

Wiktionary

The Metalexicographic and the Natural Language Processing Perspective

Christian M. Meyer



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Dissertation talk

Technische Universität Darmstadt, 28. October 2013.

Entry [Discussion](#) [Citations](#)

Read [Edit](#) [History](#)

Search



boat

English

[\[edit\]](#)

Etymology

[\[edit\]](#)





From Middle English *boot*, *bot*, *boet*, *boyt* ("boat"), from Old English *bāt* ("boat"), from Proto-Germanic **baltaz*, **baitan* ("boat, small ship"), from Proto-Indo-European **bheid-* ("to break, split"). Cognate with Old Norse *beit* ("boat").

Old Norse *bātr* (whence Icelandic *bátur*, Norwegian *båt*), Dutch *boot*, German *Boot*, and French *bateau* are all ultimately borrowings from the Old English word.

Pronunciation

[\[edit\]](#)

- (RP) enPR: bōt, IPA: /bəʊt/, X-SAMPA: /b@ʊt/
- Rhymes: -əʊt
- (GenAm) enPR: bōt, IPA: /boʊt/, X-SAMPA: /boʊt/
- Rhymes: -oʊt

• Audio (US)  0:00   

Noun

[\[edit\]](#)

boat (*plural* **boats**)

1. A craft used for transportation of goods, fishing, racing, recreational cruising, or military use on or in the water, propelled by oars or outboard motor or inboard motor or by wind.
2. (*poker slang*) A full house.
3. (*chemistry*) One of two possible conformations of cyclohexane rings (the other being *chair*), shaped roughly like a boat.

Usage notes

[\[edit\]](#)

There's no explicit limit, but the word **boat** usually refers to a relatively small watercraft that is generally smaller than a "ship" and



Wikipedia has an article on:
Boat



A **boat** kept on land



Motivation



TECHNISCHE
UNIVERSITÄT
DARMSTADT

New lexicographic paradigm: collaborative lexicography

New lexicographic paradigm: collaborative lexicography

Why care?

- The Web leads to quickly changing business models of dictionary publishers
 - unclear future of the Deutsches Wörterbuch (the “Grimm”)
 - downsizing of Duden incl. editorial staff
- Collaborative encyclopedias are about to replace expert-built ones
 - radical changes to Brockhaus encyclopedia
- Large amount of freely available information for language technology

The Metalexicographic Perspective



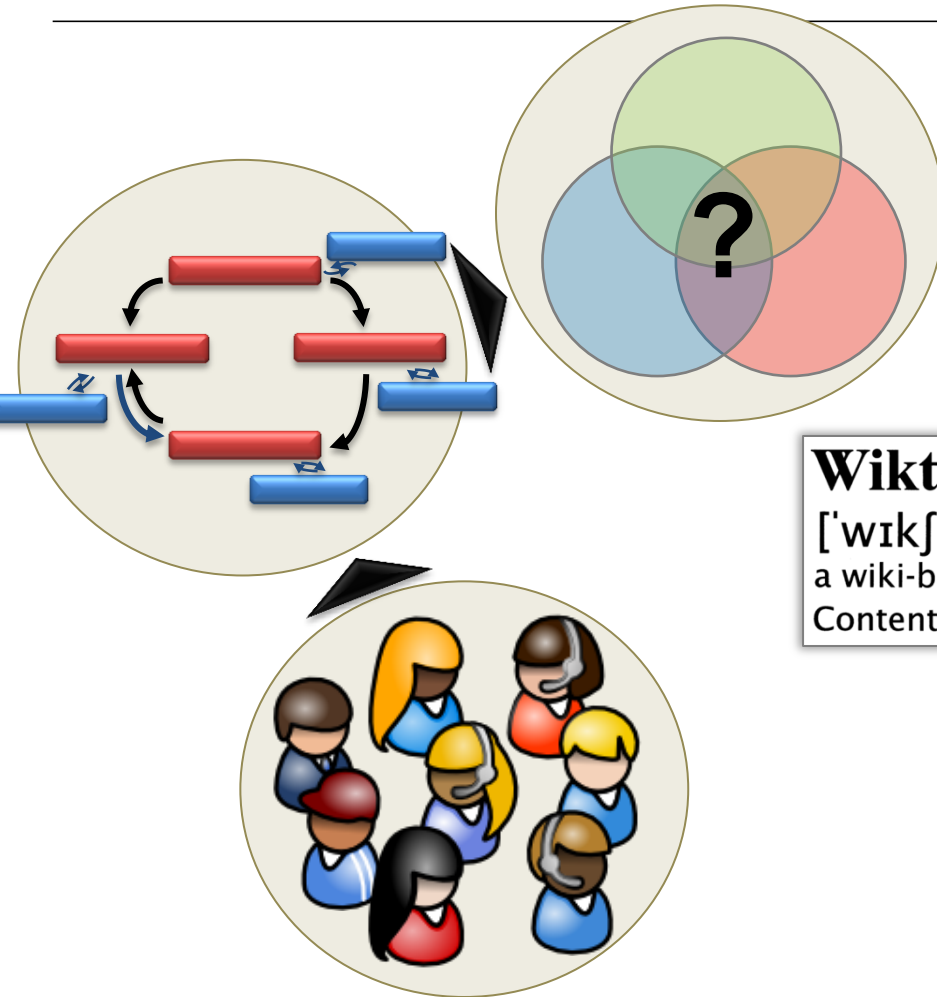
TECHNISCHE
UNIVERSITÄT
DARMSTADT

Research Question:

**Implications of collaborative
construction on lexicography**

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

The Metalexicographic Perspective



Research Question:
Implications of collaborative construction on lexicography

Contributions:

- **Comprehensive description of Wiktionary**
- **Analysis of the collaborative lexicographic process**
- **Quantitative comparison of dictionaries**

The Natural Language Processing Perspective



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Research Question:

**Harvesting linguistic
knowledge from Wiktionary**

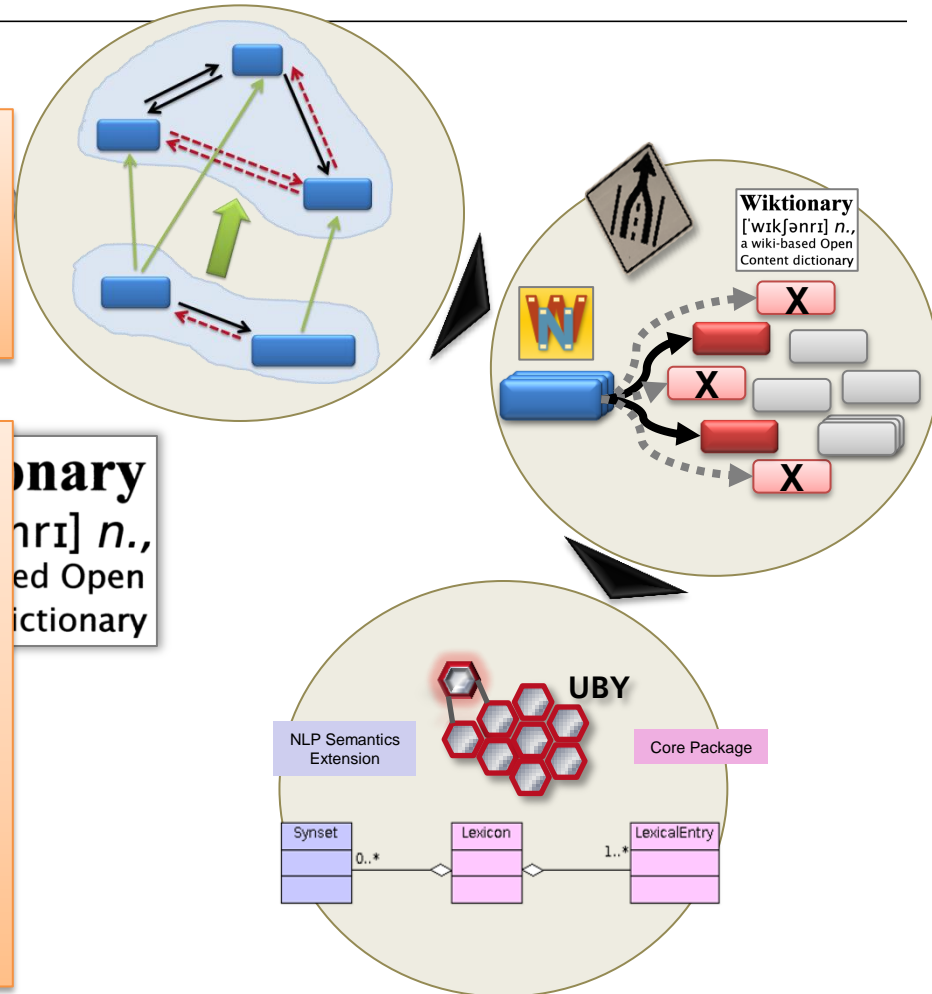
Wiktionary
[ˈwɪkʃənɹɪ] *n.*,
a wiki-based Open
Content dictionary

The Natural Language Processing Perspective

Research Question:
**Harvesting linguistic
knowledge from Wiktionary**

Contributions:

- **Extracting & disambiguating lexical-semantic information**
- **Alignment of Wiktionary and WordNet at sense level**
- **Standardized representation based on LMF**



