Adjusting sense representations for knowledge-based word sense disambiguation and automatic pun interpretation

TECHNISCHE UNIVERSITÄT DARMSTADT

Tristan Miller

Presented at:
School of Data Analysis and Artificial Intelligence
National Research University – Higher School of Economics
25 May 2017

Biography

TECHNISCHE UNIVERSITÄT DARMSTADT

Technische Universität Darmstadt

Biography

TECHNISCHE UNIVERSITÄT DARMSTADT

UBIQUITOUS KNOWLEDGE PROCESSING

- argumentation mining
- language technology for the digital humanities
- lexical-semantic resources and algorithms
- text mining and analytics
- writing assistance and language learning

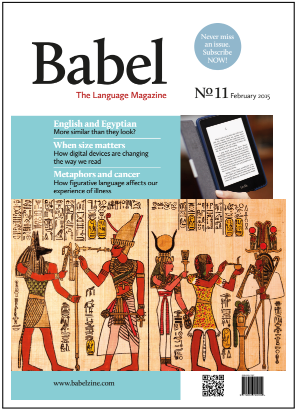
Biography

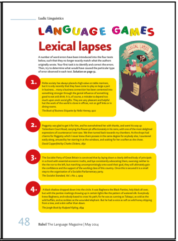
University of Regina

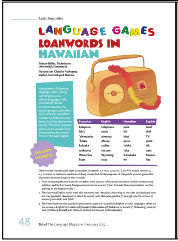
University of Toronto

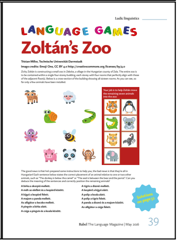
Biography

TECHNISCHE
UNIVERSITÄT
DARMSTADT









Agenda

Introduction

Knowledge-based word sense disambiguation

Pun interpretation

Conclusion

Introduction

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Polysemy is a characteristic of all natural languages.

Introduction

Polysemy is a characteristic of all natural languages.

"He hit the ball with the bat."

Introduction

Polysemy is a characteristic of all natural languages.

"He hit the ball with the bat."

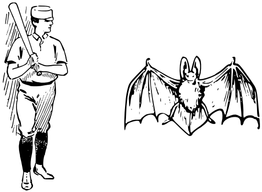
Introduction

Polysemy is a characteristic of all natural languages.

"He hit the ball with the bat."

Polysemy is a characteristic of all natural languages.

"He hit the ball with the bat."

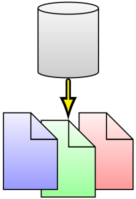


Word sense disambiguation (WSD) is the task of determining which of a word's senses is intended in a given context.

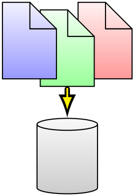
Applications of word sense disambiguation

bat { Fledermaus, Schläger, Schlagstock, Brandschiefer, Brennbuch, schlagen, blinzeln

Machine translation



Information retrieval



Information extraction

Spelling correction

Why is WSD hard?
Many different formulizations and parameterizations

TECHNISCHE UNIVERSITÄT DARMSTADT

high-coverage
or
high-precision?

fixed inventory
or
induced inventory?

coarse-grained
or
fine-grained?

WSD

domain-independent
or
domain-specific?

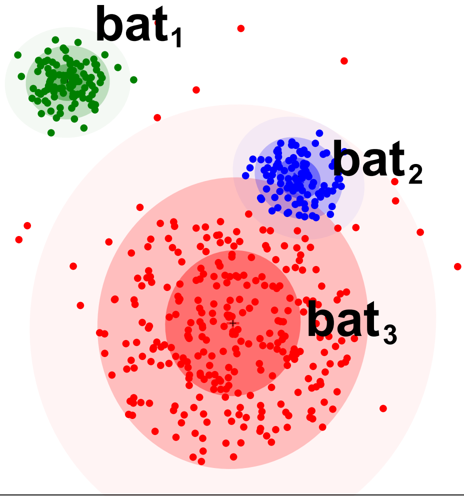
lexical sample
or
all-words?

language-independent
or
language-specific?

monolingual
or
crosslingual?

Why is WSD hard?
Nature of word senses unclear

TECHNISCHE UNIVERSITÄT DARMSTADT

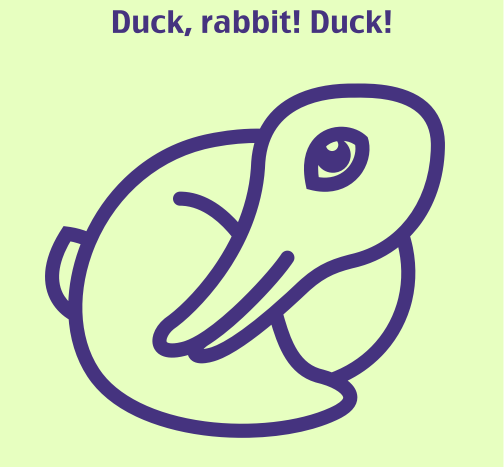


Why is WSD hard?
Knowledge acquisition bottleneck

TECHNISCHE UNIVERSITÄT DARMSTADT

Why is WSD hard?
Some usages are deliberately ambiguous

TECHNISCHE UNIVERSITÄT DARMSTADT



Approaches to word sense disambiguation

	supervised	knowledge-based
input	▸ manually annotated training examples	▸ machine readable dictionaries (MRDs) ▸ lexical semantic resources (LSRs) ▸ unannotated corpora
pros	▸ better performance	▸ wider applicability
cons	▸ knowledge acquisition bottleneck	▸ informational gap problem

Motivation and contributions

TECHNISCHE UNIVERSITÄT DARMSTADT

Problem: Low accuracy of knowledge-based WSD due to informational gap

Contribution 1: Bridge the gap through distributional semantics.

Contribution 2: Bridge the gap by aligning lexical-semantic resources.

Motivation and contributions

TECHNISCHE UNIVERSITÄT DARMSTADT

Problem: Low accuracy of knowledge-based WSD due to informational gap

Contribution 1: Bridge the gap through distributional semantics.

Contribution 2: Bridge the gap by aligning lexical-semantic resources.

