

# Beyond the Synset: Synonyms in Collaboratively Constructed Semantic Resources

Michael Matuschek and Iryna Gurevych



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

# Outline



- Introduction
  - Wiktionary
  - OmegaWiki
  - Wikipedia
- Explicit encoding of synonyms
- Implicit encoding of synonyms
- Inference of synonymy from context

# Introduction I

- Dictionaries/thesauri are important tools for linguists
- In the past: made by **experts only!**



Source: [www.duden.de](http://www.duden.de), © Bibliographisches Institut GmbH, 2010

# Introduction II

- Now, this paradigm has changed
- People easily collaborate and construct resources on the Web...



- ... and challenge “classical” resources in size and quality!
- So what can we learn about synonyms here?

# Wiktionary



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- A free, Wiki-based online dictionary
- Over 300,000 English entries right now
  - Also available in various other languages
  - German, French and others with > 50,000 entries
- Over 400,000 users, **over 20,000** actively editing
- Users can freely add and edit...
  - Word senses and definitions
  - Etymology
  - Pronunciation
  - Lexical relations
    - **Synonymy**, antonymy, hyponymy, hypernymy
  - ...
  - **Problem:** There are guidelines and templates, but these are not followed consistently

a multilingual tree  
encyclopedia

## Wiktionary

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

Wilek [ˈwɪl kɔʔ]

- A free, **multilingual** dictionary
- Over 420.000 expressions in 255 languages
- Over 40.000 **language-independent** concepts
- Around 3,000 users
- Goals
  - Overcome Wiktionary's structural inconsistencies
  - Create a resource for translations/synonyms which is easily accessible and maintainable
- **Consequence:** a fixed database schema
  - Users can only contribute if they stick to the predefined structure
    - ...but the price is a loss in expressiveness



# Wikipedia



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- A free, multilingual encyclopedia
- Over 3.000.000 English articles right now
  - German, French and many others with > 500,000
- Over 13,000,000 users, **over 130,000 active**
- Each article describes a distinct concept
- Goal
  - A collaboratively created source of **encyclopedic** knowledge...
  - ...NOT **linguistic** knowledge
- However, we can mine this knowledge on various levels:
  - Links
  - Article history
  - Simple Wikipedia



**WIKIPEDIA**  
*The Free Encyclopedia*



# PART 1

## Explicit encoding of synonyms

# Wiktionary



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Over 400,000 English lexemes
  - Almost twice as much as WordNet
- Only about 20,000 (unidirectional) synonymy relations for English
  - WordNet has over 1,000,000!
- But: Other languages have far more
  - German Wiktionary: 50,000 lexemes, almost 40,000 synonym links
    - Cf. **[Meyer & Gurevych 2010a]**
  - Reason is unclear

a multilingual tree  
encyclopedia

**Wiktionary**

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

וִיקִיוֹנָרִי

# Wiktionary Characteristics



- Each user is free to add/edit synonyms
- A template is provided, but **not mandatory**
  - Consequence: inconsistencies => hard to use

====Synonyms====

- \* `{{sense|private vehicle that moves independently}}` `[[auto]]`, `[[motorcar]]`, `[[vehicle]]`; `[[automobile]]` `{{qualifier|US}}`, `[[motor]]` `{{qualifier|British colloquial}}`, `[[carriage]]` `{{qualifier|obsolete}}`
- \* `{{sense|non-powered part of a train}}` `[[railcar]]`, `[[wagon]]`
- \* `{{sense|unit of quantity}}` `[[carload]]`, `[[wagonload]]`
- \* `{{sense|passenger-carrying light rail unit}}` `[[carriage]]`
- \* `{{sense|part of an airship}}` `[[gondola]]`, `[[basket]]` `{{qualifier|balloons only}}`
- \* See also `[[Wikisaurus:automobile]]`

a multilingual tree  
encyclopedia

**Wiktionary**  
[ˈwɪkʃənɹɪ] *n.*,  
a wiki-based Open  
Content dictionary  
Wilek ɪˈwɪɪ kəɪɪ

# Example



Template is also not perfect:

