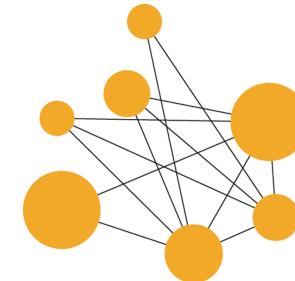
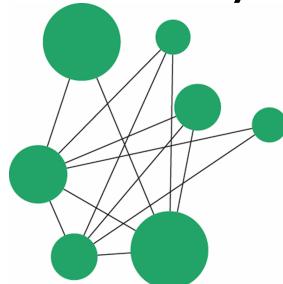




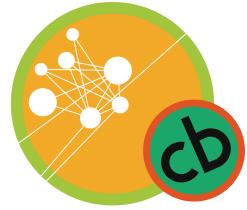
Tutorial: Reviewing Introduction sections

Kevin Bretonnel Cohen

Director, Biomedical Text Mining Group,
University of Colorado School of Medicine;
D'Alembert Chair in Natural Language
Processing for the Biomedical Domain,
Université Paris-Saclay



kevin.cohen@gmail.com
http://compbio.ucdenver.edu/Hunter_lab/Cohen



Karën Fort, Margot Mieskes, and Aurélie Névéol

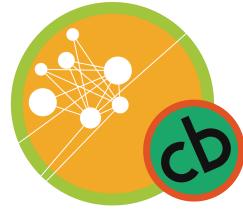


Kevin Bretonnel Cohen,
UCSOM

Karën Fort,
Sorbonne
Université /
Loria

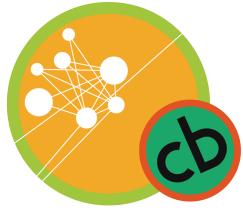
Margot Mieskes,
h_da Darmstadt

Aurélie Névéol,
Université
Paris Saclay,
CNRS, LIMSI

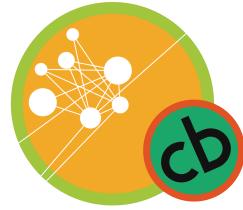


For in-depth discussion of all of the topics in this tutorial, see:

**COHEN, FORT, NÉVÉOL,
AND MIESKES (2020) HOW
TO REVIEW A PAPER
ABOUT NATURAL
LANGUAGE PROCESSING**



**DISCLAIMER: I HAVE MADE
EVERY MISTAKE THAT I
WILL DESCRIBE TODAY**



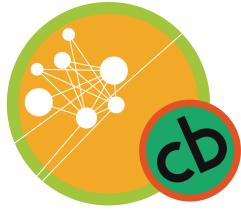
Sins I have committed

...of writing

1. Boring introductions
2. Duplicate publication of results
3. Massive lack of details about methodology
4. Completely inappropriate data
5. Over-extrapolation of conclusions

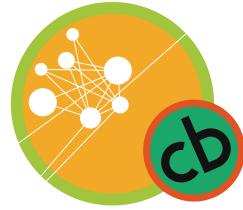
...of reviewing

1. Late without warning editor
2. High recommendation score without details in written comments
3. Not responding to emails
4. Getting irritated
5. Saying mean things



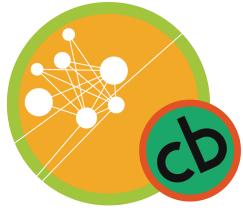
We are talking about the Introduction section because...

**THE INTRODUCTION IS
THE MOST-HEAVILY
WEIGHTED PART OF THE
PAPER IN THE
QUANTITATIVE SCORES**

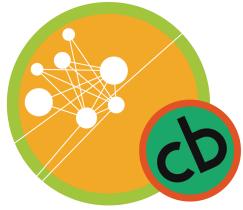


What is the function of the *Introduction* section?

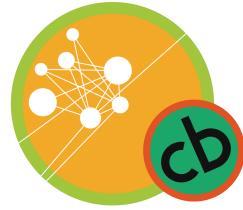
**THE INTRODUCTION IS
WHERE THE PAPER
CONVINCES YOU THAT IT
WAS WORTH WRITING**



PRIMARY REVIEWING QUESTION #1: IS THE WORK SIGNIFICANT?

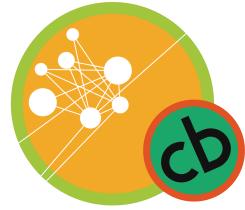


PRIMARY REVIEWING QUESTION #2: IS THE WORK INNOVATIVE?



Quantitative criteria based (primarily) on the *Introduction* section

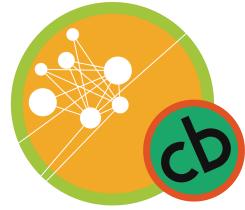
1. How significant is the paper?
2. How innovative is the paper?
3. How good is the discussion of related work?
4. How clear is the paper?



