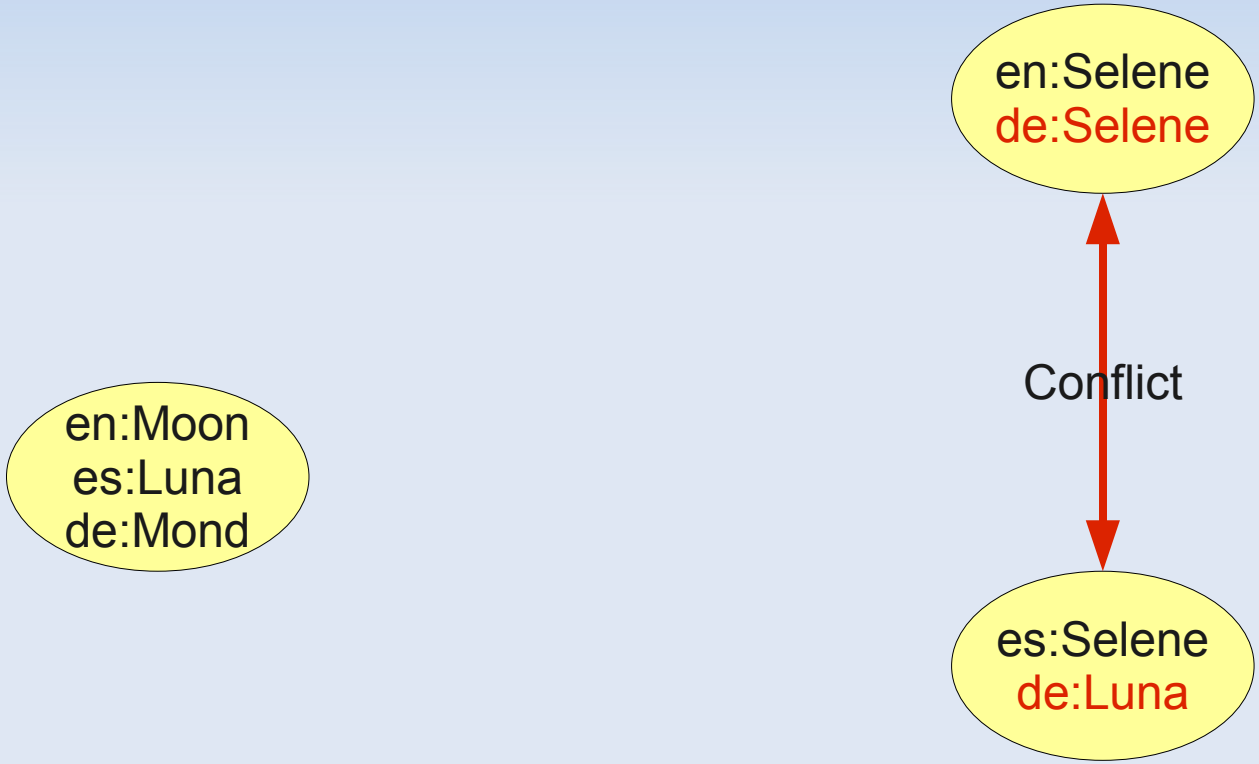
Merging

If two concepts reference each other with language links, merge them
Remember which languages are covered by a merged concept