Problem: Sense distinctions are too subtle for accurate WSD.

Contribution 3: Use the alignments to coarsen the original sense inventory.

Motivation and contributions

TECHNISCHE UNIVERSITÄT DARMSTADT

Problem: Low accuracy of knowledge-based WSD due to informational gap

Contribution 1: Bridge the gap through distributional semantics.

Contribution 2: Bridge the gap by aligning lexical-semantic resources.

Problem: Sense distinctions are too subtle for accurate WSD.

Contribution 3: Use the alignments to coarsen the original sense inventory.

Problem: Traditional WSD is incapable of processing intentional ambiguity.

Contribution 4: Adapt WSD to puns using the above three contributions.

Agenda

TECHNISCHE UNIVERSITÄT DARMSTADT

Introduction

Knowledge-based word sense disambiguation

Pun interpretation

Conclusion

Knowledge-based disambiguation: The Lesk family of algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

Simplified Lesk: overlap between context and dictionary definitions

Knowledge-based disambiguation: The Lesk family of algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

Simplified Lesk: overlap between context and dictionary definitions

"He hit the ball with the bat."

bat
1. A small, nocturnal flying mammal of order *Chiroptera*.
2. A wooden club used to hit a ball in various sports.

Knowledge-based disambiguation: The Lesk family of algorithms

Simplified Lesk: overlap between context and dictionary definitions

"He **hit** the **ball** with the bat."

bat
1. A small, nocturnal flying mammal of order *Chiroptera*.
2. A wooden club used to **hit** a **ball** in various sports.

Knowledge-based disambiguation: The Lesk family of algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

Simplified Lesk: overlap between context and dictionary definitions

"He **hit** the **ball** with the bat."

bat
1. A small, nocturnal flying mammal of order *Chiroptera*.
2. A wooden club used to **hit** a **ball** in various sports.

Lexical gap problem: Because the context and definitions are usually quite short, it is often the case that there are no overlapping words at all.

Knowledge-based disambiguation: The Lesk family of algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

Simplified Lesk: overlap between context and dictionary definitions

"He **hit** the **ball** with the bat."

bat
1. A small, nocturnal flying mammal of order *Chiroptera*.
2. A wooden club used to **hit** a **ball** in various sports.

Lexical gap problem: Because the context and definitions are usually quite short, it is often the case that there are no overlapping words at all.

"The loan interest is paid monthly."

interest
1. A fixed charge for borrowing money.
2. A sense of concern with something.

Knowledge-based disambiguation: The Lesk family of algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

Simplified Lesk: overlap between context and dictionary definitions

"He **hit** the **ball** with the bat."

bat
1. A small, nocturnal flying mammal of order *Chiroptera*.
2. A wooden club used to **hit** a **ball** in various sports.

Lexical gap problem: Because the context and definitions are usually quite short, it is often the case that there are no overlapping words at all.

"The loan interest is paid monthly."

interest
1. A fixed charge for borrowing money.
2. A sense of concern with something.

How can we bridge the lexical gap?

Bridging the lexical gap, Solution 1: Lexical expansion

The loan <u>interest</u> is paid monthly.

interest 1. a fixed charge for borrowing money

Bridging the lexical gap, Solution 1: Lexical expansion

TECHNISCHE UNIVERSITÄT DARMSTADT

The	loan	interest	is	paid	monthly.
	mortgage			paying	annual
	loans			pay	weekly
	debt			pays	yearly
	financing			owed	quarterly
	mortgages			generated	hefty
	credit			invested	daily
	lease			spent	regular
	bond			collected	additional
	grant			raised	substantial
	funding			reimbursed	recent

interest	1.	a	fixed	charge	for	borrowing	money
			solved	charges		spending	dollars
			hefty	counts		borrow	cash
			resolved	charging		lending	funds
			monthly	cost		borrowed	billions
			additional	conviction		debt	monies
			existing	allegation		investment	millions
			reduced	pay		raising	trillions
			done	suspicion		inflows	funding
			current	count		investing	resources
			substantial	part		borrowings	donations

Bridging the lexical gap, Solution 1: Lexical expansion

The	loan	interest	is	paid	**monthly.**
	mortgage			paying	annual
	loans			**pay**	weekly
	debt			pays	yearly
	financing			owed	quarterly
	mortgages			generated	**hefty**
	credit			invested	daily
	lease			spent	regular
	bond			collected	**additional**
	grant			raised	**substantial**
	funding			reimbursed	recent

interest	1.	a	fixed	charge	for	borrowing	money
			solved	charges		spending	dollars
			hefty	counts		borrow	cash
			resolved	charging		lending	funds
			monthly	cost		borrowed	billions
			additional	conviction		**debt**	monies
			existing	allegation		investment	millions
			reduced	**pay**		raising	trillions
			done	suspicion		inflows	**funding**
			current	count		investing	resources
			substantial	part		borrowings	donations

Distributional similarity

TECHNISCHE UNIVERSITÄT DARMSTADT

- A **distributional thesaurus** (DT) provides a ranked list of similar words for every word in a lexicon

Distributional similarity

TECHNISCHE UNIVERSITÄT DARMSTADT

- A **distributional thesaurus** (DT) provides a ranked list of similar words for every word in a lexicon
- Distributional hypothesis: words that tend to appear in similar contexts have similar meanings (Firth, 1957)

Distributional similarity

- A **distributional thesaurus** (DT) provides a ranked list of similar words for every word in a lexicon
- Distributional hypothesis: words that tend to appear in similar contexts have similar meanings (Firth, 1957)
- Syntagmatic and paradigmatic relations between signs (de Saussure, 1916)

Distributional thesauri

- Heretofore used in WSD only as a heuristic ("one sense per collocation", *etc.*) or to construct (dense) vector representations (LSA, LDA, *etc.*)

Distributional thesauri

TECHNISCHE UNIVERSITÄT DARMSTADT

- Heretofore used in WSD only as a heuristic (“one sense per collocation”, *etc.*) or to construct (dense) vector representations (LSA, LDA, *etc.*)
- Advantages of DTs over dense vector representations:
  - Easy to retrieve the top *n* most similar terms
  - Sparse vectors too inefficient; dense vectors inherently lossy
  - Symbolic, interpretable representations
  - Similarity lists not polluted by infrequent terms
  - No sampling errors when representing rare topics

Construction of the distributional thesaurus

TECHNISCHE UNIVERSITÄT DARMSTADT

- 10-million-sentence, automatically parsed English news corpus
- Use collapsed dependencies to extract features for words
- Count frequency of each feature for each word
- Rank features by significance, prune to 300 per word
- Word similarity = count of common features
- Final DT contains $\geqslant 5$ similar terms for a vocabulary of over 150 000.