Research Goals

Metalexicographic Perspective

- Dictionary development
- New theoretical insights

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

Natural Language Processing Perspective

- Resource development
- Improved NLP systems

Interdisciplinary Relationships

Potential of collaboratively
created dictionaries

Metalexicographic
Perspective

- Dictionary development
- New theoretical insights

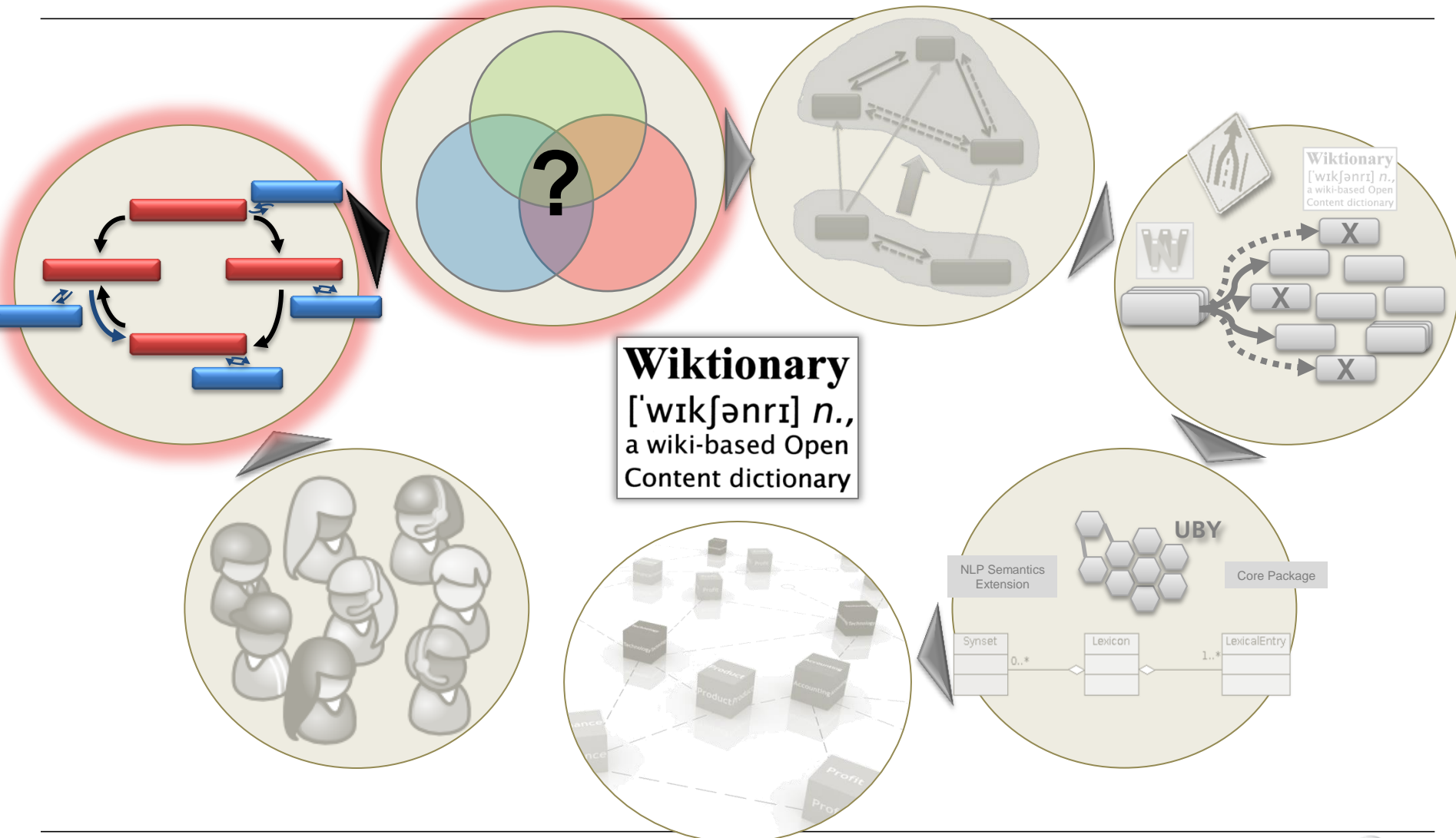
Wiktionary
[ˈwɪkʃənɹɪ] *n.*,
a wiki-based Open
Content dictionary

Natural Language
Processing
Perspective

- Resource development
- Improved NLP systems

Tools for qualitative and
quantitative analysis

Agenda



Dictionary Analysis and Comparison



TECHNISCHE
UNIVERSITÄT
DARMSTADT



<http://www.wordle.net> – Font size represents article count (logarithmic scale)

Dictionary Analysis and Comparison



(diff | hist) . . తీర్థము: 14:51 . . (+109) . . Rajasekhar1961 (Talk | contribs) (→Telugu)

(diff | hist) . . horseshoe; 14:50 . . (+24) . . 115.132.182.28 (Talk) (t+ms:ladam t-balance (Assisted))

(diff | hist) . . **N** армада; 14:50 . . (+132) . . Stephen G. Brown (Talk | contribs) (→Noun)

(diff | hist) . . excite; 14:50 . . (+18) . . Jc86035 (Talk | contribs) (+rhyme -art)

(diff | hist) . . **N** despellejar: 14:50 . . (+208) . . Ultimateria (Talk | contribs) (Created page with "===
skin # to criticize =====Conjugation===== {{es-conj-ar|despellej}} =====Synonyms===== * {{/es|des

(diff | hist) . . **N** вооружение; 14:49 . . (+102) . . Stephen G. Brown (Talk | contribs) (→Noun)

(diff | hist) . . **N** तीर्थ: 14:49 . . (+928) . . Rajasekhar1961 (Talk | contribs) (Created page with "===Sa
{{Hinduism|lang=sa}} a passage, way, road, ford, stairs for landing or for descent into a river, bath

(diff | hist) . . **N** ковчег завета; 14:48 . . (+163) . . Stephen G. Brown (Talk | contribs) (→Noun)

(diff | hist) . . **m** femme fatale; 14:48 . . (+39) . . Norm mit (Talk | contribs) (Add pronunciation in transcription)

(diff | hist) . . **N** geschwommen; 14:47 . . (+80) . . Sae1962 (Talk | contribs) (Added German entry)

(diff | hist) . . **N** арность; 14:47 . . (+114) . . Stephen G. Brown (Talk | contribs) (→Noun)

(diff | hist) . . **m** gegangen; 14:47 . . (+5) . . Sae1962 (Talk | contribs)

(diff | hist) . . **N** pellejos; 14:47 . . (+99) . . Ultimateria (Talk | contribs) (Creating plural form of pellejo (Accelerated))

(diff | hist) . . **N** schwamm; 14:47 . . (+169) . . Sae1962 (Talk | contribs) (Added German entry)

(diff | hist) . . **m** عمو: 14:47 . . (+80) . . Dick Laurent (Talk | contribs) (→Noun: fa-poss-u need)

(diff | hist) . . **N** аристотелизм; 14:46 . . (+142) . . Stephen G. Brown (Talk | contribs) (→Noun)

(diff | hist) . . Aristotelianism; 14:46 . . (+8) . . Stephen G. Brown (Talk | contribs) (→Translations)

(diff | hist) . . **N** Арситофан; 14:45 . . (+128) . . Stephen G. Brown (Talk | contribs) (→Proper noun)

(diff | hist) . . **m** floated; 14:45 . . (+8) . . Sae1962 (Talk | contribs) (→Verb: Added a template)

(diff | hist) . . **m** basketeer; 14:44 . . (+37) . . Equinox (Talk | contribs) (→English: ===Etymology=== {{suffix|basket|eer}})

(diff | hist) . . **N** prašānu; 14:44 . . (+126) . . Pereru (Talk | contribs) (Created page with "===Latvian=== =====Noun===== {{head|lv|noun form|g=f}} #
{{lv-inflection of|prašānalacis}} # {{lv-inflection of|prašānals}}")

Multilingualism

Size

Up-to-dateness

Dictionary Analysis and Comparison



- (diff | hist) . తీర్థము; 14:51 .. (+109) .. Rajasekhar1961 (Talk | contribs) (→Telugu)
- (diff | hist) . horseshoe; 14:50 .. (+24) .. 115.132.182.28 (Talk) (t+ms:ladam t-balance (Assisted))
- (diff | hist) . **N** армада; 14:50 .. (+132) .. Stephen G. Brown (Talk | contribs) (→Noun)
- (diff | hist) . excite; 14:50 .. (+18) .. Jc86035 (Talk | contribs) (+rhyme -art)
- (diff | hist) . **N** despellejar; 14:50 .. (+208) .. Ultimateria (Talk | contribs) (Created page with "===
skin # to criticize =====Conjugation===== {{es-conj-ar|despellej}} =====Synonyms===== * {{/es|des
- (diff | hist) . **N** вооружение; 14:49 .. (+102) .. Stephen G. Brown (Talk | contribs) (→Noun)
- (diff | hist) . **N** तीर्थ; 14:49 .. (+928) .. Rajasekhar1961 (Talk | contribs) (Created page with "===Sa
{{Hinduism|lang=sa}} a passage, way, road, ford, stairs for landing or for descent into a river, bath
- (diff | hist) . **N** ковчег завета; 14:48 .. (+163) .. Stephen G. Brown (Talk | contribs) (→Noun)
- (diff | hist) . **m** femme fatale; 14:48 .. (+39) .. Norm mit (Talk | contribs) (Add pronunciation in transcription)
- (diff | hist) . **N** geschwommen; 14:47 .. (+80) .. Sae1962 (Talk | contribs) (Added German entry)
- (diff | hist) . **N** арность; 14:47 .. (+114) .. Stephen G. Brown (Talk | contribs) (→Noun)
- (diff | hist) . **m** gegangen; 14:47 .. (+5) .. Sae1962 (Talk | contribs)
- (diff | hist) . **N** pellejos; 14:47 .. (+99) .. Ultimateria (Talk | contribs) (Creating plural form of pellejo (Accelerated))
- (diff | hist) . **N** schwamm; 14:47 .. (+169) .. Sae1962 (Talk | contribs) (Added German entry)
- (diff | hist) . **m** عمو; 14:47 .. (+80) .. Dick Laurent (Talk | contribs) (→Noun: fa-poss-u need)
- (diff | hist) . **N** аристотелизм; 14:46 .. (+142) .. Stephen G. Brown (Talk | contribs) (→Noun)
- (diff | hist) . Aristotelianism; 14:46 .. (+8) .. Stephen G. Brown (Talk | contribs) (→Translations)
- (diff | hist) . **N** Арситофан; 14:45 .. (+128) .. Stephen G. Brown (Talk | contribs) (→Proper noun)
- (diff | hist) . **m** floated; 14:45 .. (+8) .. Sae1962 (Talk | contribs) (→Verb: Added a template)
- (diff | hist) . **m** basketeer; 14:44 .. (+37) .. Equinox (Talk | contribs) (→English: ===Etymology=== {{suffix|basket|eer}})
- (diff | hist) . **N** prašanu; 14:44 .. (+126) .. Pereru (Talk | contribs) (Created page with "===Latvian=== =====Noun===== {{head|lv|noun form|g=f}} #
{{lv-inflection of|prašān|acc|s}} # {{lv-inflection of|prašān|ins|s}}")

Multilingualism

Size

Up-to-dateness

History

Dictionary Analysis and Comparison

Sprachvarietäten

- **Dialekte:** Liste deutscher Dialekte • Hunsrücker Mundart • Kärnter Mundart • Wäller Platt
- **Außereuropäische Varietäten:** Außereuropäische Varietäten des Deutschen
- **Fachsprache:** Fliegersprache • Hip-Hop-Jargon • Jägersprache • Netzjargon Polizeijargon • Soldatensprache • Soldatensprache (Schweizer Armee) • **Seemannssprache**

Wissenschaft

- Astronomie • Geowissenschaften • Informatik • Linguistik • Mathematik • Medizin • Pharmazie • Philosophie • Philosophie • Physik • Psychologie • Rechtswissenschaften
- **Biologie:** Ökologie • Zoologie • Zytologie und Genetik
- **Chemie:** Anorganische Stoffe • Organische Stoffe

Listen von Namen

- Liste von Staatennamen
- Liste der Männlichen Vornamen
- Liste der Weiblichen Vornamen
- Liste der häufigsten weiblichen Vornamen Deutschlands
- Liste der häufigsten männlichen Vornamen Deutschlands
- Liste geschlechtsneutraler Vornamen
- Liste der Kosenamen und Kosewörter
- Liste der häufigsten Nachnamen Deutschlands / Österreichs

Gesellschaft

- Berufe • Politik
- *Essen und Trinken:* Getränke • Speisen • Tätigkeiten und Eigenschaften • Deutschsprachige Küchenbegriffe nach Regionen • Küche und Küchengeräte • Weinbau
- *Sport und Freizeit:* Sportarten • **Fußball** • Golf • Radsport • Schach • Tanzen