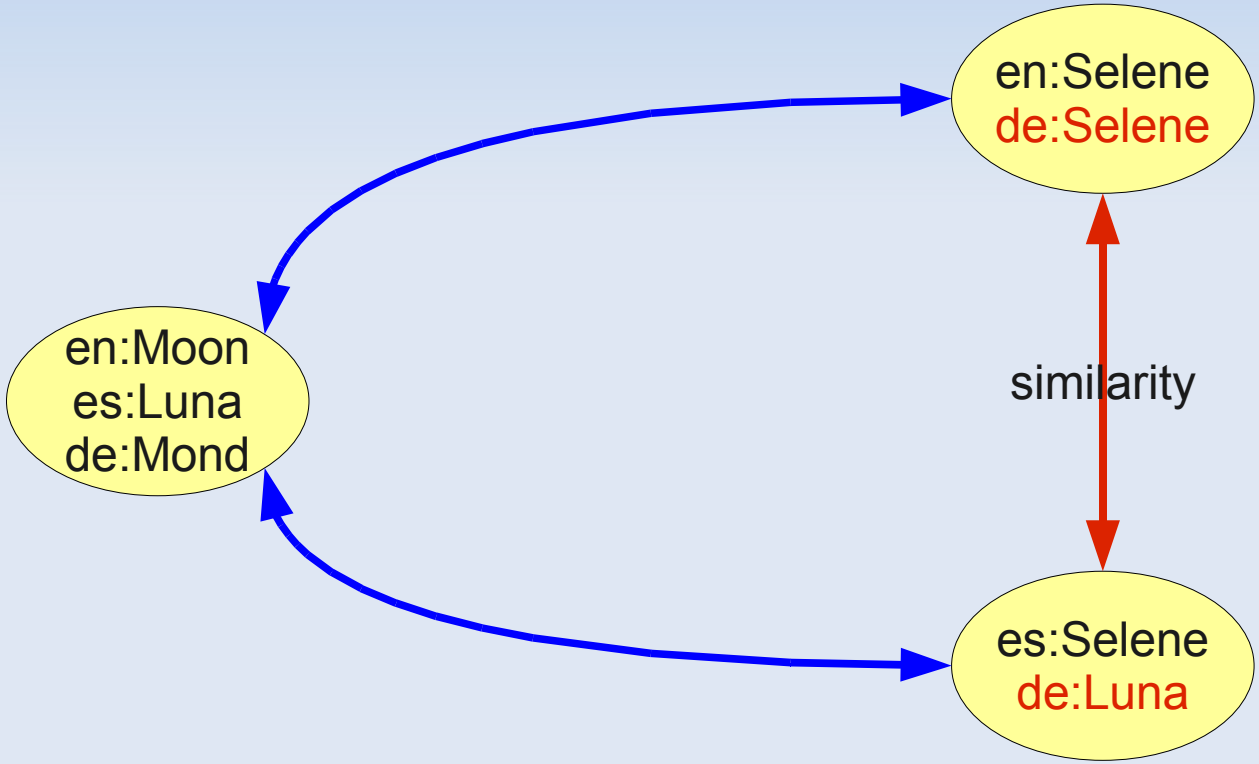
Merging

- Each concept can cover each language only once

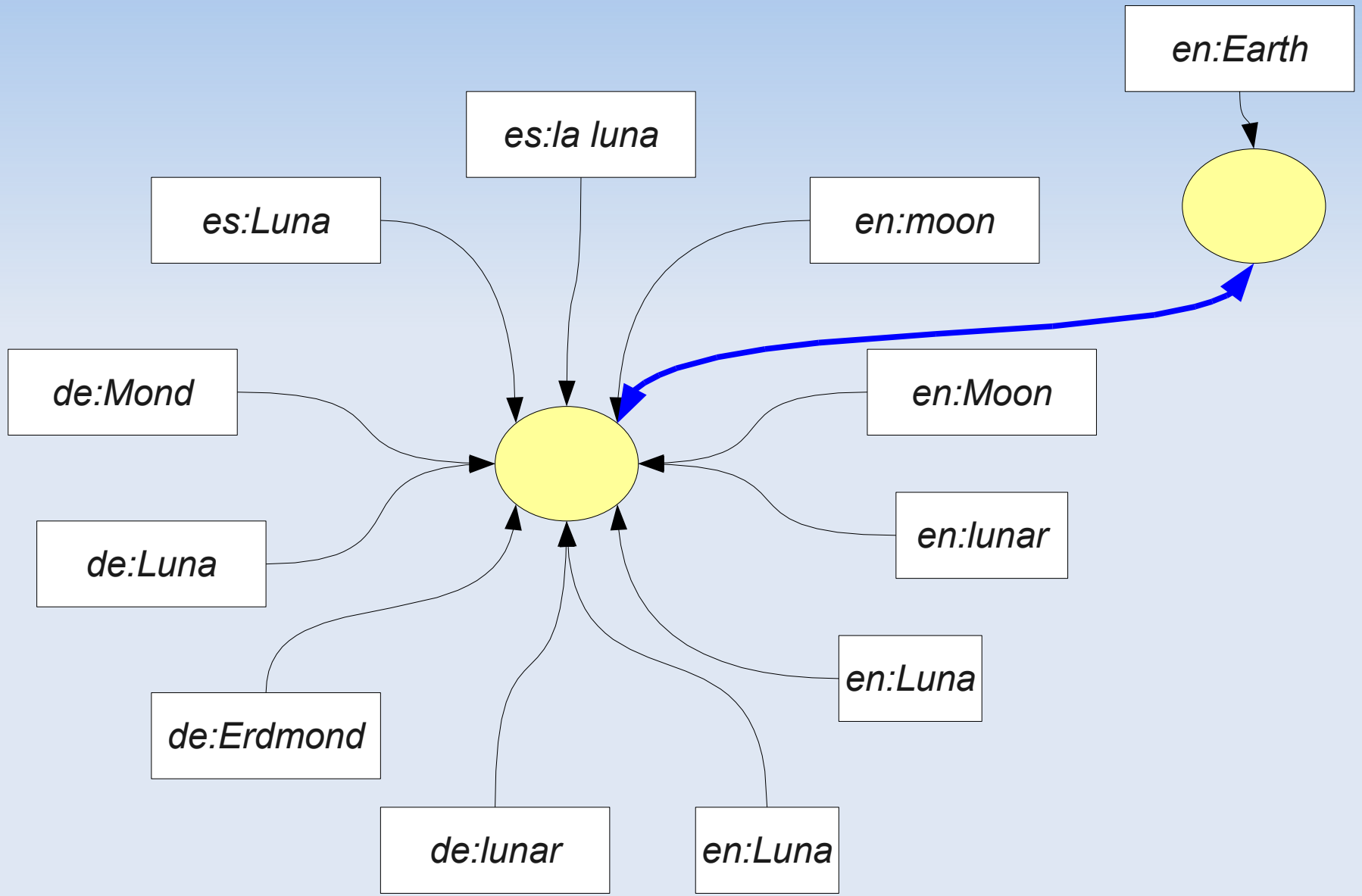


Merging

