Finding Keystone Citations for Constructing Validity Chains among Research Papers

Yuanxi Fu[1], Jodi Schneider[1] and Catherine Blake[1,2]

School of Information Sciences[1]

Department of Computer Science[2]

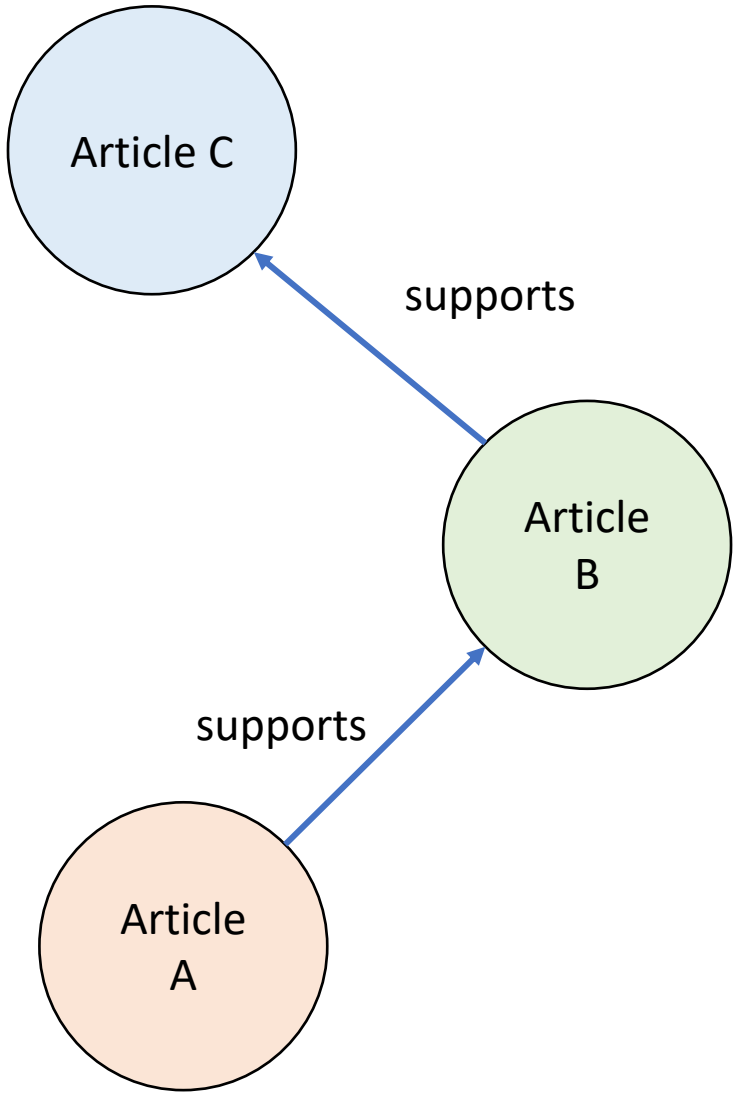
University of Illinois at Urbana-Champaign