Weitere Wortlisten

- **Abkürzungen:** Liste der Abkürzungen im Internet • ASCII-Smileys • Liste

Multilingualism

und Übersichten: Liste der Wörter,

Size

Bedeutung in verschiedenen

Up-to-dateness

Eisenbahnwesen • Fahrzeugtechnik •

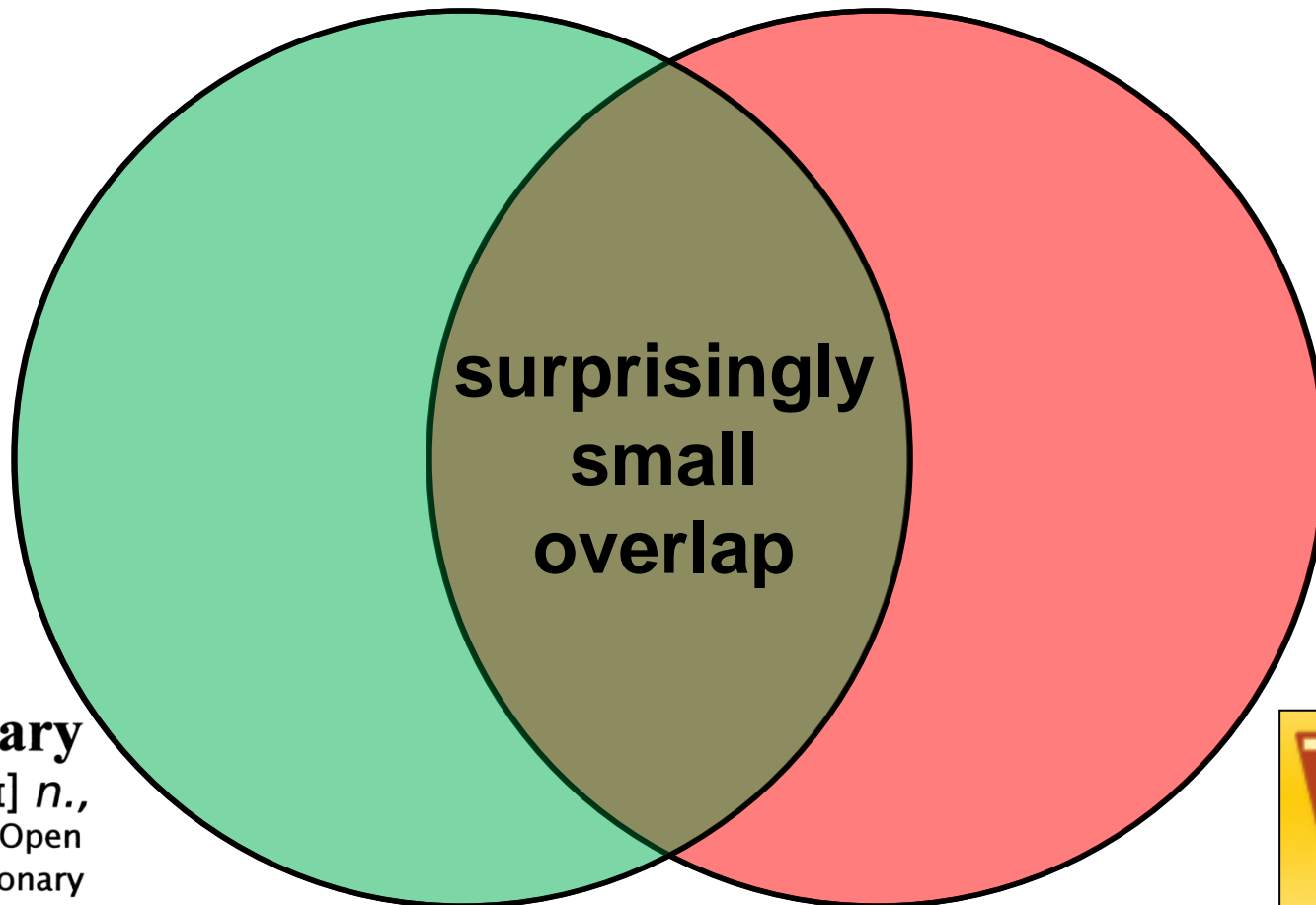
History

Topic & Variety

Overlap of Lexical Entries



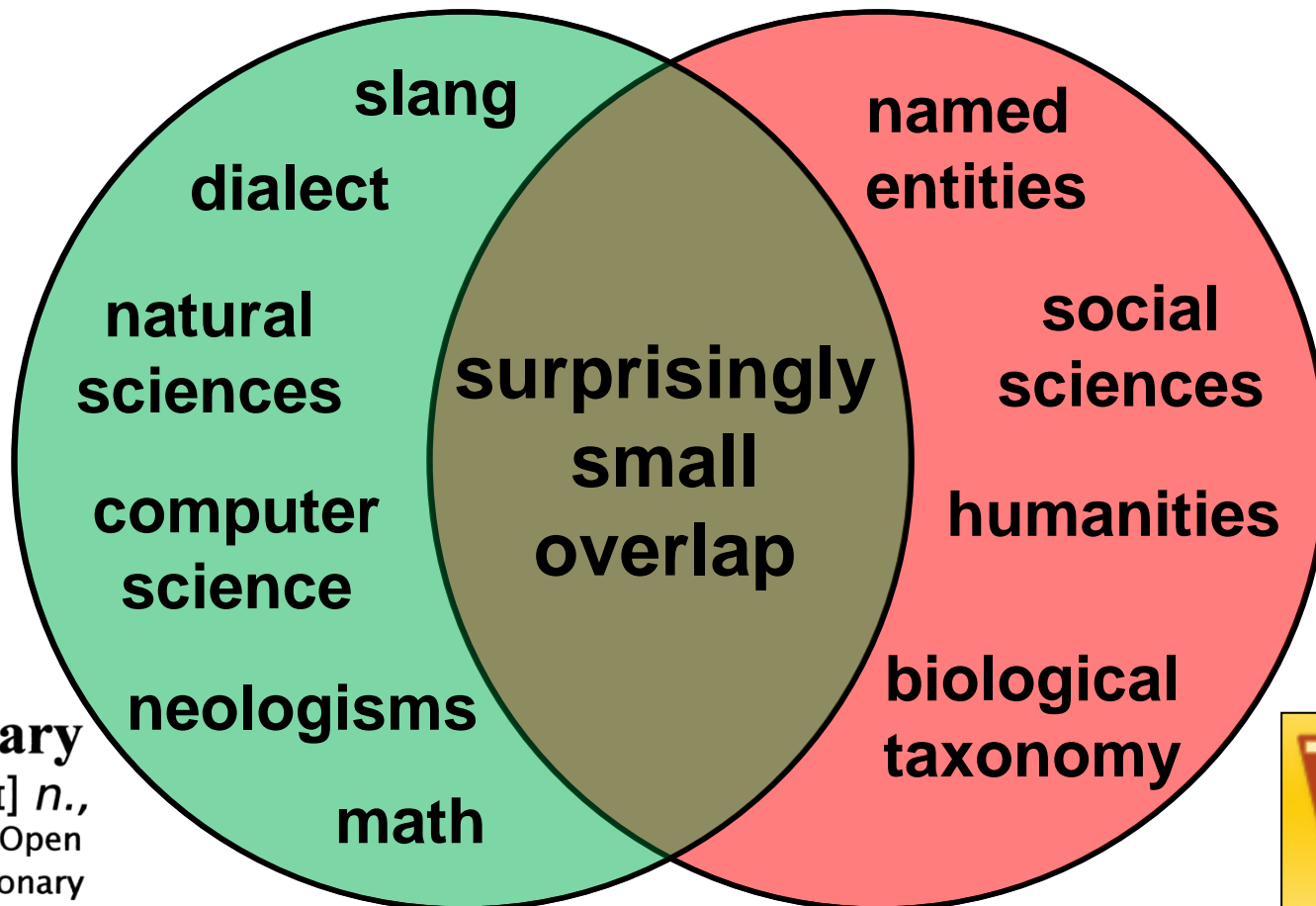
TECHNISCHE
UNIVERSITÄT
DARMSTADT



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



Overlap of Lexical Entries



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




But...

Wiktionary is far from being perfect

- Different developmental stages of articles
- Inconsistencies
- Spelling errors
- Old-fashioned and overly general definitions

No systematic corpus analysis

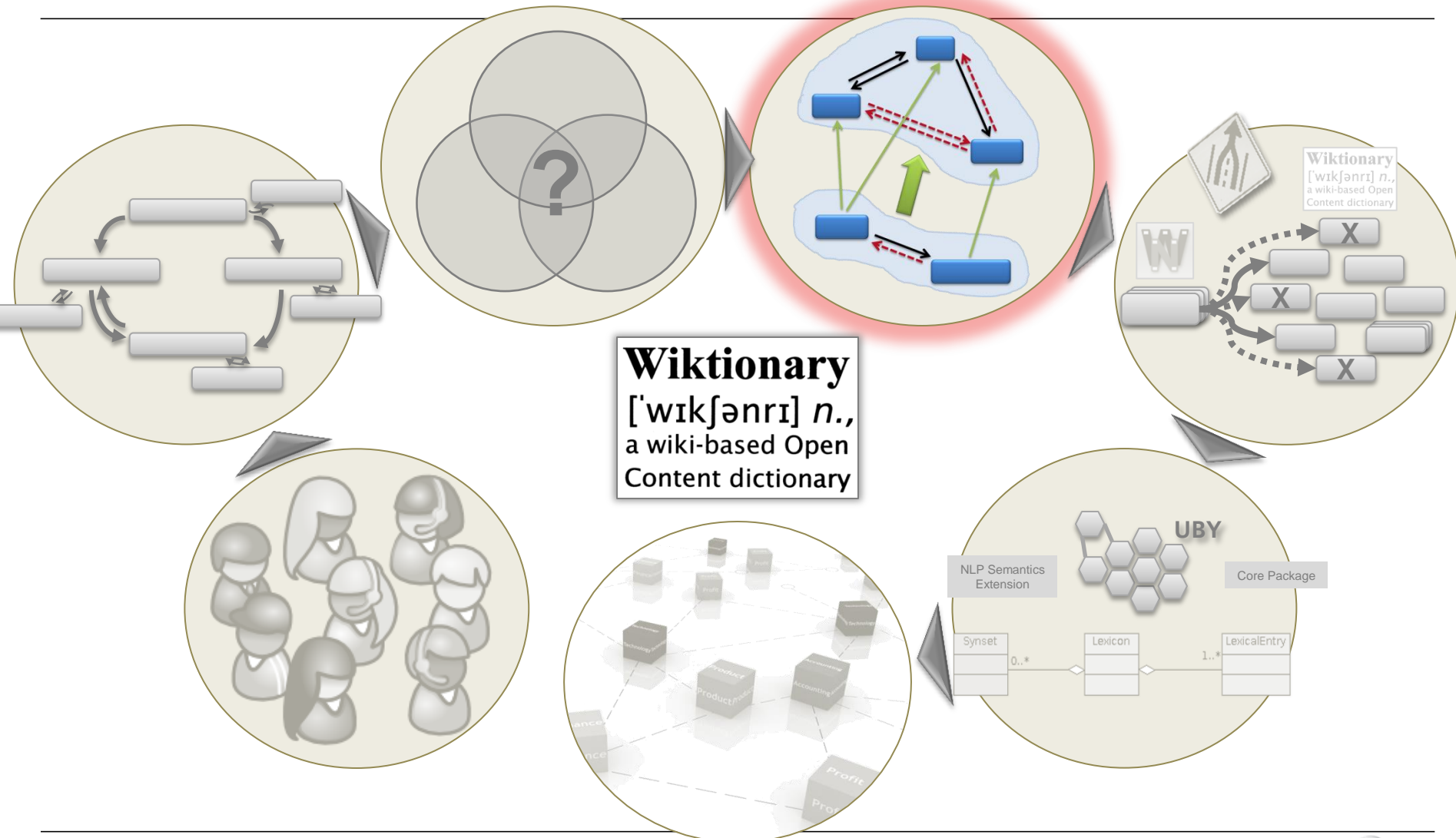
- Many references to expert-built dictionaries
- Sum of subjective opinions of the authors



A user suggests that this entry be cleaned up, giving the reason: “clean up music definitions”.