Excerpt of the DT entry for the noun *paper*

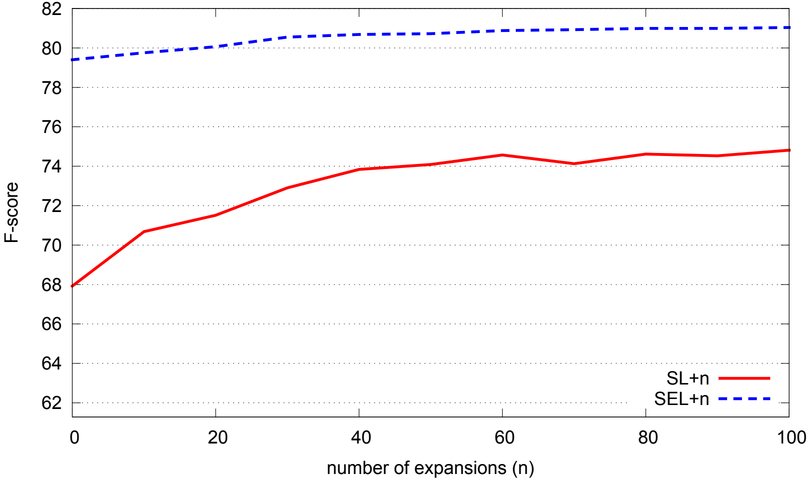
TECHNISCHE UNIVERSITÄT DARMSTADT

<table>
<tr><th>term</th><th>score</th><th>shared features</th></tr>
<tr><td>newspaper</td><td>45/300</td><td>told VBD -dobj, column NN -prep|in, local JJ amod, editor NN -poss, edition NN -prep|of, editor NN -prep|of, hometown NN nn, industry NN -nn, clips NNS -nn, shredded JJ amod, pick VB -dobj, news NNP appos, daily JJ amod, writes VBZ -nsubj, write VB -prep|for, wrote VBD -prep|for, wrote VBD -prep|in, wrapped VBN -prep|in, reading VBG -prep|in, reading VBG -dobj, read VBD -prep|in, read VBD -dobj, read VBP -prep|in, read VB -dobj, read VB -prep|in, record NN prep|of, article NN -prep|in, reports VBZ -nsubj, reported VBD -nsubj, printed VBN amod, printed VBD -nsubj, printed VBN -prep|in, published VBN -prep|in, published VBN partmod, published VBD -nsubj, sunday NNP nn, section NN -prep|of, school NN nn, saw VBD -prep|in, ad NN -prep|in, copy NN -prep|of, page NN -prep|of, pages NNS -prep|of, morning NN nn, story NN -prep|in</td></tr>
<tr><td>book</td><td>33/300</td><td>recent JJ amod, read VB -dobj, read VBD -dobj, reading VBG -dobj, edition NN -prep|of, printed VBN amod, industry NN -nn, described VBN -prep|in, writing VBG -dobj, wrote VBD -prep|in, wrote VBD rcmod, write VB -dobj, written VBN rcmod, written VBN -dobj, wrote VBD -dobj, pick VB -dobj, photo NN nn, co-author NN -prep|of, co-authored VBN -dobj, section NN -prep|of, published VBN -dobj, published VBN -nsubjpass, published VBD -dobj, published VBN partmod, copy NN -prep|of, buying VBG -dobj, buy VB -dobj, author NN -prep|of, bag NN -nn, bags NNS -nn, page NN -prep|of, pages NNS -prep|of, titled VBN partmod</td></tr>
</table>

Experiments: Overlap and use of distributional information

- Remove occurrences of the disambiguation target
- For each content word in the context and sense definition, retrieve the *n* most similar terms from the DT and add them to the text
- Separate runs for $n = 0, 10, 20, \ldots, 100$
- No lemmatization or stop word filtering
- Expanded context and definitions treated as bags of words
- Overlap is the cardinality of the intersection between the two bags of words
- Ties between senses broken by choosing randomly

WSD accuracy (F_1) on SemEval-2007 by number of expansions



SemEval-2007 accuracy by part of speech, and comparison with state of the art and baselines

	part of speech				
system	**adj.**	**noun**	**adv.**	**verb**	**all**
MFS baseline	84.25	77.44	87.50	75.30	78.89
random baseline	68.54	61.96	69.15	52.81	61.28
SL+0	75.32	69.71	69.75	59.46	67.92
SL+100	82.18	76.31	78.85	66.07	74.81
SEL+0	87.19	81.52	74.87	**72.26**	79.40
SEL+100	**88.40**	**83.45**	80.29	72.25	**81.03**
Anaya-Sánchez *et al.*, 2007	78.73	70.76	74.04	62.61	70.21
Li *et al.*, 2010	82.04	80.05	**82.21**	70.73	78.14
Ponzetto & Navigli, 2010	—	79.4	—	—	—
Chen *et al.*, 2014	—	81.6	—	—	75.8

SemEval-2007 accuracy by part of speech, and comparison with state of the art and baselines

TECHNISCHE UNIVERSITÄT DARMSTADT

system	adj.	noun	adv.	verb	all
	part of speech				
MFS baseline	84.25	77.44	87.50	75.30	78.89
random baseline	68.54	61.96	69.15	52.81	61.28
SL+0	75.32	69.71	69.75	59.46	67.92
SL+100	82.18	76.31	78.85	66.07	74.81
SEL+0	87.19	81.52	74.87	**72.26**	79.40
SEL+100	**88.40**	**83.45**	80.29	72.25	**81.03**
Anaya-Sánchez *et al.*, 2007	78.73	70.76	74.04	62.61	70.21
Li *et al.*, 2010	82.04	80.05	**82.21**	70.73	78.14
Ponzetto & Navigli, 2010	—	79.4	—	—	—
Chen *et al.*, 2014	—	81.6	—	—	75.8