It is the paper's job to convince the reader that the work is significant and innovative. It does this via discussion of related research.

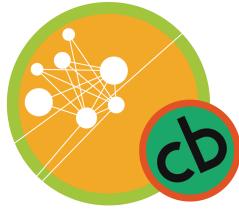
**NICE DRAFT. NOW HOW
ABOUT ADDING SOME
CITATIONS SO WE KNOW
THAT YOU'RE NOT JUST
MAKING ALL THIS UP?**

--Pierre Zweigenbaum, to me



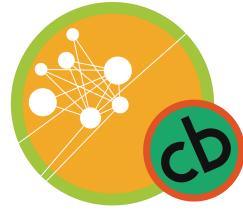
How to review the Introduction: Look for the rhetorical structure

- What the paper is about
- Why that is important
- What we are doing
- What is new about that



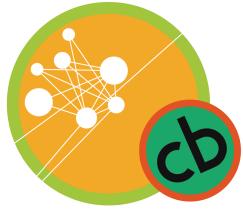
Look for the rhetorical structure **and explanations**

- What the paper is about
 - *The work reported here tackles the problem of -----, which is [DEFINITION].*
- Why that is important
 - *That is an important problem because -----.*
- What we are doing
 - *Here the problem is approached by -----, which is [DESCRIPTION].*
- What is new about that
 - *Previous approaches have -----, which is good, but have not -----, which is a missed opportunity because -----.*



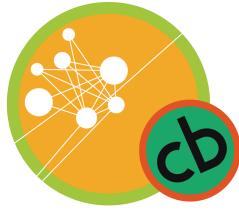
Look for the rhetorical structure, and explanations **supported by citations**

- What the paper is about
 - *The work reported here tackles the problem of -----, which is [DEFINITION] [CITATION].*
- Why that is important
 - *That is an important problem because ----- [CITATION, CITATION, CITATION].*
- What we are doing
 - *Here the problem is approached by -----, which is [DESCRIPTION] [CITATION].*
- What is new about that
 - *Previous approaches have -----, which is good [CITATION, CITATION, CITATION], but have not -----, which is a missed opportunity because ----- [CITATION, CITATION, CITATION].*



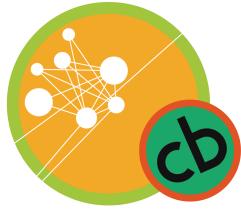
Exercise: Which paper would you expect to get a higher score for significance? Why?

1. Researchers have been interested in information extraction since the wide-spread recognition that true language understanding by machine is not achievable. The state of the art is an F-measure in the low 0.90s.
2. True language understanding by machines would be an amazing triumph for artificial intelligence, but because it is not achievable yet, information extraction has emerged as a challenging and worthwhile intermediate goal.



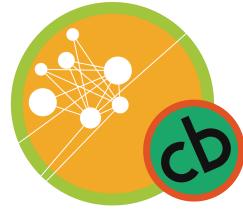
Exercise: Which paper would you expect to get a higher score for discussion of related work? Why?

1. Researchers have been interested in information extraction since the wide-spread recognition that true language understanding by machine is not achievable [1,2,3]. The state of the art is an F-measure in the low 0.90s [4,5,6].
2. True language understanding by machines would be an amazing triumph for artificial intelligence, but because it is not achievable yet, information extraction has emerged as a challenging and worthwhile intermediate goal [7].
1. A major textbook of NLP
2. An influential whitepaper
3. A *Last Words* column by a past president of the ACL
4. A recent paper
5. A review article in *Computational Linguistics*
6. An update on the industry from *J. Nat. Lang. Engineering*
7. An unreviewed preprint on arXiv

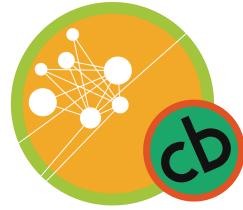


Exercise: which paper would you give a higher score for clarity?
Why?

1. Loss functions are an important part of neural networks [citation].
2. Loss functions, defined as ---, are an important part of neural networks [citation].
3. Loss functions are an important part of neural networks because --- [citation].