Presentation for Sci-K 2021, 2021-04-13

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

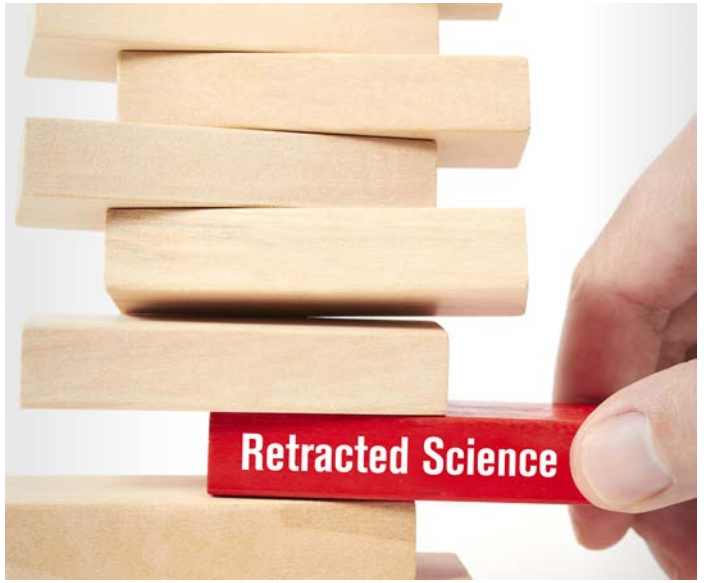
Keystone Citations

- Gilbert (1977): “References are used to provide authoritative grounds to persuade readers of the validity and significance of the argument in the paper.”
- Argued against by Small (1978), who thinks that “we should not give a causal explanation why author-scientists cite certain papers or not.”
- We developed a framework to identify citations that can make or break a paper’s arguments (Fu & Schneider, 2020).



Motivations

- When DOES the citation matter to the validity of the scientific arguments of a paper?
- Retracted papers kept being cited, predominantly, as valid work (Kochan & Budd, 1992, Budd et al., 1999, Budd et al., 2011, Bar-Ilan & Halevi, 2017, Hamilton, 2019, Bolboacă et al.,2019, Schneider et al., 2020). So, under what circumstances should papers citing retracted work also be examined too?



The Keystone Framework

- Combine **argument-based document models** and **citation context analysis.**
- Guide users through a process to find citations that are a “keystone” to the citing paper’s arguments.