Please see the discussion on [Requests for cleanup^{\(+\)}](#) or the [talk page](#) for more information and remove this template after the problem has been dealt with.

Agenda



Harvesting Linguistic Knowledge from...



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Machine-readable Dictionaries?

- Usually high quality
- Weakly structured; often not freely available



Wordnets?

- Commonly used; well-structured
- Small coverage, in particular multilingual ones



Harvesting Linguistic Knowledge from...



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Machine-readable Dictionaries?

- Usually high quality
- Weakly structured; often not freely available



Wordnets?

- Commonly used; well-structured
- Small coverage, in particular multilingual ones



Wikipedia?

- High coverage of factual knowledge
- No linguistic knowledge (word usage, verbs,...)



WIKIPEDIA
The Free Encyclopedia

Wiktionary!

- Freely available, high coverage, all parts of speech

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

Adding Structure to Wiktionary

Structure

Dictionary

Ontology

puppy

Contents [show]

English

Pronunciation

- Audio (US)  0:00   MENU
- Rhymes: -ʌpi

Noun

puppy (*plural* **puppies**)

1. A young dog.
2. A young rat.
3. A young seal.
4. (*slang, usually in the plural*) A woman's **breast**.
5. (*informal*) A (generic) **thing**; particularly something that
*I have another two dozen of these **puppies** to finish*

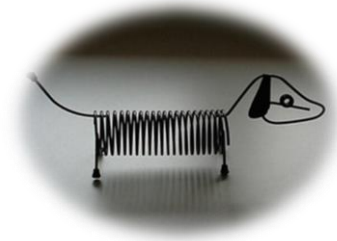
{dog₁, hound₁, canine₁}

Concepts

IS-A

Relation

{puppy₁, juvenile dog₁}



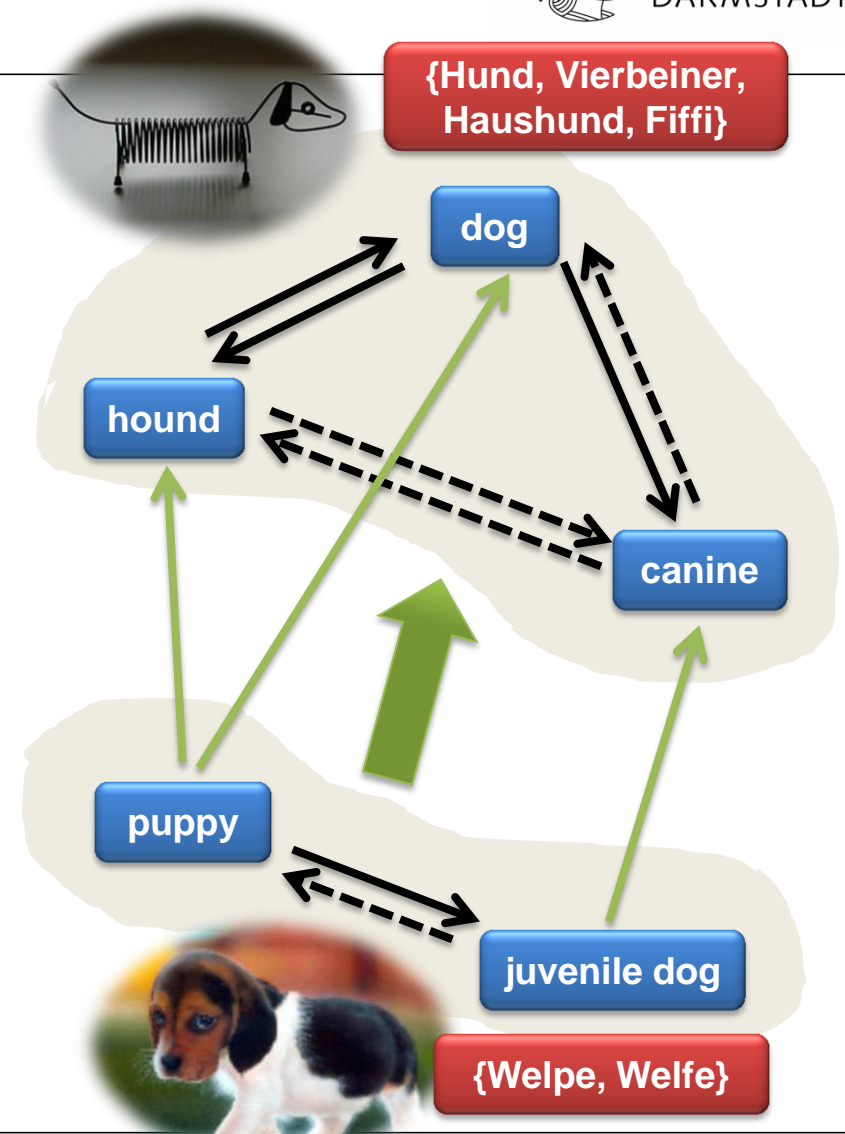
1. Concept formation

- Start with Wiktionary senses
- Create synonymy graph
- Calculate the transitive hull
- Synsets → concepts

2. Conceptual relations

- From lexical relations to conceptual relations

3. Multilingual Lexicalizations



Disambiguation of Relations and Translations



boat (plural **boats**)

1. A **craft** used for **transportation** of goods, fishing, racing, recreational cruising, or military use on or in the **water**, **propelled** by **oars** or **outboard motor** or **inboard motor** or by **wind**.
2. (*poker*, *hang*) A full house.
3. A **vehicle**, **utensil**, or **dish** somewhat resembling a boat in shape.
*a stone **boat**; a gravy **boat***
4. (*chemistry*) One of two possible **conformations** of **cyclohexane** rings (the other being **chair**), shaped roughly like a boat.
5. (*Australia*, *politics*, *informal*) The refugee boats arriving in Australian waters, and by extension, refugees generally.

Synonyms

- (*craft on or in water*): **craft**, ship, vessel

Translations

water craft [hide ▲]

Select targeted languages

• Afrikaans: boot	• Lojban: bloti
• Ainu: チブ (cip)	• Luxembourgish: <i>please add this translation if you can</i>
• Albanian: varkë <i>f</i>	

Disambiguation of Relations and Translations

boat (plural **boats**)

1. A **craft** used for **transportation** of goods, fishing, racing, recreational cruising, or military use on or in the **water**, propelled by **oars** or **outboard motor** or **inboard motor** or by **wind**.
2. (*poker, slang*) A full house.
3. A **vehicle**, **utensil**, or **dish** somewhat resembling a boat in shape.
*a stone **boat**; a gravy **boat***
4. (*chemistry*) One of two possible **conformations** of a molecule, especially one that is shaped roughly like a boat.
5. (*Australia, politics, informal*) The refugee boat extension, refugees generally.

Synonyms

- (*craft on or in water*): **craft**, **ship**, **vessel**

Translations

water craft

- Afrikaans: *boot*
- Ainu: チブ (cip)
- Albanian: *varkë f*

Automatic
Disambiguation

craft (plural **craft** or -)

- ✗ The **skilled practice** of a practical occupation.
*She represented the **craft** of brewers.*
- ✗ (*nautical, whaling*) Implements used in catching fish, such as **net**, **line**, or **hook**. Modern use primarily in whaling, as in **harpoons**, **hand-lances**, etc. [quotations ▼]
- ✓ (*nautical*) Boats, especially of smaller size than **ships**. Historically primarily applied to vessels engaged in loading or unloading of other vessels, as **lighters**, **hoys**, and **barques**.
- ✗ (*nautical, British Royal Navy*) Those vessels attendant on a **fleet**, such as **cutters**, **schooners**, and **gun-boats**, generally commanded by **lieutenants**.
- ✗ A **vehicle** designed for **navigation** in or on water or air or through **outer space**.
- ✗ A **particular kind of skilled work**.

Disambiguation Features (Selection)



craft (*countable and uncountable; plural craft or crafts*)

[1] The skilled practice of a practical occupation.

definition

*She represented the **craft** of brewers.*

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

[3] (*nautical*) Boats used primarily a loading or unloading of other vessels, as lighters, hovers, and barges.

lemma of the source relation

[4] (*nautical, British Royal Navy*) Those such as cutters generally commanded by lieutenants.

pragmatic labels

[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.

Disambiguation Features (Selection)



craft (*countable and uncountable; plural craft or crafts*)

[1] The skilled practice of a practical occupation.

definition

She represented the craft of brewers.

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

[3] (*nautical*) Boats especially those used for the purpose of assisting in the loading or unloading of other vessels, as lighters, hovers, and barges.

lemma of the source relation

[4] (*nautical, British Royal Navy*) Those officers of the Royal Navy who are generally commanded by lieutenants.

pragmatic labels

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

[6] A particular kind of skill.