- Links are unidirectional (no synsets a la WordNet)
- Synonyms are not directly attached to senses (only via the gloss)

## Synonyms

- (*private vehicle that moves independently*): [auto](#), [motorcar](#), [ve](#)  
*colloquial*), [carriage](#) (*obsolete*)
- (*non-powered part of a train*): [railcar](#), [wagon](#)
- (*unit of quantity*): [carload](#), [wagonload](#)
- (*passenger-carrying light rail unit*): [carriage](#)
- (*part of an airship*): [gondola](#), [basket](#) (*balloons only*)
- See also [Wikisaurus:automobile](#)

a multilingual tree  
encyclopedia

**Wiktionary**

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

Wileon Γ'wɪl kəɪl

# Example



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Synonyms for “flexible”:

## Synonyms

- ductile
- inconstant
- manageable

- obsequious
- pliant
- pliable

- supple
- tractable
- wavering

Deviation from the standard!

a multilingual tree  
encyclopedia

**Wiktionary**

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

Wilek ʾwɪʾl kəʾɪʾl

# Ambiguities in Wiktionary links



Synonym links lead to whole entries, not word senses

<b>boat</b>	<b>craft</b> ( <i>countable and uncountable; plural craft or crafts</i> )
<b>English</b>	
Most common English words	
<b>Pronunciation</b>	
<ul style="list-style-type: none"><li>(RP) enPR: bōt, IPA: /bəʊt/</li><li>(GenAm) enPR: bōt, IPA: /t/</li><li>Audio (US)<sup>help</sup>.file</li><li>Rhymes: -əʊt</li></ul>	
<b>Etymology</b>	
From Old English <i>bāt</i> < Proto-Germanic <i>bātr</i> , <i>beit</i> (Icelandic: <i>bátur</i> ), <i>Re</i>	
<b>Noun</b>	
boat (plural <b>boats</b> )	
<ul style="list-style-type: none"><li>[1] A craft used for transport propelled by oars or another</li><li>[2] (<i>poker slang</i>) A full house</li><li>[3] (<i>chemistry</i>) One of two</li></ul>	
<b>Synonyms</b>	
[1] craft, ship, vessel	
<b>Hyponyms</b>	
<ul style="list-style-type: none"><li>[1] ark, bangca, barge, canoe, catamaran, caravel, carrack, coracle, cruiser, cutter, dhow, dinghy, dory, dragon boat, Dutch barge, East Indiaman, felucca, ferry, ferryboat, fishing boat, folding boat, galley, galleon, gig, go-fast boat, houseboat, hovercraft, hydrofoil, hydroplane, inflatable boat, inflatable raft, jetboat, jetski, junk, kayaaki,</li></ul>	

[1] The skilled practice of a practical occupation.  
*She represented the craft of brewers.*

[2] (*nautical, whaling*) Implements used in catching fish, such as net, line as in harpoons, hand-lances, etc.

[3] (*nautical*) Boats, especially of smaller size than ships. Historically for loading or unloading of other vessels, as lighters, hoys, and barges.

[4] (*nautical, British Royal Navy*) Those vessels attendant on a fleet, such as gunboats, generally commanded by lieutenants.

[5] A vehicle designed for navigation in or on water or air or through outer space.

[6] A particular kind of skilled work.  
*He learned his craft as an apprentice.*

[7] Shrewdness as demonstrated by being skilled in deception.

Cf. [Meyer & Gurevych, 2010b]

a multilingual tree  
encyclopedia

**Wiktionary**

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

וויקישאָנאָר



- Word senses: 45,000
- Expressions: 420,000
  - Around 40,000 in English
- Average no. of translations for a sense: 10.73
- **But where are the synonyms?**
  - Encoded as translations within the same language!
  - Senses with at least two English synonyms: 6051
  - Average no. of synonyms for these: 2.74
- Numbers for German and French are comparable
- Not very large
  - But the structure is worth a look!