- Relations between concepts are accumulated



Multilingual Thesaurus



Using WikiWord

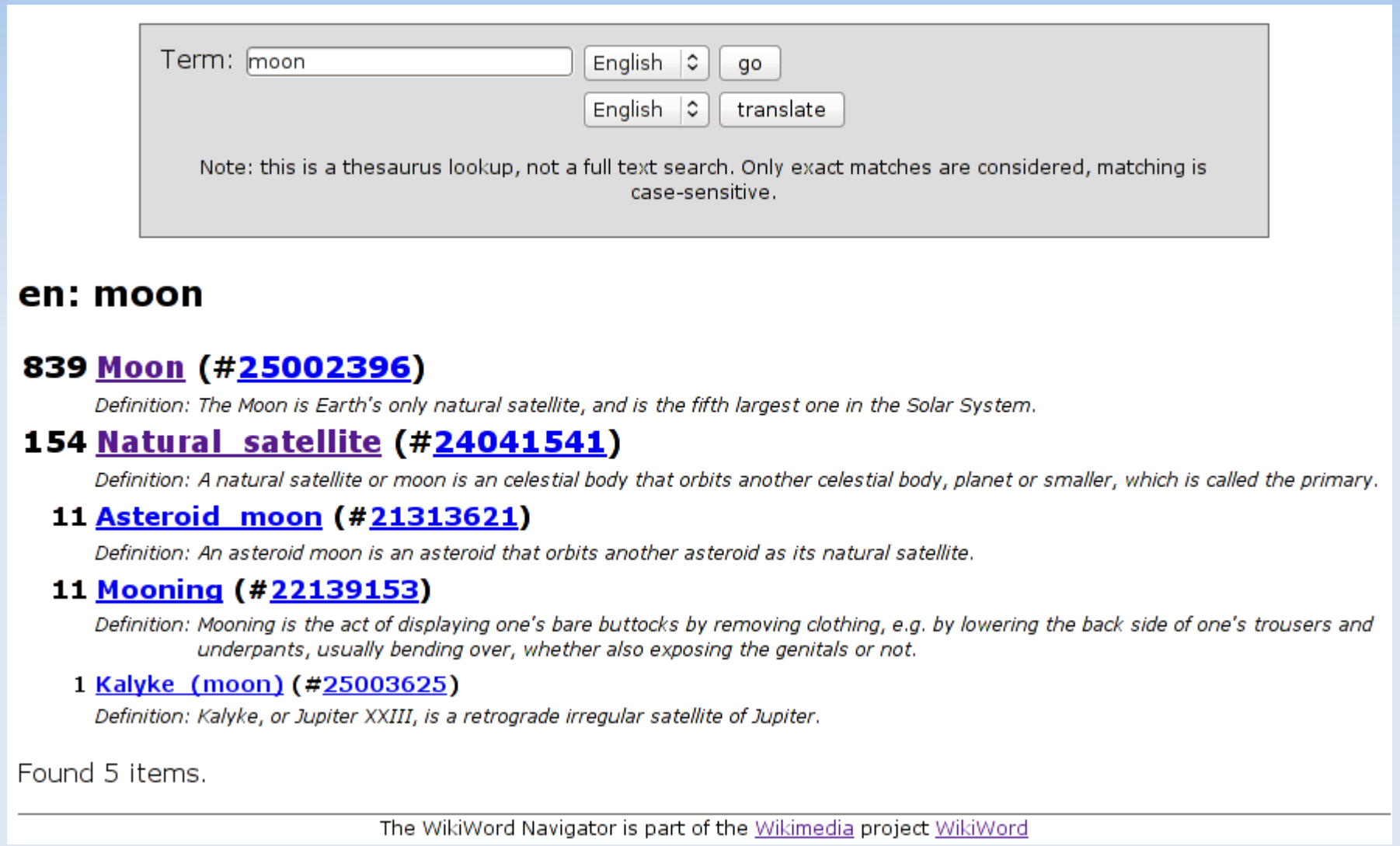
- Term → Concepts
- Concept → terms, properties, images
- Lookup, navigation, translation

- Commons: pretend...
  - Commons is a language
  - `{{Commons:Foo}}` is a language link