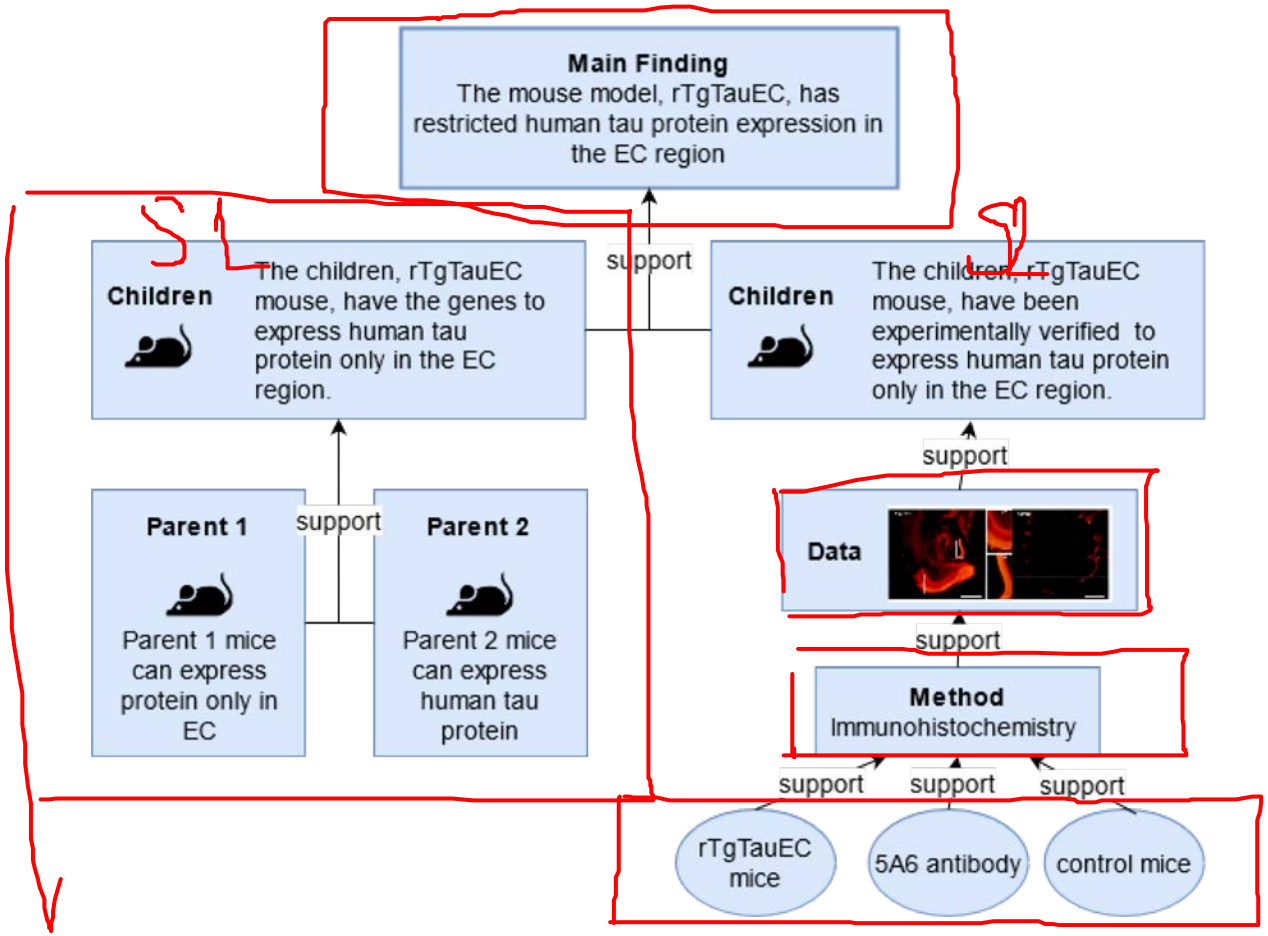
Document Modeling

Citation Context Analysis

Classification

Modeling the Argument

In our JCDL paper, we modeled the argument supporting a main finding in an Alzheimer research paper (de Calignon et al., 2011) using the Micropublication Ontology (Clark et al., 2014).



(Fu & Schneider, 2020)