# OmegaWiki entry for “car”



Language	Spelling	Identical meaning?	Annotation
Basque	automobil	<input checked="" type="checkbox"/>	
Basque	auto	<input checked="" type="checkbox"/>	
Bokmål	bil	<input checked="" type="checkbox"/>	
Breton	gwetur	<input checked="" type="checkbox"/>	
Breton	tangarr	<input checked="" type="checkbox"/>	
Breton	karr-tan	<input checked="" type="checkbox"/>	
Breton	karr	<input checked="" type="checkbox"/>	

English	car	<input checked="" type="checkbox"/>
English	automobile	<input checked="" type="checkbox"/>

Castilian	coche	<input checked="" type="checkbox"/>	<a href="#">Annotation »</a>
Czech	automobil	<input checked="" type="checkbox"/>	
Danish	bil	<input checked="" type="checkbox"/>	
Dutch	wagen	<input checked="" type="checkbox"/>	
Dutch	auto	<input checked="" type="checkbox"/>	
Dutch	kar	<input checked="" type="checkbox"/>	
Dutch	automobiel	<input checked="" type="checkbox"/>	
English	car	<input checked="" type="checkbox"/>	<a href="#">Annotation «</a>
English	automobile	<input checked="" type="checkbox"/>	
Esperanto	aŭto	<input checked="" type="checkbox"/>	

Property  Value  
part of speech noun



# OmegaWiki Characteristics



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- A fixed, WordNet-like structure (“synsets”)
  - Volunteers only => has to be **enforced** by a DB structure!
- Pro: Easy maintenance and access
  - Simple SQL is enough
- Con: Less flexibility
  - Central idea: **language-independent** concepts
  - All translations and synonyms are treated equally
    - “Identical meaning” can be unchecked, but this is hardly ever done
- In fact, this encodes **absolute synonymy**
  - Rarely seen in real life!
  - But might still be good enough for users...



# Example for Problematic Senses

- Are these words really identical in meaning?

Language ▲	Spelling	Identical meaning?
Italian	aggraziato	<input checked="" type="checkbox"/>
German	anmutig	<input checked="" type="checkbox"/>
French	gracieux	<input checked="" type="checkbox"/>
English	comely	<input checked="" type="checkbox"/>
English	good-looking	<input checked="" type="checkbox"/>
English	beautiful	<input checked="" type="checkbox"/>
English	pretty	<input checked="" type="checkbox"/>
English	handsome	<input checked="" type="checkbox"/>



# Cross-lingual Synonymy

- We could also study **cross-lingual synonymy**
  - That's what OmegaWiki was made for anyway!
- But: The low expressive power leads to even more problems here
- Translations are rarely unambiguous/obvious
  - => Word sense disambiguation, only worse
- Cf. [**Sinha et al., 2010**]

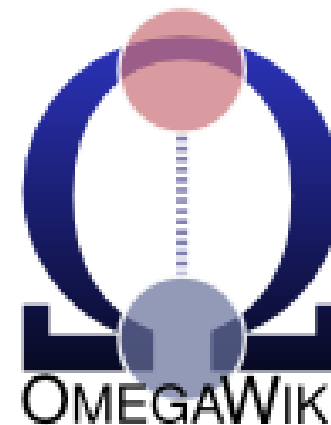


# Wiktionary & OmegaWiki: Interesting Aspects

If they are flawed, why even bother using them?

**Because of their fundamental idea!**

- Edited by “regular” people, not experts
- Continuous validation by the crowd
- If a link persists, users seem to be ok with it
  - The synonymy is perceived as valid...  
=> **cognitive synonymy**
  - ...by a large community, not just some experts  
=> **“collective cognitive synonymy”**
  - Cf. [Cruse 1986]



a multilingual tree  
encyclopedia

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

וִיקִיִּוֹנָרִי [ˈwɪkʃənri]

# Collective Cognitive Synonymy

## Conclusion:

- This could be a gold mine for research!
- We see the “**people’s choice**” of synonyms
- There are a lot of questions to ask:
  - Why are two entries linked here, but not in other resources?
  - What “synsets” have been created by the crowd?
  - Why are some links in Wiktionary unidirectional?
  - What synsets exist across languages in OmegaWiki?
  - ...