SemEval-2007 accuracy by part of speech, and comparison with state of the art and baselines

	part of speech				
system	**adj.**	**noun**	**adv.**	**verb**	**all**
MFS baseline	84.25	77.44	87.50	75.30	78.89
random baseline	68.54	61.96	69.15	52.81	61.28
SL+0	75.32	69.71	69.75	59.46	67.92
SL+100	82.18	76.31	78.85	66.07	74.81
SEL+0	87.19	81.52	74.87	**72.26**	79.40
SEL+100	**88.40**	**83.45**	80.29	72.25	**81.03**
Anaya-Sánchez *et al.*, 2007	78.73	70.76	74.04	62.61	70.21
Li *et al.*, 2010	82.04	80.05	**82.21**	70.73	78.14
Ponzetto & Navigli, 2010	—	79.4	—	—	—
Chen *et al.*, 2014	—	81.6	—	—	75.8

Bridging the lexical gap, Solution 2: Enrich sense definitions by aligning complementary LSRs

Full Definition of CHILD

plural **chil·dren** \ˈchil-drən, -dərn\

1 **a** : an unborn or recently born person
b *dial* : a female infant

2 **a** : a young person especially between infancy and youth
b : a childlike or childish person
c : a person not yet of age

3 usually **childe** \ˈchī(-ə)ld\ *archaic* : a youth of noble birth

4 **a** : a son or daughter of human parents
b : DESCENDANT

child (*plural* **children** *or* (*dialectal or archaic*) **childer**)

1. A person who has not yet reached adulthood, whether natural (puberty), cultural (initiation), or legal (majority); (*obsolete*, *specifically*) a female child, a girl. [quotations ▼]
   *Go easy on him: he is but a **child**.*
2. (*with possessive*) One's son or daughter, regardless of age.
   *My youngest **child** is forty-three.*
3. (*with possessive*) One's descendants, regardless of age.
   *The **children** of Israel.*

Bridging the lexical gap, Solution 2: Enrich sense definitions by aligning complementary LSRs

Full Definition of CHILD

plural **chil·dren** \ˈchil-drən, -dərn\

1 **a** : an unborn or recently born person
b *dial* : a female infant

2 **a** : a young person especially between infancy and youth
b : a childlike or childish person
c : a person not yet of age

3 usually **childe** \ˈchī(-ə)ld\ *archaic* : a youth of noble birth

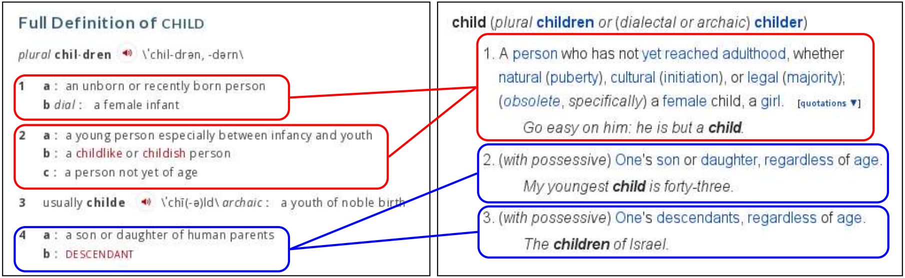
4 **a** : a son or daughter of human parents
b : DESCENDANT

child (*plural* **children** or (*dialectal or archaic*) **childer**)

1. A person who has not yet reached adulthood, whether natural (puberty), cultural (initiation), or legal (majority); (*obsolete*, *specifically*) a female child, a girl. [quotations ▼]
   *Go easy on him: he is but a **child**.*
2. (*with possessive*) One's son or daughter, regardless of age.
   *My youngest **child** is forty-three.*
3. (*with possessive*) One's descendants, regardless of age.
   *The **children** of Israel.*

Bridging the lexical gap, Solution 2: Enrich sense definitions by aligning complementary LSRs

TECHNISCHE UNIVERSITÄT DARMSTADT



Solution: Automatically merge existing pairwise alignments

Step A: Collect pairwise alignments

TECHNISCHE UNIVERSITÄT DARMSTADT

Wiktionary to WordNet (Meyer & Gurevych, 2011)	**Wikipedia to WordNet** (Matuschek & Gurevych, 2013)
3198:0:2 ⟷ 09828216n	Child ⟷ 09918248n
3198:0:3 ⟷ 01322221n	Child ⟷ 09918554n
3198:0:3 ⟷ 09918554n	Child ⟷ 09918762n
4487:0:1 ⟷ 09918248n	Infant ⟷ 09827363n
4487:0:2 ⟷ 09918762n	Infant ⟷ 09827519n
	Infant ⟷ 09827683n
	Infant ⟷ 09828216n
	Infant ⟷ 10353016n

