A Robust Approach to Aligning Heterogeneous Lexical Resources

Mohammad Taher Pilehvar
Roberto Navigli
Lexical Resource

- WordNet
- BabelNet
- FrameNet
- VERBNet
- UBY
- Wiktionary
- Wikipedia
Lexical Resource

- WordNet
- FrameNet
- BabelNet
- UBY
- Wiktionary
  - a multi-lingual tree encyclopedia
  - Wiktionary [ˈwɪkʃənəri] n., a wiki-based Open Content dictionary
  - Wiko [ˈwɪk oʊ]
Why combine resources?

✓ Improved **word** and **concept** coverage
  ➢ e.g., named entities, new senses

✓ Improved **domain** coverage

✓ Improved **multilinguality**
  ➢ dozens of new languages

✓ Expert-made **relations preserved**
  ➢ e.g., Hyponymy, meronymy, etc.
Why combine resources?

Provides complementary knowledge

Applications:

Semantic Parsing
Shi and Mihalcea, 2005

Semantic Role Labeling
Palmer et al., 2010

WSD and entity linking
Moro et al., TACL 2014
Difficulty of resource alignment

Fine granularity of lexical resources

plant

WordNet
4 senses

Wiktionary
15 senses
How resource alignment works?

- Usually measures the similarity of two concepts
How resource alignment works?

- Usually measures the similarity of two concepts
- And aligns two concepts if their similarity exceeds a certain threshold

\[ \text{WKT: plant\#n\#1} \quad ? \quad \text{WN: plant\#n\#1} \]
How resource alignment works?

Alignment approaches **differ** in the way they calculate this similarity.
Denfinitional similarity

Noun

- **S**: (n) plant, **works**, **industrial plant** (buildings for carrying on industrial labor) "they built a large plant to manufacture automobiles"

- **S**: (n) plant, **flora**, **plant life** ((botany) a living organism lacking the power of locomotion)
- **S**: (n) plant (an actor situated in the audience whose acting is rehearsed but seems spontaneous to the audience)
- **S**: (n) plant (something planted secretly for discovery by another) "the police used a plant to trick the thieves"; "he claimed that the evidence against him was a plant"

**plant** *(plural** plants** or **plantae**)*

1. An **organism** that is not an animal, especially an organism capable of photosynthesis. Typically a small or herbaceous organism of this kind, rather than a **tree**. *(quotations ▼)*

   *The garden had a couple of trees, and a cluster of colourful plants around the border.*

2. *(botany)* An **organism** of the kingdom Plantae; now specifically, a living organism of the Embryophyta (land plants) or of the Chlorophyta (green algae), a **eukaryote** that includes double-membraned **chloroplasts** in its cells containing **chlorophyll** a and b, or any organism closely related to such an organism.

3. *(ecology)* Now specifically, a multicellular **eukaryote** that includes **chloroplasts** in its **cells**, which have a cell wall.
Denfinitional similarity

• Strong baseline

• Fall short when
  • Different wordings are used for same concepts
  • When two words lack quality glosses

plant -- Buildings for carrying on industrial labor.

plant -- The necessary infrastructure used in support and maintenance of a given facility.
Contributions

A novel concept similarity measure

+ Denfinitional similarity

An effective ontologization approach

A robust technique for alignment of heterogeneous resources
Our approach: SemAlign
Our approach: SemAlign

Definition similarity
Our approach: SemAlign

Structural similarity
SemAlign: structural similarity

WordNet
SemAlign: core

Modeling concepts through Semantic Signatures
Semantic Signature of a concept

Personalized PageRank

some concept
Semantic Signature of a concept
Personalized PageRank

Distributional representation over all concepts in the semantic network
SemAlign
SemAlign: signature unification
SemAlign: signature unification

Find concepts associated with monosemous words
SemAlign: signature unification

Truncate vectors to the overlapping concepts
SemAlign: signature unification

The reliability of leveraging monosemous words

~ 60% (72,000)

WordNet Synsets containing at least one monosemous word
Semantic Signature Comparison

![Diagram of Semantic Signature Comparison](image)
Semantic Signature Comparison

Weighted Overlap
(Pilehvar et al., ACL 2013)

\[ Sim(S_{v1}, S_{v2}) = \frac{\sum_{i=1}^{|T|} (r_i^1 + r_i^2)^{-1}}{\sum_{i=1}^{|T|} (2i)^{-1}} \]
SemAlign: score combination
Ontologization of lexical resources
Ontologization of lexical resources