# PART 2

## Implicit encoding of synonyms

# Wiktionary I



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Synonyms can be hidden in a gloss

**pretty** (*comparative prettier, superlative prettiest*)

1. **Cunning; clever, skilful.** [from 9th c.] [quotations ▼]
2. **Pleasant** in sight or other senses; **attractive**, especially of women or children.  
c.] [quotations ▼]
3. Of objects or things: nice-looking, **appealing.** [from 15th c.] [quotations ▼]

- Mining these is probably not trivial!
  - Antonyms etc. would also be ok here
- We haven't tried yet, though 😊

a multilingual tree  
encyclopedia

**Wiktionary**

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

וִיקִי־אֵנְצִיקְלוֹפֵדִיָּה

# Wiktionary II



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Synonyms could also be inferred by the link structure
  1. Add backwards direction for unidirectional links (symmetry!)
  2. Calculate the transitive closure⇒ Equivalence classes  
⇒ **Pseudo synsets**
- “The synonym of my synonym is also my synonym”
  - Or is it?
  - Links are set manually => there might be a reason why they're missing!

a multilingual tree  
encyclopedia

**Wiktionary**

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

וִיקִי־לֵשׁוֹן

# Interesting Examples



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Synonym list for the verb die:

## Synonyms

[edit]

- *(to stop living)*: bite the dust, buy the farm, check out, cross over, expire, succumb, give up the ghost, pass, pass away, pass on, be no more, cease to be, go to meet one's maker, be a stiff, push up the daisies, hop off the twig, kick the bucket, shuffle off this mortal coil, join the choir invisible

- These would all be linked!
- Not all of them are synonyms...
  - ...but common traits seem to be enough!
- Using these might be hard
  - Minor traits are lost
- But: We get to see **which traits matter** in the collective mind

a multilingual tree  
encyclopedia

**Wiktionary**

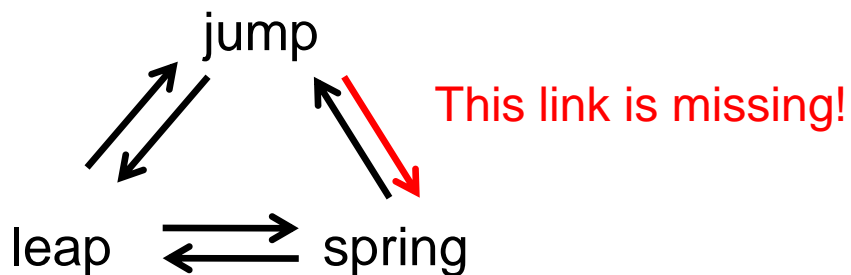
['wɪkʃənri] *n.*,  
a wiki-based Open  
Content dictionary

Wilek ɪ'wɪl kəʀɪl

# More Examples



- Same holds for “link chains”:
  - liberty → freedom → exemption → dispense
  - Again, the capital trait is the same
    - Minor traits are not
- But remember: These are **directed links**
  - The other direction might be missing for a reason!
  - What traits are crucial for **denying** synonymy?
- Careful, however: Sometimes it’s just a mistake after all:



a multilingual tree  
encyclopedia

**Wiktionary**

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

Wileon Γ'ωτλ κερτλ

# More elaborate approaches



- Extended idea:
  - Add links between neighbors in the graph ("clusters")
  - Use translation links
    - Shared translations => Synonyms!
  - Discussed by **[Navarro et al., 2009]**
- Result:
  - Better coverage...
  - ...but a sharp drop in precision
- Consequence:
  - Don't link automatically, it's not reliable enough!
  - Better: Just make suggestions to users
  - Browser plug-in discussed in **[Sajous et al., 2010]**

a multilingual tree  
encyclopedia

**Wiktionary**

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

Wilek ɪˈwɪɪ kəɪɪ

# Wikipedia Links



- Another example: Wikipedia link anchors
- Each article stands for a concept
  - => a link anchor leading there might just be a different label for it!
- Moreover: links are embedded in context
  - => **propositional synonymy**

In recent years, the paid program or **infomercial** has become common. Some **drug companies** and other businesses have even entered the industry as **video news releases**, paying program directors to use **product placement**. For example, a character could be drinking a certain

Pharmaceutical company



# Wikipedia Redirects



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Redirects link different terms to the same concept
  - Goal: avoid redundancy
- Different names for the same thing?
  - Smells like synonyms!