Step B: Build a graph of aligned senses

01322221n

09827363n

09827519n

3198:0:2

09827683n

3198:0:3

Child

09828216n

4487:0:1

Infant

09918248n

4487:0:2

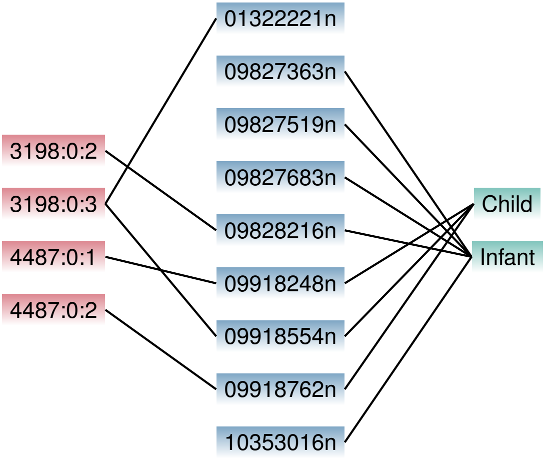
09918554n

09918762n

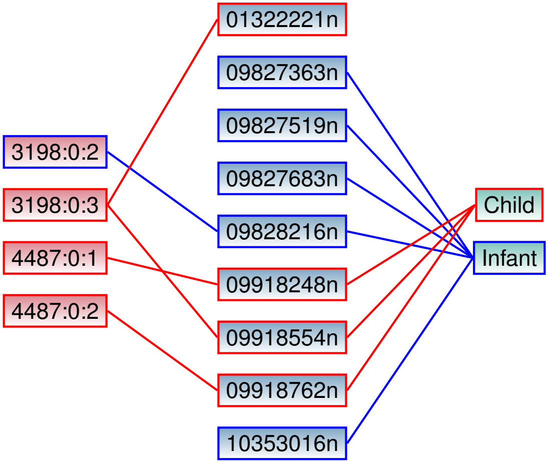
10353016n

Step B: Build a graph of aligned senses

TECHNISCHE UNIVERSITÄT DARMSTADT



Step C: Find connected components to cluster word senses



Results on Senseval-3 all-words data set

TECHNISCHE UNIVERSITÄT DARMSTADT

glosses	coverage	precision	recall	F-score
SL+0	26.85	**69.23**	18.59	29.30
SL+0 with enriched glosses	**29.17**	67.26	**19.62**	**30.38**
SEL+30	98.61	**53.46**	**52.71**	**53.08**
SEL+30 with enriched glosses	**98.76**	51.07	50.44	50.75

- Improvement to simplified Lesk is modest but statistically significant (McNemar's $\chi^2 = 6.22$, df $= 1$, $\chi^2_{1,0.95} = 3.84$)
- Method is not compatible with the lexical expansion method, particularly for rarer and more polysemous words

Agenda

Introduction

Knowledge-based word sense disambiguation

Pun interpretation

Conclusion

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Traditional WSD assumes every word carries a single meaning
- In punning, words are used in a deliberately ambiguous manner:

Motivation

- Traditional WSD assumes every word carries a single meaning
- In punning, words are used in a deliberately ambiguous manner:

The electric company to a customer:
"We would be delighted if you send in your bill.
However, if you don't, you will be."
(Aarons, 2012)

Motivation

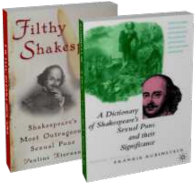
TECHNISCHE UNIVERSITÄT DARMSTADT

- Traditional WSD assumes every word carries a single meaning
- In punning, words are used in a deliberately ambiguous manner:

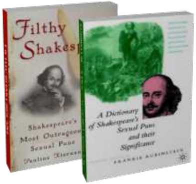
The electric company to a customer:
“We would be **delighted** *if you send in your bill.*
However, if you don’t, you will be.”
(Aarons, 2012)

Motivation

Motivation


Filthy
A Dictionary of Shakespeare's Sexual Puns and their Significance

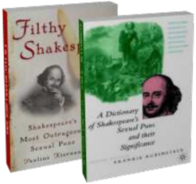
Motivation



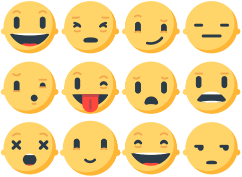
A Dictionary
of Shakespeare's
Sexual Puns
and their
Significance

HAL ROACH presents STAN
LAUREL
OLIVER
HARDY
A
Metro-
Goldwyn
Mayer
ALL
TALKING
COMEDY
ANOTHER FINE MESS

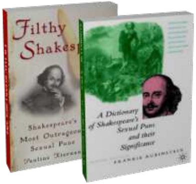
Motivation


Filthy
Shakes
A Dictionary
of Shakespeare's
Sexual Puns
and their
Significance


HAL ROACH presents STAN
LAUREL
OLIVER
HARDY
Metro-
Goldwyn
Mayer
ALL
TALKING
COMEDY
ANOTHER FINE
MESS

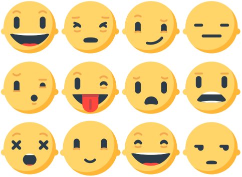


Motivation



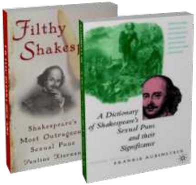
Digital humanities

Machine(-assisted) translation



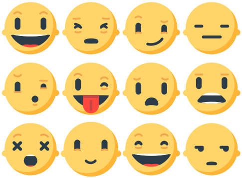
Sentiment analysis

Motivation



Digital humanities

Machine(-assisted) translation

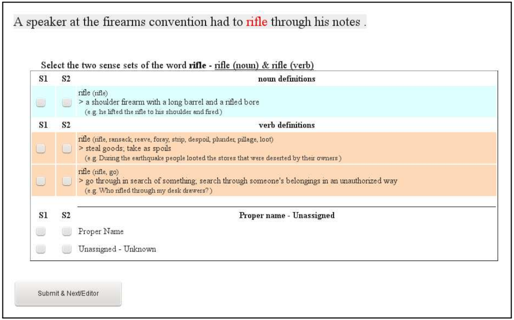


Sentiment analysis

Human–computer interaction

Data set

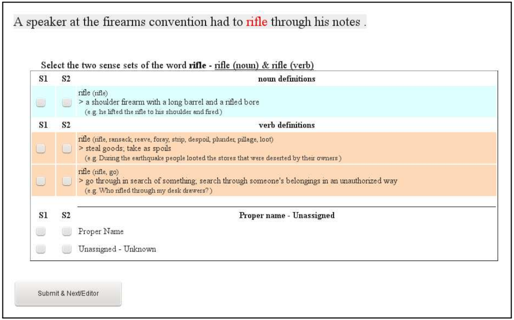
- 1607 short punning jokes, each with one homographic pun
- WordNet 3.1 annotations applied by three human judges



Data set

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- 1607 short punning jokes, each with one homographic pun
- WordNet 3.1 annotations applied by three human judges
- Good interannotator agreement (Krippendorff's $\alpha = 0.777$)
- Pun senses often transcend part of speech



A speaker at the firearms convention had to rifle through his notes .