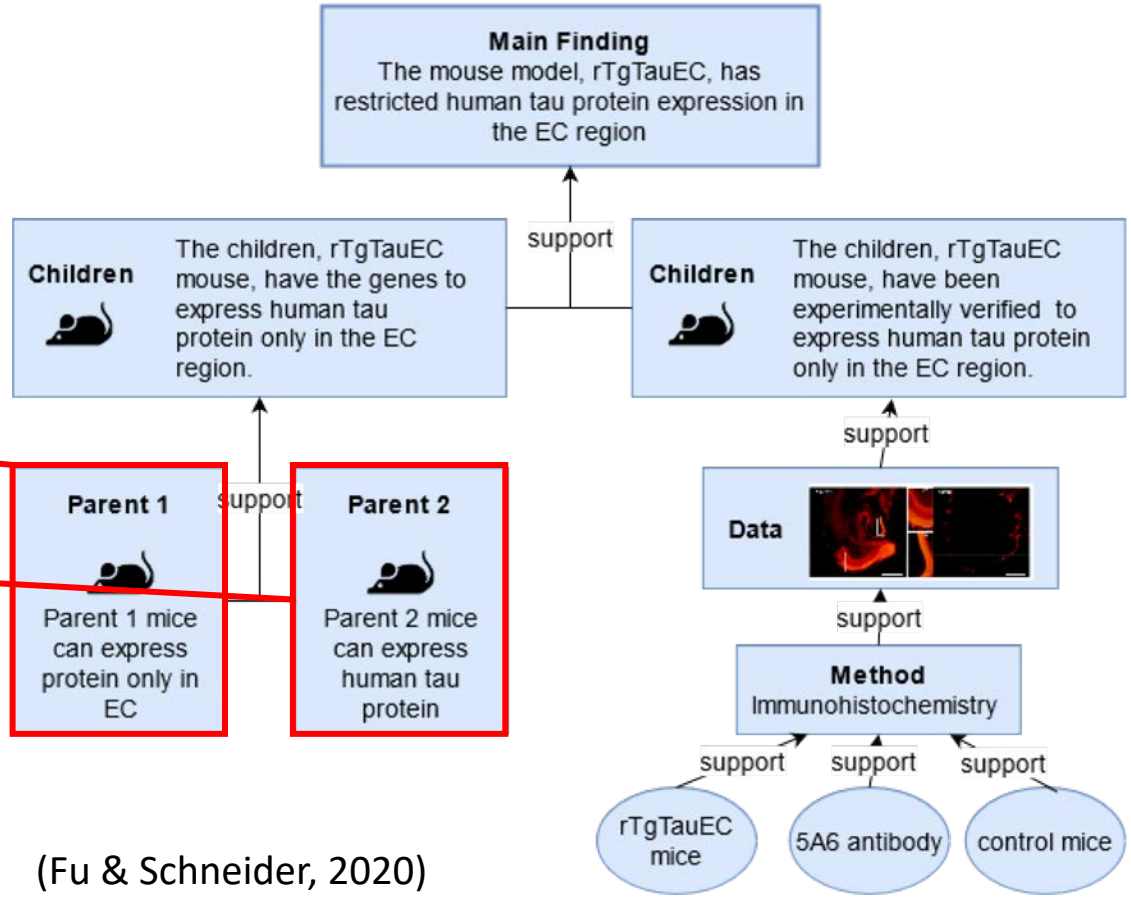
Citation Context Analysis

RESULTS

Restricted Transgene Expression and Age-Dependent Pathology in the EC of rTgTauEC Mice

We generated a mouse line that reversibly expresses human variant tau P301L primarily in EC-II, the rTgTauEC mouse (Figure 1A). We took advantage of a mouse line in which expression of a tet transactivator transgene is under control of the neuropsin gene promoter (Yasuda and Mayford, 2006). This line was crossed with the Tg(tetO-tau$_{P301L}$)4510 line that only expresses human tau carrying the P301L frontotemporal dementia mutation in the presence of a tet transactivator (Santacruz et al., 2005). Human tau expression in bigenic rTgTauEC mice is limited largely to the superficial layers of medial EC and the closely related pre- and parasubicular cortices (Figures 1B and 1C).

(de Calignon et al., 2011)



(Fu & Schneider, 2020)