Synonyms

• (*skill at work*): craftsmanship, workmanship

shared relations

• (*nautical sense*): boat

inverse relation

• (*vehicle*):

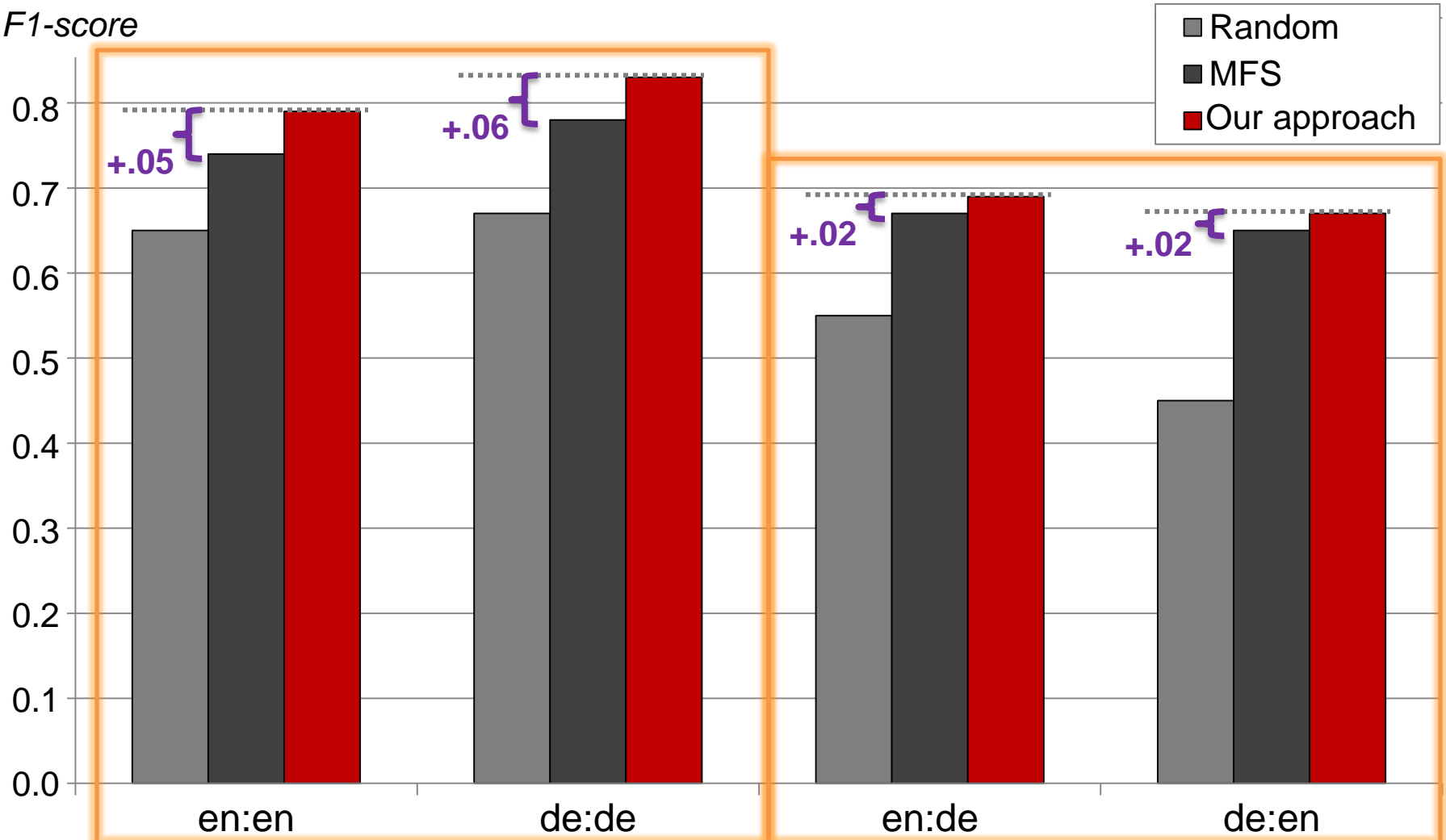
• (*kind of skilled work*): trade

• (*shrewdness*): craftiness, cunning, foxiness, guile, slyness, wiliness

Hyponyms

Evaluation Results

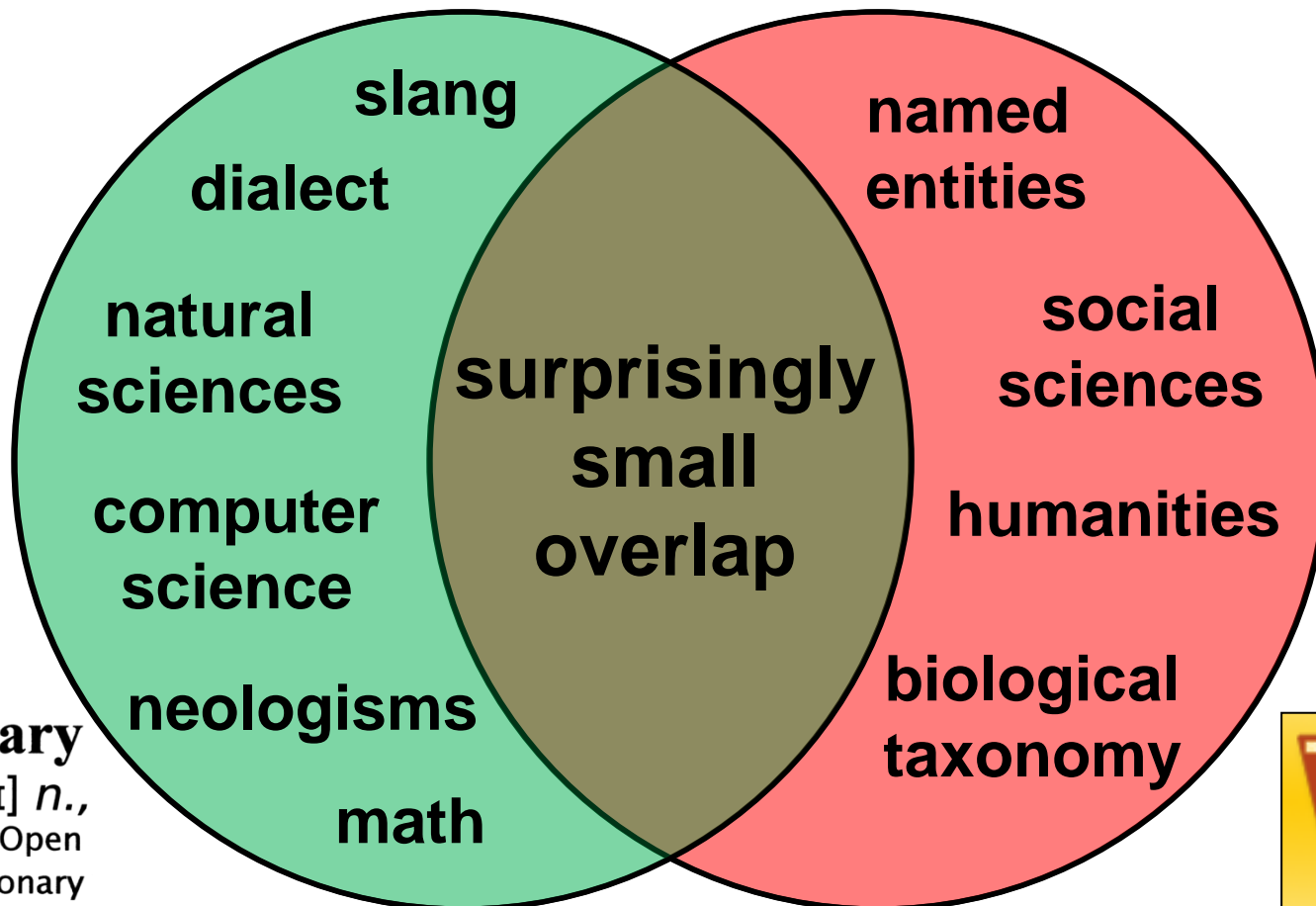
F1-score



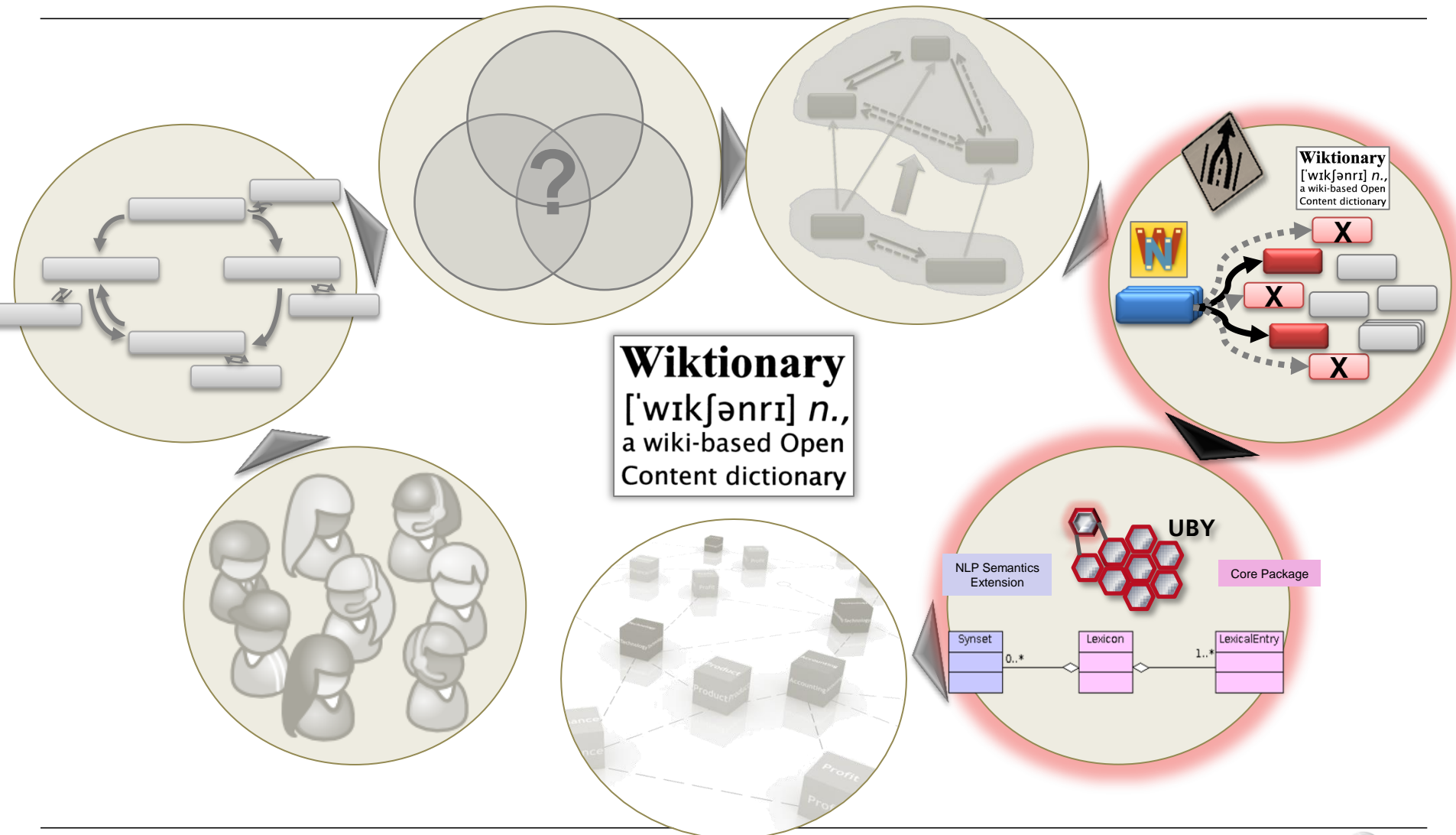
Overlap of Lexical Entries – Revisited



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Agenda



Word Sense Alignment

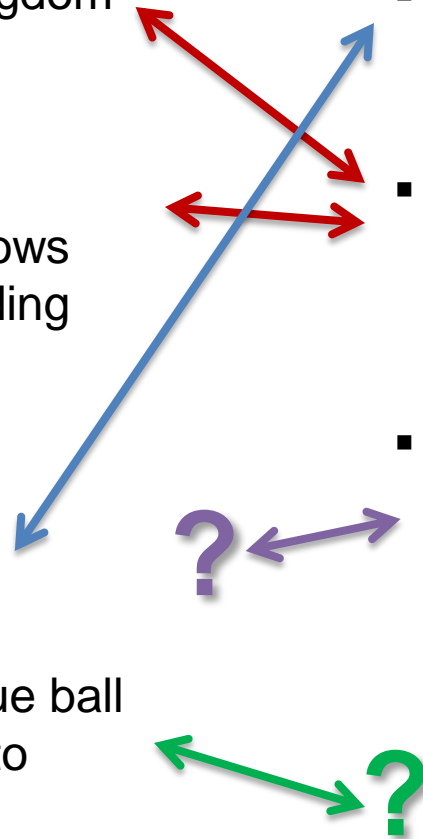


plant in Wiktionary

- (**botany**) An organism of the kingdom Plantae [...]
- (proscribed as **biologically** inaccurate) Any creature that grows on soil or similar surfaces, including plants and fungi.
- A **factory** or other industrial or institutional building or facility.
- (**snooker**) A play in which the cue ball knocks one (usually red) ball onto another [...]