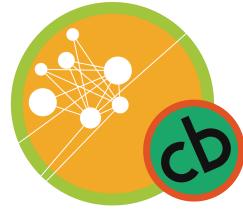


**THERE ARE MANY WAYS
TO BE SIGNIFICANT**

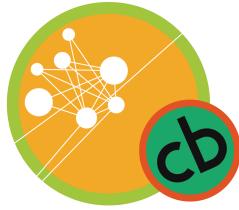


From Anna Rogers's list of 10 ways to be significant

1. Computational efficiency
 2. Data efficiency
 3. Stable over large range of hyperparameters and/or random initializations
 4. Generalizability (less biased, able to avoid learning from data artifacts, better generalizing across datasets and domains, more adversarially robust);
 5. Interpretability
 6. Simplicity
 7. Cognitive plausibility
- <https://hackingsemantics.xyz/2020/reviewing-models/#solution-guidelines-on-what-constitutes-an-acceptable-contribution>

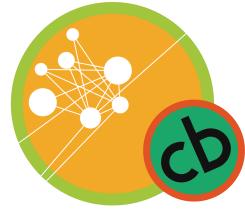


**THERE ARE MANY WAYS
TO BE INNOVATIVE**

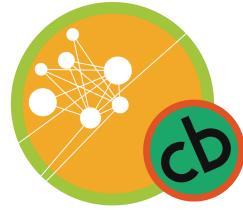


1. New problem
2. New task
3. New domain
4. Technique/model from another field
5. Taking an unpopular position
6. Addressing confounds in previous work

Cohen, Kevin Bretonnel (2021) *Writing about data science research-with examples from machine learning and natural language processing*. Cambridge University Press

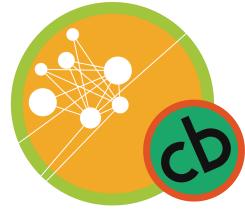


**WHAT HAVEN'T I TALKED
ABOUT?**

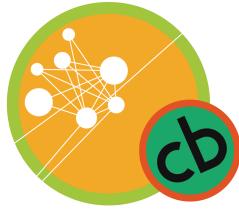


Source: Kenneth Church

**THE BETTER THE
NUMBERS ARE, THE MORE
IMPORTANT IT IS TO
REJECT THE PAPER. WE
CAN'T AFFORD PAPERS
THAT REPORT RESULTS
WITHOUT INSIGHTS.**

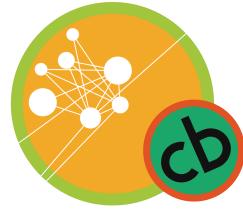


**RELATED WORK CAN BE
DISCUSSED WITHOUT A
“RELATED WORK”
SECTION**



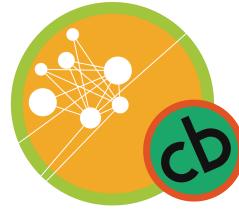
In contrast with Muggleton (1994) and Shavlik (1990) our method is empirical and does not use an explanation or proof structure. It also differs from the empirical method of Zelle and Mooney (1994) in the use of the syntactic restrictions. While the scope of our predicates is more limited we are able to guarantee correctness and efficiency.

Khardon, Roni. "Learning to take actions." *Machine Learning* 35.1 (1999): 57-90.



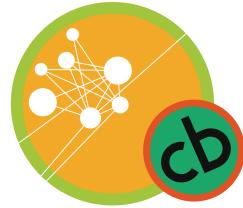
The bad news:

**THERE IS A LIMITED
NUMBER OF WAYS TO BE
CLEAR**



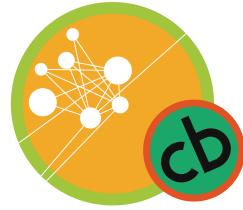
You can help improve clarity

- Definitions:
 - -: *The paper does not define some of its important terms, such as [specific examples]. Its clarity could be improved by giving definitions.*
 - +: *A real strength of the paper is that it defines important terms, which increases its clarity.*
- Explanations:
 - -: *The paper lacks explanations of why [the data set was used/the methods were used/the results were interpreted as they were]. I noticed that words like “because, since, due to” occurred only n times. Its clarity could be improved by explaining the rationale behind [the unexplained points].*
 - +: *A real strength of the paper is that it explains [why the data set was used/the methods were used/the results were interpreted as they were].*



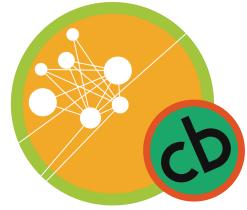
Source: Kenneth Church

**THE BETTER THE
NUMBERS ARE, THE MORE
IMPORTANT IT IS TO
REJECT THE PAPER. WE
CAN'T AFFORD PAPERS
THAT REPORT RESULTS
WITHOUT INSIGHTS.**



For in-depth discussion of all of the topics in this tutorial, see:

**COHEN, FORT, NÉVÉOL,
AND MIESKES (2020) HOW
TO REVIEW A PAPER
ABOUT NATURAL
LANGUAGE PROCESSING**



Karën Fort, Margot Mieskes, and Aurélie Névéol



Kevin Bretonnel Cohen,
UCSOM

Karën Fort,
Sorbonne
Université /
Loria

Margot Mieskes,
h_da Darmstadt

Aurélie Névéol,
Université
Paris Saclay,
CNRS, LIMSI