## Neologism

From Wikipedia, the free encyclopedia

(Redirected from [Coined word](#))



WIKIPEDIA  
*The Free Encyclopedia*

# What's in the link structure



- Yes...
  - *coined term, coined word, new word, word coinage, coining*
- ...but there's also a whole lot more!
  - Spelling variations/errors:
    - *neo-logism, neoligism, neolism*
  - Related/derived terms:
    - *neologist, neologistic*
  - And others 😊:
    - *Liberty Cabbage*
- **Conclusion:**
  - Lots of propositional synonyms here!
  - (New) labels for concepts, emerging from the collective mind
  - But, a lot of cleaning needs to be done first
    - Cf. **[Nakayama et al., 2008]**





# PART 3

## Inference of synonymy from context

# What does “context” mean here?



- So far: Exploiting the resource structure to mine/examine synonyms
- But especially Wikipedia gives us even more!
  1. Wikipedia Revision History
    - Changes for all articles are saved, not only the current version
    - A “look over the shoulder” of authors
  2. Simple English Wikipedia
    - A Wikipedia for non-native English speakers, children etc.
    - Not as big as the regular one (~ 60,000 articles)...
    - ...but a “case study” for the use of language!



# Wikipedia Example



WIKIPEDIA  
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article

Interaction

- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia
- Donate to Wikipedia
- Help

Article **Discussion**

Read **View source** | **View history**

Search

## Japan

From Wikipedia, the free encyclopedia

*This article is about the country. For other uses, see [Japan \(disambiguation\)](#).*

**Japan** (日本 *Nihon* or *Nippon*), officially the **State of Japan** (日本国 ◀ *Nippon-koku* or *Nihon-koku*), is an [island nation](#) in [East Asia](#).<sup>[9]</sup> Located in the [Pacific Ocean](#), it lies to the east of the [Sea of Japan](#), [China](#), [North Korea](#), [South Korea](#) and [Russia](#), stretching from the [Sea of Okhotsk](#) in the north to the [East China Sea](#) and [Taiwan](#) in the south. The [characters](#) that make up [Japan's name](#) mean "sun-origin" (because it lies to the east of nearby countries), which is why Japan is sometimes referred to as the "[Land of the Rising Sun](#)".

Japan is an [archipelago](#) of 6,852 islands.<sup>[10]</sup> The four largest islands are [Honshū](#), [Hokkaidō](#), [Kyūshū](#) and [Shikoku](#), together accounting for 97% of Japan's land area. Most of the islands are [mountainous](#), many [volcanic](#); for example, Japan's highest peak, [Mount Fuji](#), is a volcano. Japan has the world's [tenth-largest population](#), with over 127 million people. The [Greater Tokyo Area](#), which includes the *de facto* [capital city](#) of [Tokyo](#) and several surrounding [prefectures](#), is the [largest metropolitan area](#) in the world, with over 30 million residents

**State of Japan**  
**Shinjitai:** 日本国  
**Kyujitai:** 日本國  
***Nippon-koku* or *Nihon-koku***

	
Flag	Imperial Seal

**Anthem:**



“Most of the islands are mountainous, many volcanic.”