plant in WordNet

- buildings for carrying on **industrial labor**
- (**botany**) a living organism lacking the power of locomotion
- an **actor** situated in the audience whose acting is rehearsed but seems spontaneous to the audience



Aligning Wiktionary and WordNet

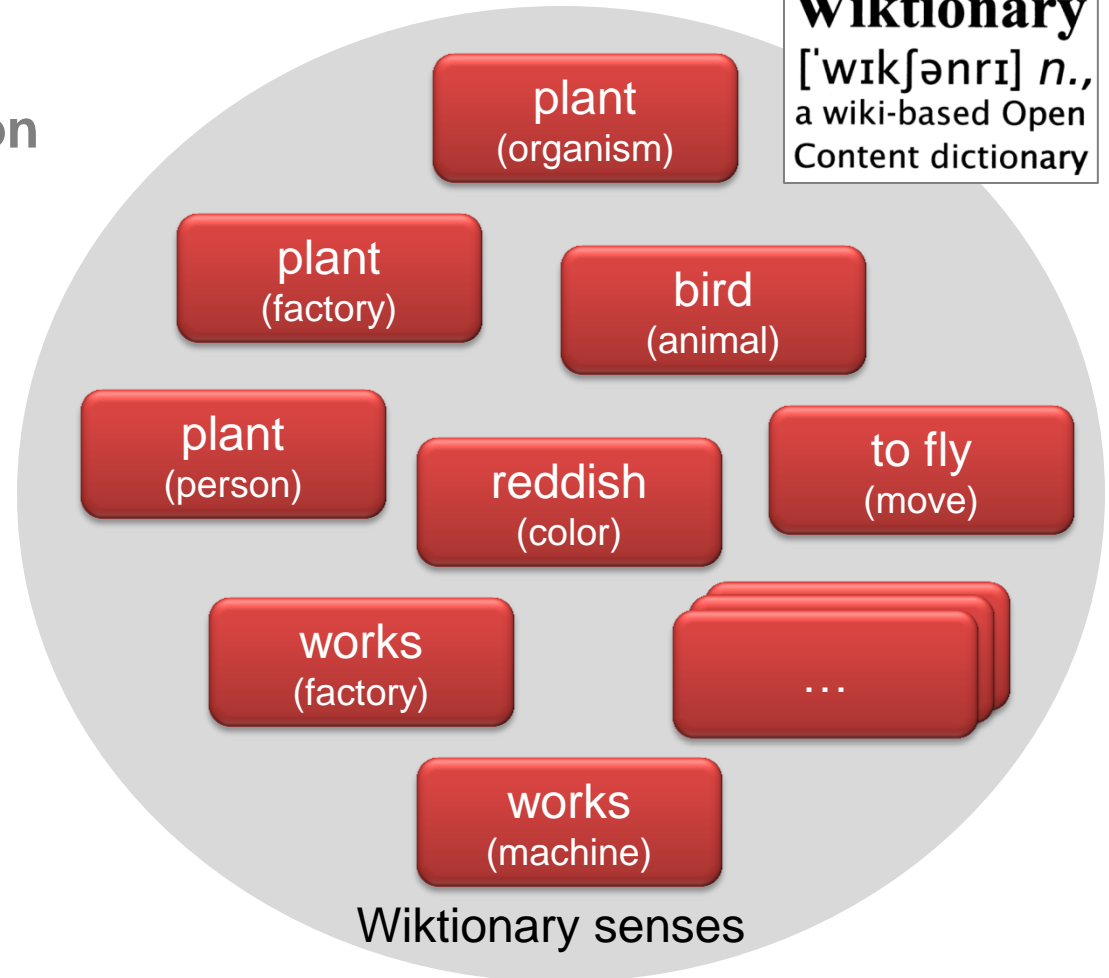
A two-step approach:

1. Candidate extraction
2. Candidate disambiguation



WordNet synsets

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



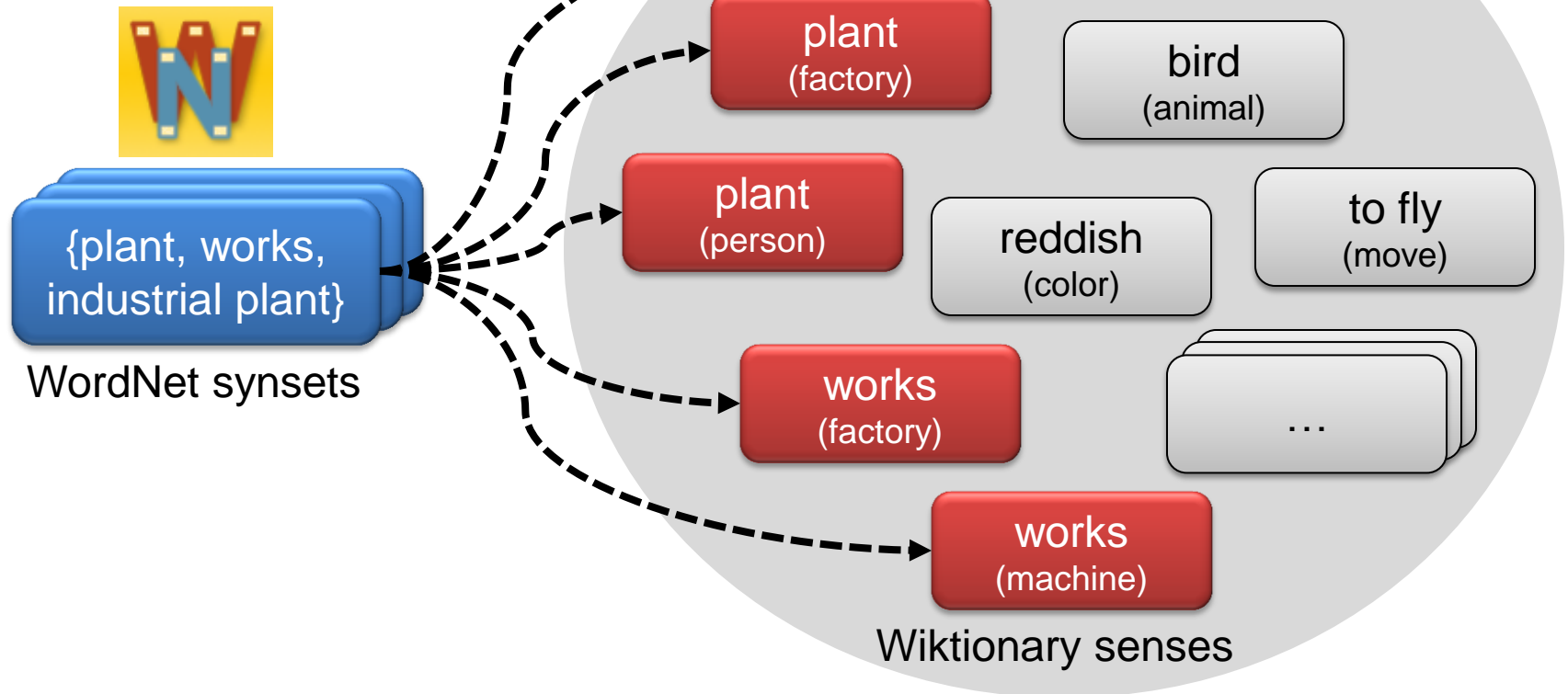
Wiktionary senses

Aligning Wiktionary and WordNet

A two-step approach:

1. Candidate extraction
2. Candidate disambiguation

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

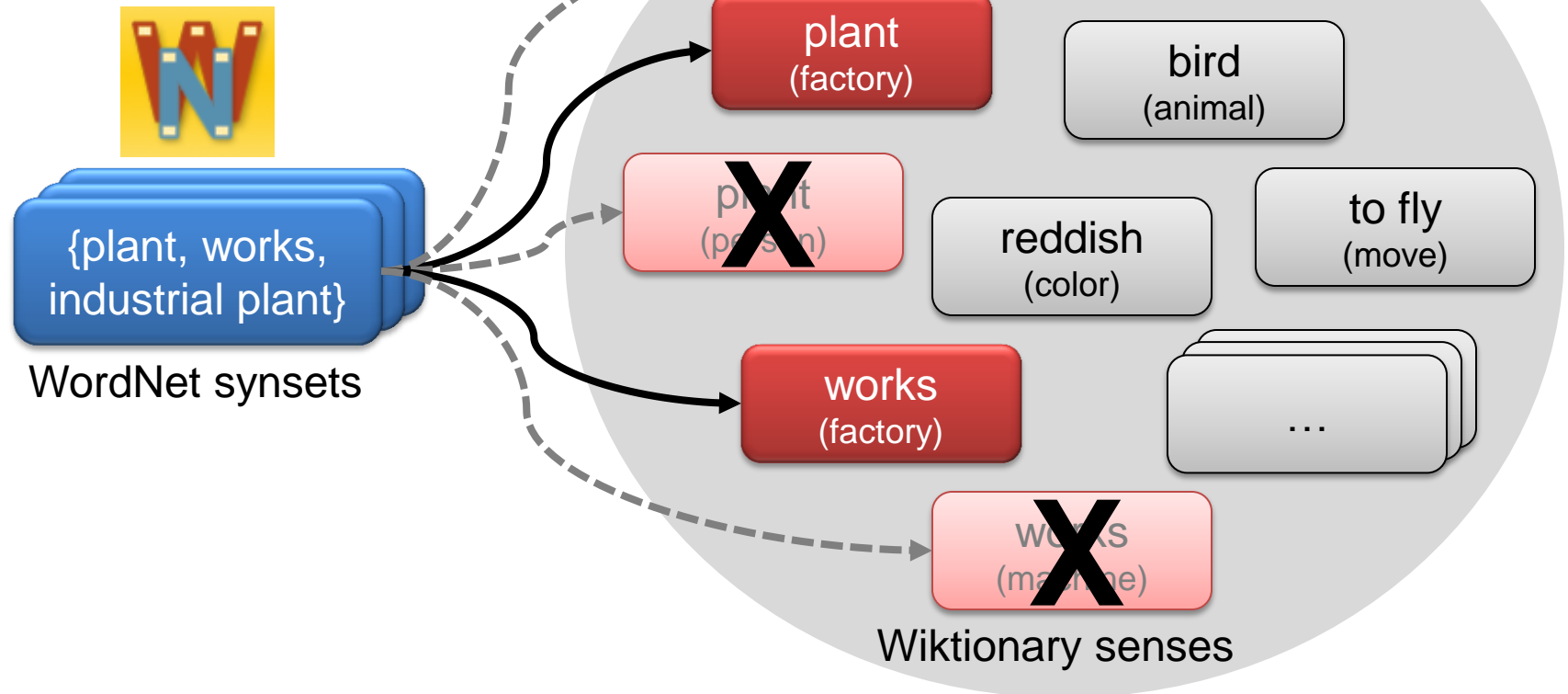


Aligning Wiktionary and WordNet

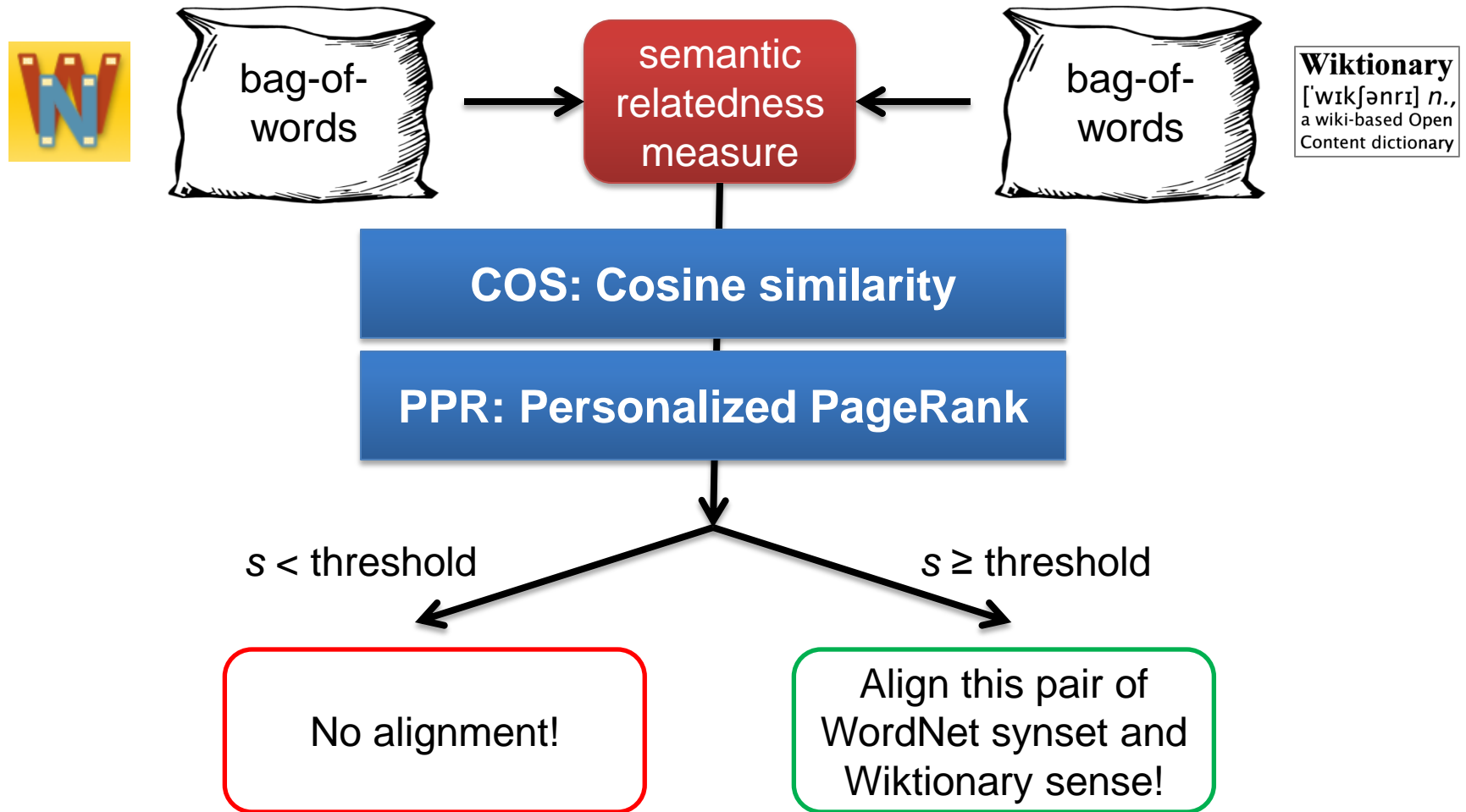
A two-step approach:

1. Candidate extraction
2. **Candidate disambiguation**

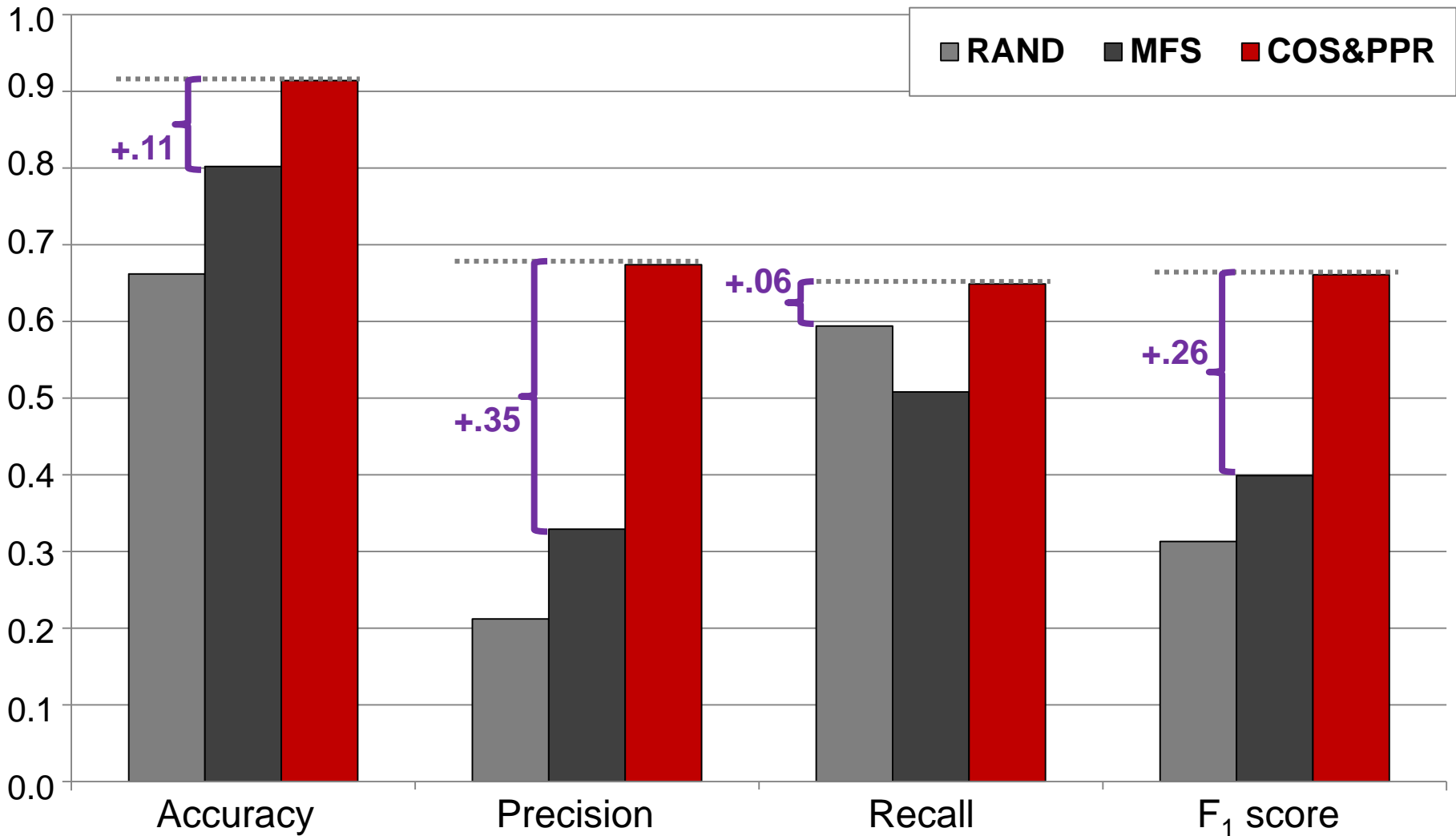
Wiktionary
[ˈwɪkʃənri] *n.*,
a wiki-based Open
Content dictionary