Searching with WikiWord

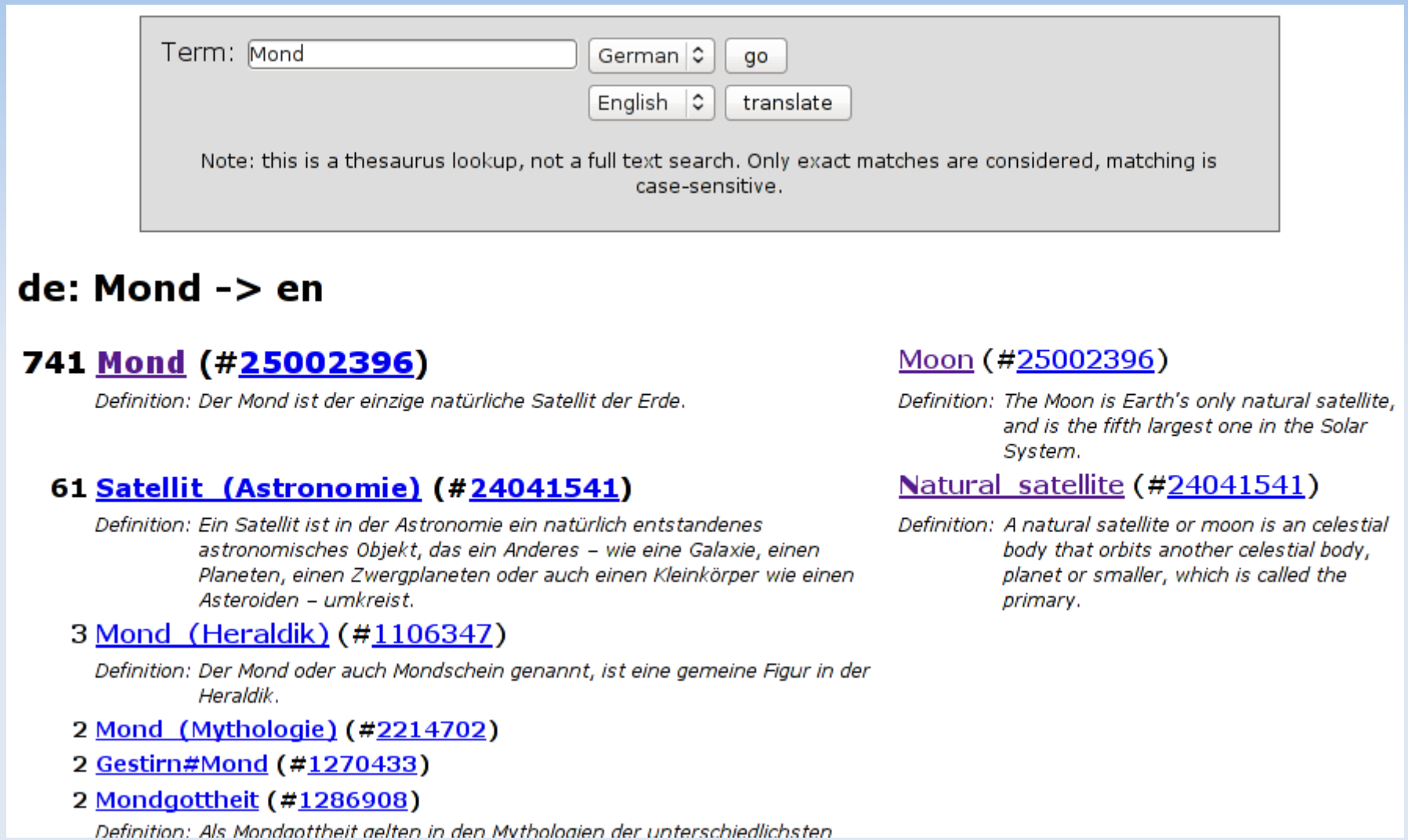
- List concepts for a term in a given language
- Allow navigation to related topics