Select the two sense sets of the word **rifle** - rifle (noun) & rifle (verb)

S1	S2	noun definitions
		rifle (rifle) > a shoulder firearm with a long barrel and a rifled bore (e.g. he lifted the rifle to his shoulder and fired)
S1	**S2**	**verb definitions**
		rifle (rifle, ransack, reave, foray, strip, despoil, plunder, pillage, loot) > steal goods; take as spoils (e.g. During the earthquake people looted the stores that were deserted by their owners)
		rifle (rifle, go) > go through in search of something; search through someone's belongings in an unauthorized way (e.g. Who rifled through my desk drawers?)
S1	**S2**	**Proper name - Unassigned**
		Proper Name
		Unassigned - Unknown

Submit & Next/Editor

Algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

- Lack of training data rules out supervised approaches
- Naïve adaptations of SL, SEL, SL+100, and random/MFS baselines (select the top *two* senses returned by the algorithm)

Algorithms

- Lack of training data rules out supervised approaches
- Naïve adaptations of SL, SEL, SL+100, and random/MFS baselines (select the top *two* senses returned by the algorithm)
- Pun-specific adaptations of SEL:

  POS: Favour senses that match the pun's putative POS

Algorithms

TECHNISCHE UNIVERSITÄT DARMSTADT

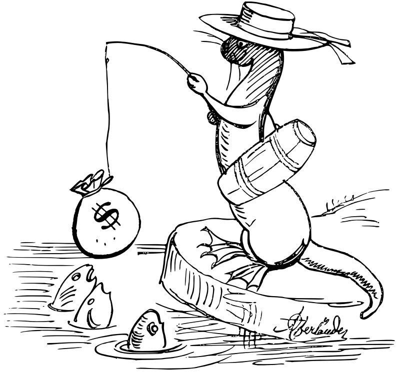
- Lack of training data rules out supervised approaches
- Naïve adaptations of SL, SEL, SL+100, and random/MFS baselines (select the top *two* senses returned by the algorithm)
- Pun-specific adaptations of SEL:

  POS: Favour senses that match the pun's putative POS

  cluster: Also ensure the two senses are in different clusters
- We tested our 3-way WordNet–Wikipedia–Wiktionary clustering as well as a 2-way WordNet–OmegaWiki clustering (Matuschek *et al.*, 2014)

Using LSR alignments for sense clustering

Where do otters keep their money? At the **bank**!



Using LSR alignments for sense clustering

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Where do otters keep their money? At the **bank**!

Senses

09213565n	sloping land (especially the slope beside...
09213434n	a long ridge or pile
08462066n	an arrangement of similar objects in a row...
08420278n	a financial institution that accepts deposits...
02787772n	a building in which the business of banking...
00169305n	a flight maneuver; aircraft tips laterally...

Using LSR alignments for sense clustering

TECHNISCHE UNIVERSITÄT DARMSTADT

Where do otters keep their money? At the **bank**!

Scores	Senses	
5	09213565n	sloping land (especially the slope beside...
2	09213434n	a long ridge or pile
1	08462066n	an arrangement of similar objects in a row...
7	08420278n	a financial institution that accepts deposits...
5	02787772n	a building in which the business of banking...
0	00169305n	a flight maneuver; aircraft tips laterally...

Using LSR alignments for sense clustering

TECHNISCHE UNIVERSITÄT DARMSTADT

Where do otters keep their money? At the **bank**!

Scores	Senses	
5	09213565n	sloping land (especially the slope beside...
2	09213434n	a long ridge or pile
1	08462066n	an arrangement of similar objects in a row...
7	08420278n	a financial institution that accepts deposits...
5	02787772n	a building in which the business of banking...
0	00169305n	a flight maneuver; aircraft tips laterally...

Results

TECHNISCHE UNIVERSITÄT DARMSTADT

System	coverage	precision	recall	F-score
SL+0	35.52	19.74	7.01	10.35
SEL+0	42.45	19.96	8.47	11.90
SL+100	**98.69**	13.43	13.25	13.34
SEL+POS	59.94	**21.21**	12.71	15.90
SEL+cluster$_3$	66.33	20.67	13.71	16.49
SEL+cluster$_2$	68.10	20.70	**14.10**	**16.77**
random	100.00	9.31	9.31	9.31
MFS	100.00	13.25	13.25	13.25

Results

TECHNISCHE UNIVERSITÄT DARMSTADT

System	coverage	precision	recall	F-score
SL+0	35.52	19.74	7.01	10.35
SEL+0	42.45	19.96	8.47	11.90
SL+100	**98.69**	13.43	13.25	13.34
SEL+POS	59.94	**21.21**	12.71	15.90
SEL+cluster$_3$	66.33	20.67	13.71	16.49
SEL+cluster$_2$	68.10	20.70	**14.10**	**16.77**
random	100.00	9.31	9.31	9.31
MFS	100.00	13.25	13.25	13.25

- Pun "disambiguation" is *much* harder than traditional WSD

Results

TECHNISCHE UNIVERSITÄT DARMSTADT

System	coverage	precision	recall	F-score
SL+0	35.52	19.74	7.01	10.35
SEL+0	42.45	19.96	8.47	11.90
SL+100	**98.69**	13.43	13.25	13.34
SEL+POS	59.94	**21.21**	12.71	15.90
SEL+cluster$_3$	66.33	20.67	13.71	16.49
SEL+cluster$_2$	68.10	20.70	**14.10**	**16.77**
random	100.00	9.31	9.31	9.31
MFS	100.00	13.25	13.25	13.25

- Pun “disambiguation” is *much* harder than traditional WSD

Results

TECHNISCHE UNIVERSITÄT DARMSTADT

System	coverage	precision	recall	F-score
SL+0	35.52	19.74	7.01	10.35
SEL+0	42.45	19.96	8.47	11.90
SL+100	**98.69**	13.43	13.25	13.34
SEL+POS	59.94	**21.21**	12.71	15.90
SEL+cluster$_3$	66.33	20.67	13.71	16.49
SEL+cluster$_2$	68.10	20.70	**14.10**	**16.77**
random	100.00	9.31	9.31	9.31
MFS	100.00	13.25	13.25	13.25

- Pun “disambiguation” is *much* harder than traditional WSD
- Pun-adapted SEL as good as supervised baseline(!)