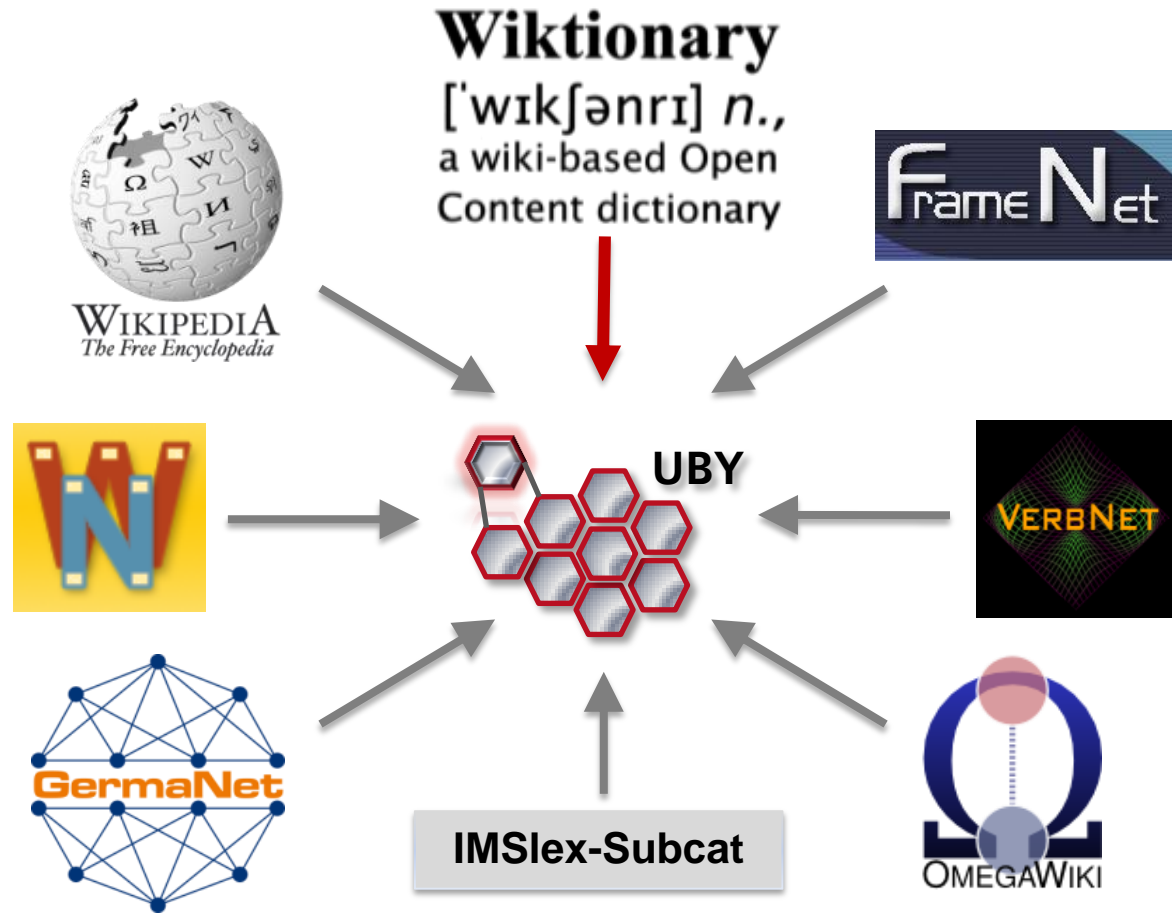
Candidate Disambiguation



Alignment Evaluation

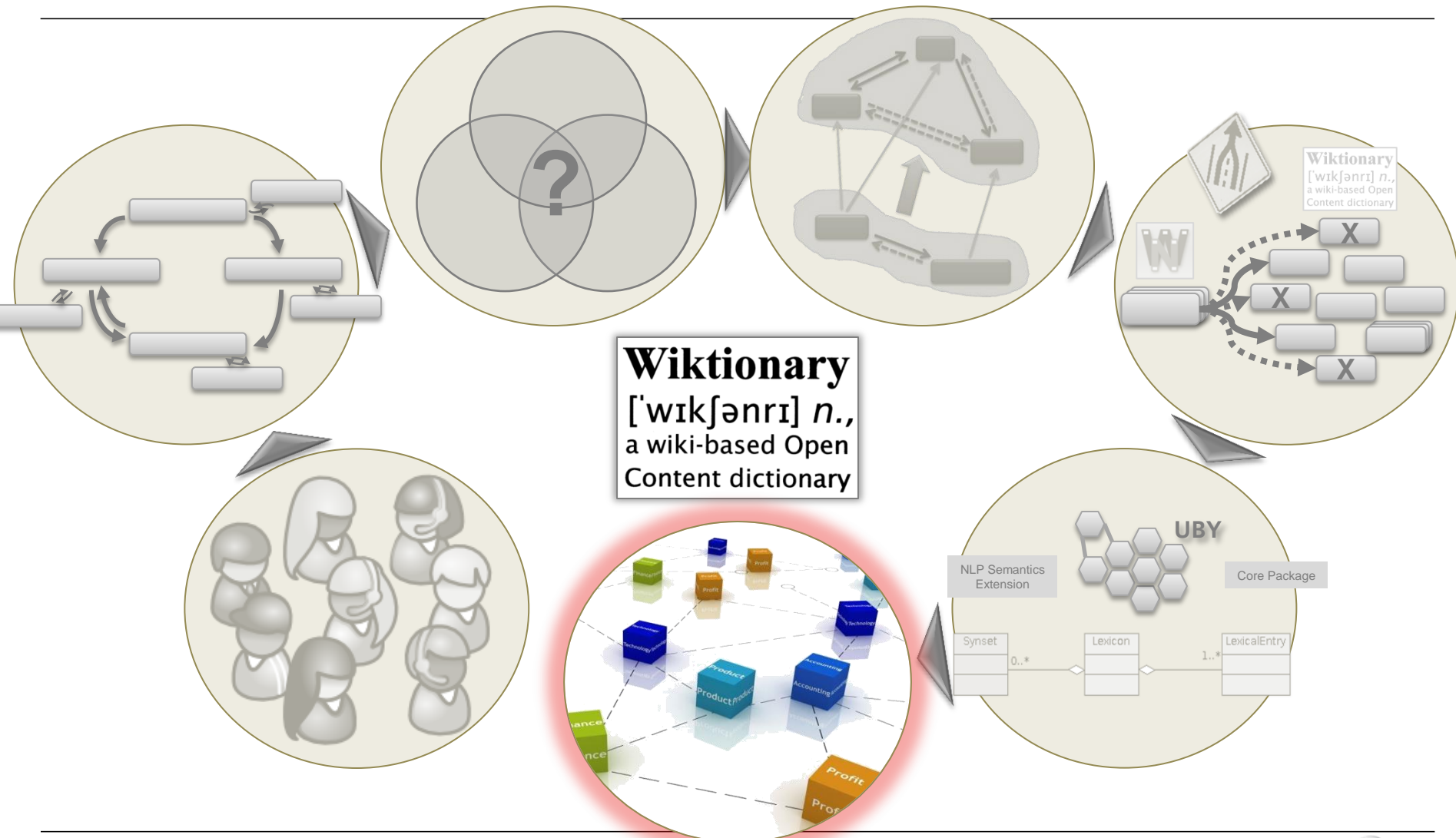


Integrated Resource UBY



Joint work of Judith Eckle-Kohler, Iryna Gurevych, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, Tri-Duc Nghiem

Agenda



Case Study: Detecting Marketing Blunders



TECHNISCHE
UNIVERSITÄT
DARMSTADT



[http://commons.wikimedia.org/wiki/File:Rolls-Royce_Silver_Shadow_I_Bj._1967_\(2005-09-17_Sp\)-2.jpg](http://commons.wikimedia.org/wiki/File:Rolls-Royce_Silver_Shadow_I_Bj._1967_(2005-09-17_Sp)-2.jpg)
(CC BY-SA 2.0 by Wikipedia users Lothar Spurzem and Sir James)

Case Study: Detecting Marketing Blunders



TECHNISCHE
UNIVERSITÄT
DARMSTADT

**Internationally advertised as “Silver Mist”
But: “Mist” means dung/manure in German!**



Case Study: Detecting Marketing Blunders



TECHNISCHE
UNIVERSITÄT
DARMSTADT

**Internationally advertised as “Silver Mist”
But: “Mist” means dung/manure in German!**








**Avoiding such blunders is challenging, since
there are no tools and only few resources**

**Wiktionary covers many languages and language varieties
(including slang, offensive and vulgar terms, etc.)**

Case Study: Detecting Marketing Blunders

Approach:








- Index translations and pragmatic labels from OntoWiktionary
- Lookup word forms for carrying a problematic meaning in any language

Mist		Animal excrement, especially that of common domestic farm animals [...]
Mist		(uncountable) Manure; animal excrement.
Mist		(mildly vulgar, slang, in the plural) Nonsense.
mist	  	(uncountable) A thick cloud that forms near the ground [..]
mist		(countable) A layer of fine droplets or particles.
...

Case Study: Detecting Marketing Blunders

Approach:

- Index translations and pragmatic labels from OntoWiktionary
- Lookup word forms for carrying a problematic meaning in any language

Mist		Animal excrement, especially that of common domestic farm animals [...]
Mist		(uncountable) Manure; animal excrement.
Mist		(mildly vulgar, slang, in the plural) Nonsense.
mist	  	(uncountable) A thick cloud that forms near the ground [..]
mist		(countable) A layer of fine droplets or particles.
...

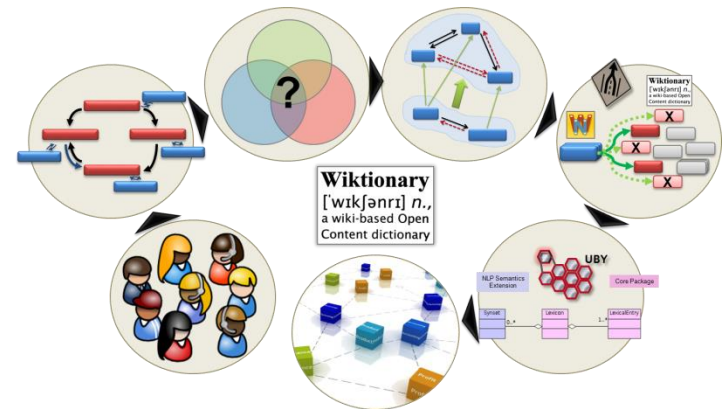
Evaluation:

- Copywriters become able to detect **35 out of the 45 marketing blunders**
→ Recall 0.78 (primary goal)
- They need to check **on average 12 clues per blunder**, of which 5 are relevant → Precision 0.41 (secondary goal)

Conclusion

Analyzing Wiktionary's potential

- neologisms, translations, language varieties, language feeling
- subjective opinions rather than corpus evidence



Conclusion

Analyzing Wiktionary's potential

- neologisms, translations, language varieties, language feeling
- subjective opinions rather than corpus evidence

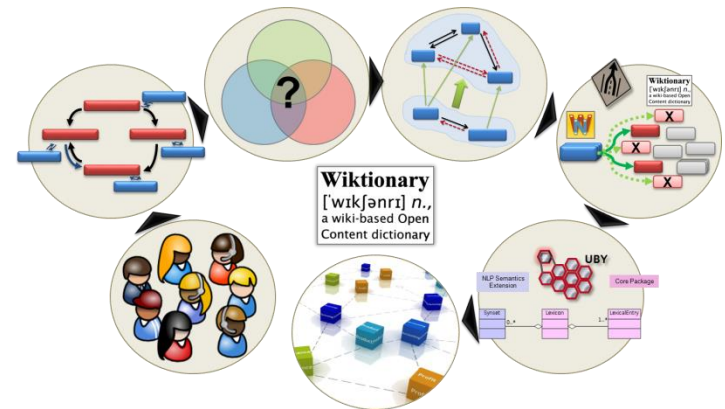
Creating OntoWiktionary

- Automatic disambiguation of relations

Integrating language resources

- Aligning Wiktionary and WordNet
- Standardized representation

Detecting marketing blunders



Publicly available data and software:

- ▶ <http://www.ukp.tu-darmstadt.de/data/lexical-resources/wiktionary/>
- ▶ <http://www.christian-meyer.org/research/publications/dissertation/data/>

Joint forces: experts and collaborative authors

Relation disambiguation:

- graph-based methods
- crowd-sourced disambiguation based on Wikidata

Word sense alignment:

- other information types
- more than two resources

Much potential for further NLP applications



Thank you for your attention!


Questions?





Kontakt / Contact

Christian M. Meyer

Technische Universität Darmstadt
Ubiquitous Knowledge Processing Lab

 Hochschulstr. 10, 64289 Darmstadt, Germany

 +49 (0)6151 16–5386

 +49 (0)6151 16–5455

 meyer (at) ukp.informatik.tu-darmstadt.de

Rechtliche Hinweise

Die Folien sind für den persönlichen Gebrauch der Vortragsteilnehmer gedacht. Im Vortrag verwendete Photographien, Illustrationen, Wort- und Bildmarken sind Eigentum der jeweiligen Rechteinhaber oder Lizenzgeber. Um Missverständnisse zu vermeiden, wäre eine kurze Kontaktaufnahme vor Weitergabe oder -nutzung der Vortragsmaterialien empfehlenswert. Sofern Sie Ihre Rechte verletzt sehen, bitte ich ebenfalls um Kontaktaufnahme zur Klärung der Sachlage.

Legal Issues

The slides are intended for personal use by the audience of the talk. Photographies, illustrations, trademarks, or logos are property of the holder of rights. To avoid any misconceptions, I would strongly recommend to get in touch before reusing or redistributing the slides or any additional material of the talk. The same applies if you consider your rights infringed – please let me know to initiate further clarification.