WordNet

Wiktionary

OmegaWiki
Ontologization of lexical resources

WordNet

Wiktionary

Wikipedia

OmegaWiki
A windmill is a machine that converts the energy of wind into rotational energy by means of vanes called sails or blades.[1][2] The reason for the name "windmill" is that the devices originally were developed for milling grain for food production; the name stuck when in the course of history, windmill machinery was adapted to supply power for many industrial and agricultural needs other than milling.[3] The majority of modern windmills take the form of wind turbines used to generate electricity, or windpumps used to pump water, either for land drainage or to extract groundwater.
Ontologization of lexical resources

Windmill sail

From Wikipedia, the free encyclopedia

Windmills are powered by their sails. Sails are found in different designs, from primitive common sails to the advanced patent sails.

Contents [hide]
1 Jib sails
2 Common sails
3 Spring sails
4 Roller reefing sails
5 Patent sails
6 Spring patent sails
7 Dutch sail types
8 Berton sails
9 Annular sails
10 Notes
11 References

Jib sails [edit]

The jib sail is found in Mediterranean countries, and consists of a simple triangle of cloth wound round a spar. The mill must be stopped in order to adjust the reefing of the sail. Though rare in the UK, at least two windmills are known to have had jib sails (St Mary's, Isle of Scilly and Cann Mills, Melbury Abbas).
Ontologization of lexical resources

WordNet

Wiktionary ['wɪkʃənri] n., a wiki-based Open Content dictionary

Wikipedia

OmegaWiki
Ontologization of lexical resources

WordNet

Wiktionary

μιλεμενή λέξη

Wikipedia

The Free Encyclopedia

OmegaWiki
Ontologization of lexical resources

windmill

1. A machine
2. The
3. A child’s
4. (base)
5. (jugg)

motion by means of adjustable vanes called sails.

State when blown by a person or by the wind.

Cicular motion before throwing the ball through the hoop.

windmill.
Definition page for sail

1. *(nautical)* A piece of *fabric* attached to a *boat* and arranged such that it causes the wind to drive the boat along. The sail may be attached to the boat via a combination of *mast*, *spars* and *ropes*. [quotations ▼]
2. *(uncountable)* The power harnessed by a sail or sails, or the use this power for travel or transport.
3. A trip in a boat, especially a *sailboat*.  
   *Let's go for a sail.*
4. *(dated)* A sailing vessel; a vessel of any kind; a craft. Plural *sail*.  
   *Twenty sail were in sight.*
5. The *blade* of a *windmill*.
6. A tower-like structure found on the dorsal (topside) surface of *submarines*.
7. The floating organ of *siphonophores*, such as the *Portuguese man-of-war*.
8. *(fishing)* A *sailfish*.
   *We caught three sails today.*
9. *(paleontology)* an outward projection of the *spine*, occurring in certain *dinosaurs* and *synapsids*
10. Anything resembling a sail, such as a *wing*. [quotations ▼]
Ontologization of lexical resources
Ontologization of lexical resources

WordNet

Wikipedia

Wiktionary

OmegaWiki
# Ontologization of lexical resources

## windmill

**windmill**: A machine for grinding or pumping driven by a set of adjustable vanes or sails

### Lexical annotations

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>word class</td>
<td>substantive</td>
</tr>
</tbody>
</table>

### Definition

<table>
<thead>
<tr>
<th>Language</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulgarian</td>
<td>Машина за смилане или изпомпване, задвижвана от система от приспособими перки и крила, въртящи от силата на вятъра.</td>
</tr>
<tr>
<td>Castilian</td>
<td>Ма́quina para moler o bombar cuya energía proviene de un sistema de paletas o de velas ajustables que son impulsadas por la fuerza del viento.</td>
</tr>
<tr>
<td>English</td>
<td>A machine for grinding or pumping driven by a set of adjustable vanes or sails that are caused to turn by the force of the wind.</td>
</tr>
</tbody>
</table>
Ontologization of lexical resources

WordNet

Wikipedia

? a multilingual tree encyclopedia

Wiktionary [ˈwɪkʃənri] n., a wiki-based Open Content dictionary

OmegaWiki
Ontologization of lexical resources

Definition page for *windmill*

1. A machine which translates linear motion of wind to rotational motion by means of adjustable vanes called sails.
Ontologization of lexical resources

Definition page for *windmill*

1. A **machine** which translates linear motion of **wind** to **rotational motion** by means of **adjustable vanes** called **sails**.
Ontologization: similarity-based disambiguation