Future directions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Work so far assumes that:
  - Location of the pun is given
  - Pun is homographic (“perfect”)

Future directions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Work so far assumes that:
  - Location of the pun is given
  - Pun is homographic (“perfect”)
- Further research problems:

Future directions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Work so far assumes that:
  - Location of the pun is given
  - Pun is homographic (“perfect”)
- Further research problems:
  - Pun detection

Future directions

TECHNISCHE UNIVERSITÄT DARMSTADT

- Work so far assumes that:
  - Location of the pun is given
  - Pun is homographic (“perfect”)
- Further research problems:
  - Pun detection
  - Processing of imperfect puns

Pun typology

TECHNISCHE UNIVERSITÄT DARMSTADT

	homophonic	**heterophonic**
homographic	A political prisoner is one who stands behind her *convictions*.	A lumberjack's world revolves on its *axes*.
heterographic	She fell through the window but felt no *pane*.	The sign at the nudist camp read, "*Clothed* until April."

Sound similarity

- Any pair of words can be characterized by their (perceived) similarity in terms of sound or pronunciation.
- Studying pairs with a phonologically constrained relationship can help us model that relationship.
- Conversely, a model that quantifies perceived sound differences between words can assess the probability of a given relationship.
- In particular, a model of sound similarity could help detect or generate puns.

Early similarity models

TECHNISCHE UNIVERSITÄT DARMSTADT

- “Predicted phonetic distance” or “PPD” (Vitz & Winkler, 1973)
  1. Optimally align two phonemic sequences
  2. Compute the relative Hamming distance (i.e., the proportion of non-matching phoneme positions)

Early similarity models

- "Predicted phonetic distance" or "PPD" (Vitz & Winkler, 1973)
  1. Optimally align two phonemic sequences
  2. Compute the relative Hamming distance (i.e., the proportion of non-matching phoneme positions)

#	∅	∅	∅	∅	∅	ɹ	ə	l	e	ʃ	n	#	*relation*
#	ʌ	n	d	ə	ɹ	ɹ	ɪ	∅	∅	t	n	#	*underwritten*

$$\text{PPD} = 9 \div 11 \approx 0.818$$

Early similarity models

TECHNISCHE UNIVERSITÄT DARMSTADT

- "Predicted phonetic distance" or "PPD" (Vitz & Winkler, 1973)
  1. Optimally align two phonemic sequences
  2. Compute the relative Hamming distance (i.e., the proportion of non-matching phoneme positions)

#	∅	∅	∅	∅	∅	ɹ	ə	l	e	ʃ	n	#	*relation*
#	ʌ	n	d	ə	ɹ	ɹ	ɪ	∅	∅	t	n	#	*underwritten*

$$\text{PPD} = 9 \div 11 \approx 0.818$$

- Method works better when it is applied separately to the syllable onset, nucleus, and coda.

Early similarity models

- “Predicted phonetic distance” or “PPD” (Vitz & Winkler, 1973)
  1. Optimally align two phonemic sequences
  2. Compute the relative Hamming distance (i.e., the proportion of non-matching phoneme positions)

#	∅	∅	∅	∅	∅	ɹ	ə	l	e	ʃ	n	#	*relation*
#	ʌ	n	d	ə	ɹ	ɹ	ɪ	∅	∅	t	n	#	*underwritten*

$$\text{PPD} = 9 \div 11 \approx 0.818$$

- Method works better when it is applied separately to the syllable onset, nucleus, and coda.
- Aligning the sequences is a nontrivial task.

Sound similarity based on phonemic features

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Many models compute similarity in terms of the classic feature matrix (Chomsky & Halle, 1968).

Sound similarity based on phonemic features

TECHNISCHE UNIVERSITÄT DARMSTADT

- Many models compute similarity in terms of the classic feature matrix (Chomsky & Halle, 1968).
- These models often fail to account for many common cases.

Sound similarity based on phonemic features

TECHNISCHE UNIVERSITÄT DARMSTADT

- Many models compute similarity in terms of the classic feature matrix (Chomsky & Halle, 1968).
- These models often fail to account for many common cases.

  Trying to preserve his savoir faire in a new restaurant, the guest looked down at the eggs the waiter had spilled in his lap and said brightly, "Well, I guess the yolk's on me!"

Sound similarity based on phonemic features

TECHNISCHE UNIVERSITÄT DARMSTADT

- Many models compute similarity in terms of the classic feature matrix (Chomsky & Halle, 1968).
- These models often fail to account for many common cases.

  > *Trying to preserve his savoir faire in a new restaurant, the guest looked down at the eggs the waiter had spilled in his lap and said brightly, "Well, I guess the yolk's on me!"*

- Variously mitigated by the use of multivalued features (Ladefoged, 1995), feature salience coefficients (Kondrak, 2002), and Optimality Theory (Lutz & Greene, 2003).

Similarity models based on puns

- Hausmann (1974) observed an absolute phonemic distance of no more than four

Similarity models based on puns

TECHNISCHE UNIVERSITÄT DARMSTADT

- Hausmann (1974) observed an absolute phonemic distance of no more than four
- Lagerquist (1980): puns tend not to insert or delete syllables, nor to change syllable stress; sound changes tend to occur on the stressed syllable

Similarity models based on puns

- Hausmann (1974) observed an absolute phonemic distance of no more than four
- Lagerquist (1980): puns tend not to insert or delete syllables, nor to change syllable stress; sound changes tend to occur on the stressed syllable
- Zwicky & Zwicky (1986): certain segments do not appear equally often in puns and targets: Y “ousts” X when Y appears as a pun substitute for the latent target X significantly more often than the reverse.

Similarity models based on puns

- Hausmann (1974) observed an absolute phonemic distance of no more than four
- Lagerquist (1980): puns tend not to insert or delete syllables, nor to change syllable stress; sound changes tend to occur on the stressed syllable
- Zwicky & Zwicky (1986): certain segments do not appear equally often in puns and targets: Y "ousts" X when Y appears as a pun substitute for the latent target X significantly more often than the reverse.
- Sobkowiak (1991): pun understandability is maximized when the consonantal skeleton is kept largely intact

How to proceed?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Past phonological analyses tend to agree

How to proceed?