Search



http://toolserver.org/~daniel/wikiword/wikiword.php

Search



http://toolserver.org/~daniel/wikiword/wikiword.php

Searching with WikiWord

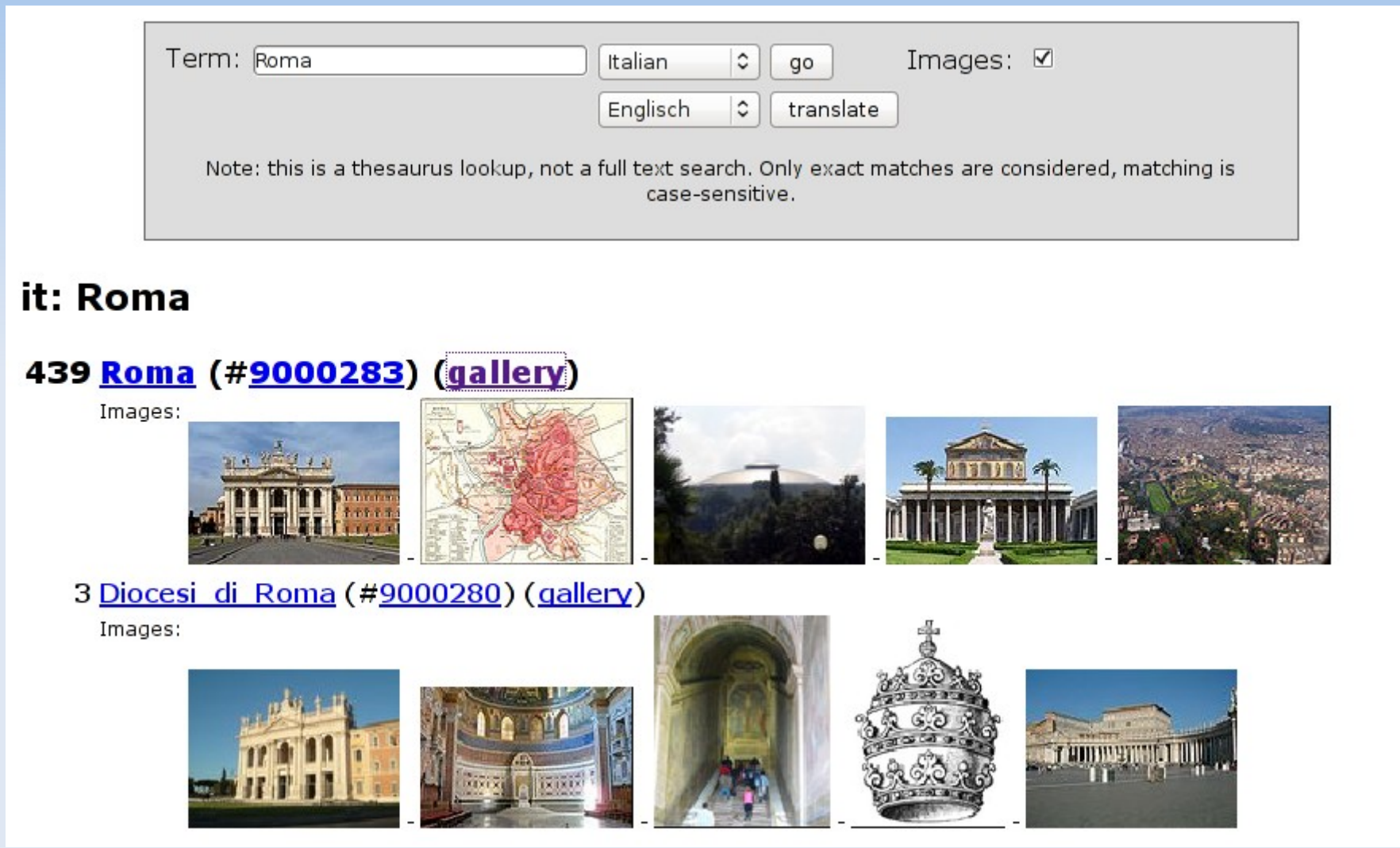
- Not a text seach
- Can't be used directly to categorize
- Still powerful and useful

Image Search with WikiWord

- Images from pages, galleries and categories
- Ranking: *used* images are best
- To do: favor featured/quality/valued images

- proof of concept implementation
  - For each concept, show images
  - Using old data, fake link to commons
  - Alpha-test with new data model: popes

Searching with WikiWord



http://toolserver.org/~daniel/wikiword-trunk/wikiword.php

Statistics

- Original WikiWord thesaurus (2007):
- en, de, fr, nl, no
- 8 million pages
- 12 million concepts (4 million ”blue”)
- 22 million term associations
- 100 million relations between concepts

- New thesaurus: nearly twice the size

Perspective

- Fresh complete thesaurus (really soon now)
  - This should be come a regular service
  - RDF/SKOS
- Usable image search (by the end of the year)
  - Alternative interface to commons

- Running on a toolserver machine
- Tricky because of huge size
- Development is ongoing

Thank You!

https://wiki.toolserver.org/view/WikiWord

http://brightbyte.de/page/WikiWord

http://toolserver.org/~daniel/wikiword/wikiword.php

WIKIMEDIA
DEUTSCHLAND