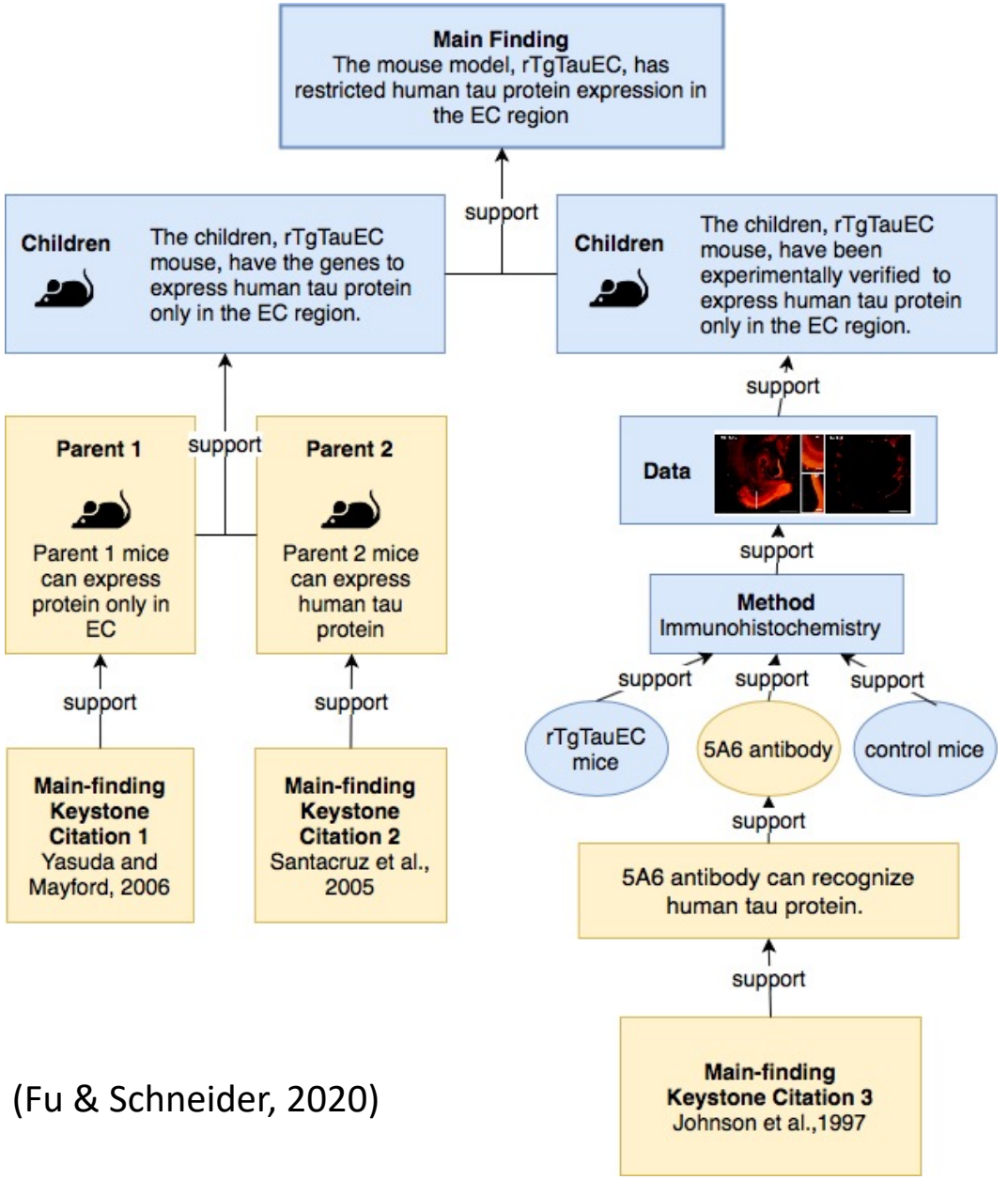
Citation Context Analysis

RESULTS

Restricted Transgene Expression and Age-Dependent Pathology in the EC of rTgTauEC Mice

We generated a mouse line that reversibly expresses human variant tau P301L primarily in EC-II, the rTgTauEC mouse (Figure 1A). We took advantage of a mouse line in which expression of a tet transactivator transgene is under control of the neuropsin gene promoter (Yasuda and Mayford, 2006). This line was crossed with the Tg(tetO-tau$_{P301L}$)4510 line that only expresses human tau carrying the P301L frontotemporal dementia mutation in the presence of a tet transactivator (Santacruz et al., 2005). Human tau expression in bigenic rTgTauEC mice is limited largely to the superficial layers of medial EC and the closely related pre- and parasubicular cortices (Figures 1B and 1C).

(de Calignon et al., 2011)



(Fu & Schneider, 2020)

Classification

A classification system that determines the impact to a citing paper.

How many items were cited?

- Singleton
- Citation cluster

Whether the cited item's main findings support the citation context?

- Main-finding support
- Pass-through support: The reference(s) of the cited paper provided the support.
- No support

(Fu & Schneider, 2020)

Can you automatically extract keystone citations?

Challenge 1: Finding keystone citations requires a global understanding of a paper.