# Simple Wikipedia Example



Simple English  
WIKIPEDIA

[Main Page](#)  
[Simple start](#)  
[Simple talk](#)  
[New changes](#)  
[Show any page](#)  
[Help](#)  
[Give to Wikipedia](#)

▼ [Print/export](#)  
[Make a book](#)  
[Download as PDF](#)  
[Page for printing](#)  
► [Toolbox](#)

[New features](#) [Log in / create account](#)

Page [Talk](#)

Read

[Change](#)

[View history](#)



## Japan

From Wikipedia, the free encyclopedia

**Japan** (日本) is a [country](#) in [Asia](#). It has many [islands](#). Four of them are big, and the biggest is one of the largest islands in the [world](#). The islands are on the east Asian coast in the western part of the [Pacific Ocean](#).<sup>[1]</sup>

Japanese [people](#) call their country "Nihon" or "Nippon",<sup>[1]</sup> which in Japanese means "the origin of the sun".

### Contents [hide]

- [1 History](#)
- [2 Geography](#)
  - [2.1 Territorial problem](#)
- [3 Transportation](#)
  - [3.1 Other pages](#)
- [4 Subdivisons](#)

### Japan



Official flag



Coat of Arms

### National information

National motto: n/a

National anthem: Kimigayo (君が代)

### About the people

“In the middle of Japan there are mountains. Most of the mountains are volcanoes.”



WIKIPEDIA  
The Free Encyclopedia

# Wikipedia Revision History Example



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Revision as of 03:43, 7 October 2003

[\(view source\)](#)

[TakuyaMurata](#) ([talk](#) | [contribs](#))

m *(add Seto Inland Sea)*

[← Previous edit](#)

Revision as of 04:02, 10 October

2003 [\(view source\)](#)

[211.5.123.232](#) ([talk](#))

*(source of sun > origin of sun (source indicates a stream of something not the origin point))*

[Next edit →](#)

Line 15:

: "Alternate meaning: [\[\[Japan \(band\)\]\]](#)"

"Japan" ("Nippon/Nihon"

&#26085;&#26412;, literal meaning:

"**Source** of Sun") is a country in [\[\[Far East Asia\]\]](#), made up of a chain of islands -

located between the [\[\[Pacific Ocean\]\]](#) and the [\[\[Sea of Japan\]\]](#), east of the [\[\[Korea\]\]](#)n

Line 15:

: "Alternate meaning: [\[\[Japan \(band\)\]\]](#)"

"Japan" ("Nippon/Nihon"

&#26085;&#26412;, literal meaning: "**Origin**

of Sun") is a country in [\[\[Far East Asia\]\]](#), made up of a chain of islands - located

between the [\[\[Pacific Ocean\]\]](#) and the [\[\[Sea of Japan\]\]](#), east of the [\[\[Korea\]\]](#)n peninsula.



WIKIPEDIA  
The Free Encyclopedia

# Wikipedia Revision Mining



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- Revisions can be mined for synonyms!
  - Done by **[Kulesa, 2008]**
  - See **[Nelken & Yamangil, 2008]** for other uses
- Approach:
  1. Identify sentences with word replacements through revisions
  2. Filter out typos, spam etc.
- Result:
  - Same context, different words (see example before)  
=> **propositional synonymy**
  - Or not?
    - **Why** was the original word replaced?
    - Was it just wrong, did another word fit better?
    - Lots of open research questions here



WIKIPEDIA  
*The Free Encyclopedia*

# Simple Wikipedia Mining I

- Simple Wikipedia can also be mined...
  - ...but it's in fact not as simple!
- Problem:
  - Sentences in simple/normal articles are not aligned naturally (as revisions are)
  - But: Alignment is crucial for the mining task!
  - Automatic alignment is error-prone
  - Some things just don't match 😊
  - Cf. [Zhu et al. 2010]
- **Example:**
  - “Japan (日本 Nihon or Nippon), officially the State of Japan (日本国 Nippon-koku or Nihon-koku), is an island **nation** in East Asia.”
  - “Japan (日本) is a **country** in Asia.”
- These have been written independently (most likely)
  - How to match them?