Definition page for windmill

1. A machine which translates linear motion of wind to rotational motion by means of adjustable vanes called sails.
Ontologization: similarity-based disambiguation

Definition page for *windmill*

1. A machine which translates linear motion of wind to rotational motion by means of adjustable vanes called **sails**.

Definition page for *sail*

1. A trip in a boat, especially a sailboat.
2. A sailing vessel; a vessel of any kind; a craft.
3. The blade of a windmill.
4. A tower-like structure found on the dorsal (topside) surface of submarines.
Ontologization: similarity-based disambiguation

Definition page for *windmill*

1. A machine which translates linear motion of wind to rotational motion by means of adjustable vanes called sails.

Definition page for *sail*

1. A trip in a boat, especially a sailboat.
2. A sailing vessel; a vessel of any kind; a craft.
3. The blade of a windmill.
4. A tower-like structure found on the dorsal (topside) surface of submarines.
Ontologization: similarity-based disambiguation

Definition page for windmill

1. A machine which translates linear motion of wind to rotational motion by means of adjustable vanes called sails.

Definition page for sail

1. A trip in a boat, especially a sailboat.
2. A sailing vessel; a vessel of any kind; a craft.
3. The blade of a windmill.
4. A tower-like structure found on the dorsal (topside) surface of submarines.
Ontologization: similarity-based disambiguation

Definition page for windmill

1. A machine which translates linear motion of wind to rotational motion by means of adjustable vanes called \textbf{sails}.

Definition page for sail

1. A trip in a boat, especially a sailboat.
2. A sailing vessel; a vessel of any kind; a craft.
3. The blade of a windmill.
4. A tower-like structure found on the dorsal (topside) surface of submarines.
Ontologization: evaluation

![Bar graph showing comparison between Our method and WKTWSD (Meyer & Gurevych, 2012)].

- Our method: Acc = 0.9, F1 = 0.8
- WKTWSD: Acc = 0.9, F1 = 0.8
Alignment Experiments: Datasets
(Matuschek & Gurevych, TACL 2013)

WN synsets manually mapped to their corresponding concepts

WordNet

320

484

315
Alignment Experiments:
System configurations

Parameter 1:
score combination parameter $\beta$

$$\beta \ Sim_{def}(S_{v1}, S_{v2}) + (1 - \beta) \ Sim_{str}(S_{v1}, S_{v2})$$
Alignment Experiments:
System configurations

Parameter 2:
similarity threshold $\theta$
Alignment Experiments: System configurations

Setting $\theta$ and $\beta$

a. Unsupervised system
Alignment Experiments: System configurations

Setting $\theta$ and $\beta$

a. Unsupervised system
b. Tuning on subset
Alignment Experiments: System configurations

Setting $\theta$ and $\beta$

a. Unsupervised system
b. Tuning on subset
c. Cross validation
Comparison system

Dijkstra-WSA

(Matuschek & Gurevych, TACL 2013)
Alignment Experiments

- **WN-WP**
  - SB+DWSA: 0.84
  - SB: 0.805
  - Unsupervised: 0.833

- **WN-WT**
  - SB+DWSA: 0.722
  - SB: 0.712
  - Unsupervised: 0.73

- **WN-OW**
  - SB+DWSA: 0.749
  - SB: 0.733
  - Unsupervised: 0.709

Legend:
- Unsupervised
- Tuning on subset
- Cross-validation
Alignment Experiments

- WN-WP: SB+DWSA = 0.824, SB = 0.824
- WN-WT: Tuning on WN-WT = 0.684
- WN-OW: Tuning on WN-OW = 0.684
SemAlign: structural similarity
Alignment Experiments

<table>
<thead>
<tr>
<th></th>
<th>DWSA</th>
<th>SemAlign</th>
</tr>
</thead>
<tbody>
<tr>
<td>WN-WP</td>
<td>0.71</td>
<td>0.83</td>
</tr>
<tr>
<td>WN-WT</td>
<td>0.39</td>
<td>0.623</td>
</tr>
<tr>
<td>WN-OW</td>
<td>0.473</td>
<td>0.627</td>
</tr>
</tbody>
</table>
Conclusions

• A novel approach for aligning lexical resources
  – Accurate even in the absence of training data
  – Robust across different resources

• An effective ontologization approach

• Experiments on aligning
  – WN to WP, WT, and OW
Future directions

• Integrating the approach into BabelNet for boosting the alignment accuracy

• Alignment across different languages
Thanks!