TECHNISCHE UNIVERSITÄT DARMSTADT

- Past phonological analyses tend to agree
- How to consolidate and implement them computationally?

How to proceed?

- Past phonological analyses tend to agree
- How to consolidate and implement them computationally?
- Hempelmann, 2003 modelled Sobkowiak's data into a cost function that could conceivably be used in a pun generator or detector, but this has not yet been done

How to proceed?

- Past phonological analyses tend to agree
- How to consolidate and implement them computationally?
- Hempelmann, 2003 modelled Sobkowiak's data into a cost function that could conceivably be used in a pun generator or detector, but this has not yet been done
- Shared task on pun detection and interpretation

Agenda

TECHNISCHE UNIVERSITÄT DARMSTADT

Introduction

Knowledge-based word sense disambiguation

Pun interpretation

Conclusion

Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

This talk has presented…

- …a method for applying lexical expansions to knowledge-based WSD, resulting in state-of-the-art performance

Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

This talk has presented. . .

- . . . a method for applying lexical expansions to knowledge-based WSD, resulting in state-of-the-art performance
- . . . a method for combining arbitrary pairwise alignments of lexical-semantic resources, useful for:
  - (slightly) improving the accuracy of knowledge-based WSD
  - inducing a clustering of senses in LSRs

Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

This talk has presented. . .

- . . . a method for applying lexical expansions to knowledge-based WSD, resulting in state-of-the-art performance
- . . . a method for combining arbitrary pairwise alignments of lexical-semantic resources, useful for:
  - (slightly) improving the accuracy of knowledge-based WSD
  - inducing a clustering of senses in LSRs
- . . . a sense-annotated data set for puns, and pioneering algorithms for “disambiguating” them
- . . . some background on the phonology of imperfect puns and ideas for implementing them computationally.

Summary

TECHNISCHE
UNIVERSITÄT
DARMSTADT

This talk has presented…

- …a method for applying lexical expansions to knowledge-based WSD, resulting in state-of-the-art performance
- …a method for combining arbitrary pairwise alignments of lexical-semantic resources, useful for:
  - (slightly) improving the accuracy of knowledge-based WSD
  - inducing a clustering of senses in LSRs
- …a sense-annotated data set for puns, and pioneering algorithms for "disambiguating" them
- …some background on the phonology of imperfect puns and ideas for implementing them computationally.

References I

- Aarons, D. (2012). *Jokes and the Linguistic Mind.* Routledge.
- Aarons, D. (2017). Puns and Tacit Linguistic Knowledge. In S. Attardo, ed., *Handbook of Language and Humor.* Routledge, pp. 80–94.
- Anaya-Sánchez, H., A. Pons-Porrata, and R. Berlanga-Llavori (2007). TKB-UO: Using Sense Clustering for WSD. In: *SemEval 2007: Proceedings of the 4th International Workshop on Semantic Evaluations*, pp. 322–325.
- Biemann, C. and M. Riedl (2013). Text: Now in 2D! A Framework for Lexical Expansion with Contextual Similarity. *Journal of Language Modelling* 1(1), pp. 55–95.
- Chen, X., Z. Liu, and M. Sun (2014). A Unified Model for Word Sense Representation and Disambiguation. In: *The 2014 Conference on Empirical Methods in Natural Language Processing: Proceedings of the Conference*, pp. 1025–1035.
- Cholakov, K., C. Biemann, J. Eckle-Kohler, and I. Gurevych (2014). Lexical Substitution Dataset for German. In *LREC 2014, Ninth International Conference on Language Resources and Evaluation*, pp. 2524–2531.

References II

TECHNISCHE UNIVERSITÄT DARMSTADT

- Firth, J. R. (1957). A Synopsis of Linguistic Theory, 1930–1955. In: *Studies in Linguistic Analysis*. Basil Blackwell, pp. 1–32.
- Hempelmann, C. and T. Miller (2017). Puns: Taxonomy and Phonology. In S. Attardo, ed., *Handbook of Language and Humor.* Routledge, pp. 95–108.
- Li, L., B. Roth, and C. Sporleder (2010). Topic Models for Word Sense Disambiguation and Token-based Idiom Detection. In: *48th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pp. 1138–1147.
- Matuschek, M. and I. Gurevych (2013). Dijkstra-WSA: A Graph-based Approach to Word Sense Alignment. In: *Transactions of the Association for Computational Linguistics* 1, pp. 151–164.
- Matuschek, M., T. Miller, and I. Gurevych (2014). A Language-independent Sense Clustering Approach for Enhanced WSD. In: *Proceedings of the 12th Edition of the KONVENS Conference*, pp. 11–21.
- Meyer, C. M. and I. Gurevych (2011). What Psycholinguists Know About Chemistry: Aligning Wiktionary and WordNet for Increased Domain Coverage. In: *Proceedings of the Fifth International Joint Conference on Natural Language Processing*, pp. 883–892.

References III

- Ponzetto, S. P. and R. Navigli (2010). Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems. In: *48th Annual Meeting of the Association for Computational Linguistics: Proceedings of the Conference*, pp. 1522–1531.
- Raskin, V. (1985). *Semantic Mechanisms of Humor.* Springer.

Credits

TECHNISCHE
UNIVERSITÄT
DARMSTADT

- TU Darmstadt S103 ErhoehtVonS208 ©2007 ThomasGP. CC BY-SA 4.0.
- Robert-Piloty-Gebäude, TU Darmstadt ©2006 S. Kasten. CC BY-SA 4.0.
- Darmstadt 2006 121 ©2006 derbrauni. CC BY-SA 4.0.
- Darmstadt TU 1 ©2011 Andreas Pfaefcke. CC BY 3.0.
- University College Front Facade ©2004 Nuthingoldstays. CC BY-SA 3.0.
- First Nations University 3 ©2013 . CC BY-SA 3.0.
- Woman and laptop ©2012 Shopware. CC BY-SA 3.0.
- Firefox OS Emjois ©2015 Mozilla Foundation. CC BY 4.0.
- Duck, Rabbit! Duck! ©2015 Taro Istok. CC BY-SA 4.0.

Thank you!

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Questions?