# Simple Wikipedia Mining II



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- **Result:**
  - Less, and noisier data
  - But it might be valuable anyway
- **Hypothesis:**
  - When simplifying, people try to keep the original meaning
  - Higher probability of synonyms?
  - In any case, new insights about the usage of words



WIKIPEDIA  
*The Free Encyclopedia*

# Wikipedia Wrap-Up



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

## Conclusion:

- Wikipedia allows mining/analysis of synonyms in several ways
- We could gather knowledge about use of language like never before
- But: research is still in its infancy
  - Alignment, filtering etc. still need a lot of work
  - ...but it might be well worth it!



WIKIPEDIA  
*The Free Encyclopedia*

# Conclusions

- Collaboratively constructed resources offer abundant possibilities!
  - For mining synonyms themselves
  - For analyzing their use in language
- Never before so many people helped building linguistic resources
  - **“The wisdom of the crowds”** in action
- Yes, the data is noisy, incomplete and otherwise flawed
  - But it’s a peek into the use and perception of language, **real-time**
  - ...which is probably more exciting than analyzing the same old resources
- See [**Zesch et al. 2008**], [**Wolf & Gurevych 2010**] for more background

# Acknowledgements

DFG



Bundesministerium  
für Wirtschaft  
und Technologie



VolkswagenStiftung



# References

- Cruse, D.A. 1986: *Lexical Semantics. Cambridge Textbooks in Linguistics, Cambridge University Press.*
- Kulesa, S. 2008: *Mining Wikipedia's Revision History for Paraphrase Extraction (Master Thesis). TU Darmstadt.*
- Meyer, C., & Gurevych, I. 2010: How Web Communities Analyze Human Language: Word Senses in Wiktionary. *Proceedings of the Second Web Science Conference*, April 2010.
- Meyer, C., & Gurevych, I. 2010: Worth its Weight in Gold or Yet Another Resource — A Comparative Study of Wiktionary, OpenThesaurus and GermaNet. *Proceedings of the 11th International Conference on Intelligent Text Processing and Computational Linguistics*, (pp. 38-49). Iași, Romania.
- Nakayama, K., Pei, M., Erdmann, M., Ito, M., Shirakawa, M., Hara, T., 2008: Wikipedia Mining: Wikipedia as a Corpus for Knowledge Extraction. *Proceedings of Annual Wikipedia Conference (Wikimania).*
- Navarro, E., Sajous, F., Gaume, B., Prévot, L., ShuKai, H., Tzu-Yi, K., Magistry, P., Chu-Ren, H., 2009. Wiktionary and NLP: improving synonymy networks. In *Proceedings of the 2009 Workshop on the People's Web Meets Nlp: Collaboratively Constructed Semantic Resources* (Suntec, Singapore, August 07 - 07, 2009).
- Sajous, F., Navarro, E., Gaume, B., Prévot, L., Chudy, Y., 2010: Semi-automatic Endogenous Enrichment of Collaboratively Constructed Lexical Resources: Piggybacking onto Wiktionary. In: *Advances in Natural Language Processing, Lecture Notes in Computer Science, 2010, Volume 6233/2010, 332-344*
- Nelken, R., & Yamangil, E. 2008: Mining Wikipedia's Article Revision History for Training Computational Linguistics Algorithms. *Proceedings of the Wikipedia and AI Workshop at the AAAI Conference. Chicago, USA.*
- Sinha, R., McCarthy, D., Mihalcea, R. 2009: SemEval-2010 Task 2: Cross-Lingual Lexical Substitution, In *Proceedings of the NAACL-HLT 2009 Workshop: SEW-2009 - Semantic Evaluations, Denver, Colorado*
- Wolf, E., & Gurevych, I. 2010: Expert-Built and Collaboratively Constructed Lexical Semantic Resources for Natural Language Processing. *Language and Linguistics Compass. (to appear)*
- Zesch, T., Müller, C. & Gurevych, I. 2008: Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. *Proceedings of the 6th International Conference on Language Resources and Evaluation. Marrakech, Morocco.*
- Zhu, Z., Bernhard, D. & Gurevych, I. 2010: A Monolingual Tree-based Translation Model for Sentence Simplification. *Proceedings of The 23rd International Conference on Computational Linguistics, Bei Jing, China. (to appear)*