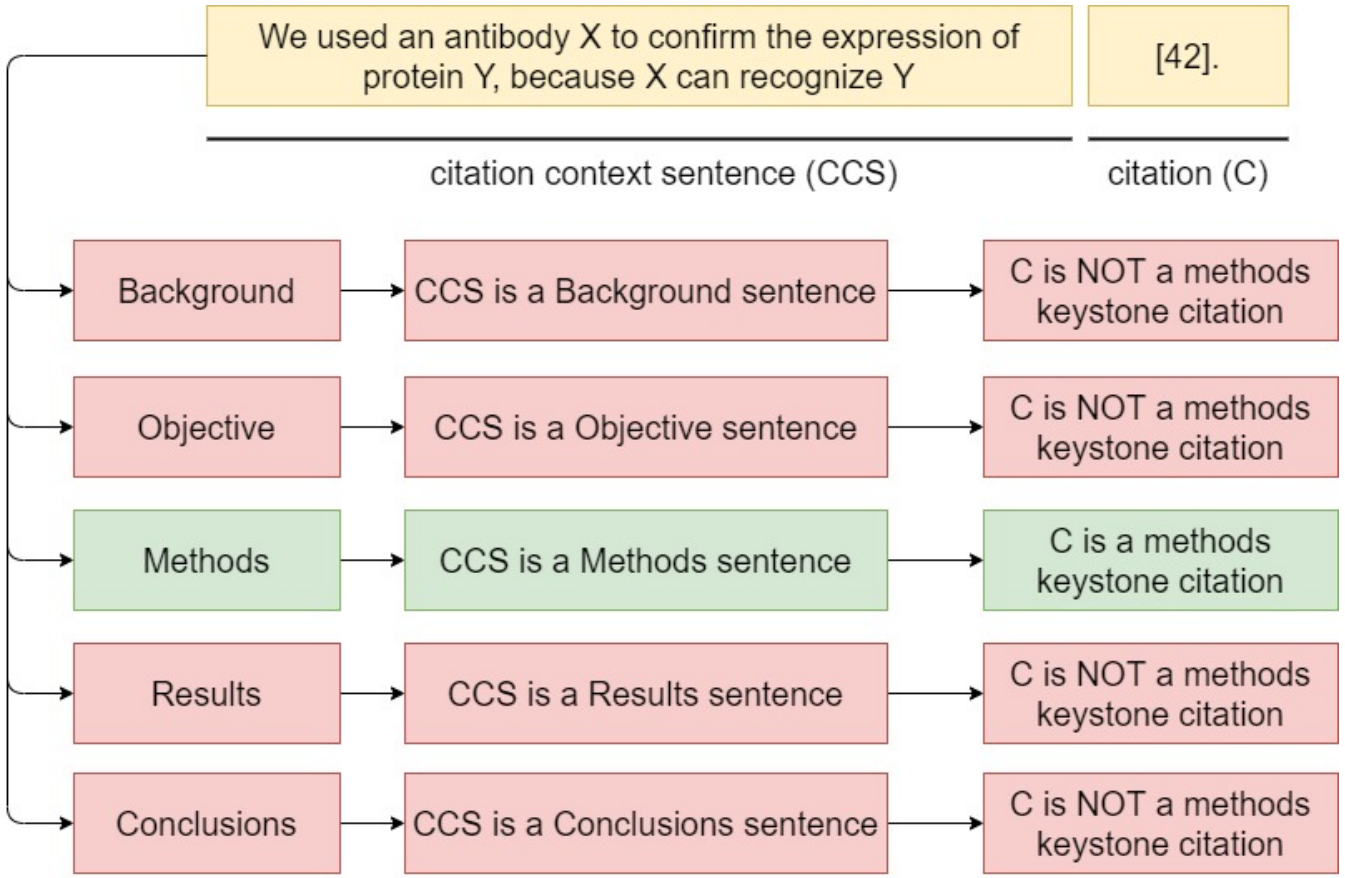
Solution: Start with a subset of the keystone citations: **Citations that support research methods and materials**, because:

(1) Under three argument-based ontologies (Micropublication, SEPIO and RDO), citations that support methods and materials will always be keystone citations.
(2) Their keystone status can be determined by using the citation context only.

Challenge 2: No training dataset.

Solution: Repurposed classifiers trained on biomedical abstract sentences with section labels (e.g., Background, Objective, Methods, Results, and Conclusions). We also got hold of a manually labeled dataset from Prof. Halil Kilicoglu.

The Proposed Strategy



- All binary classifiers
- "No hit" means a sentence received no positive classifications from any classifiers.

The Training Dataset

500 abstracts randomly selected from PubMed without sub-field specifications

All sentences in the 500 abstracts were included, except 34 sentences that were not part of the narrative.

Annotated by three bioinformatics experts.

5517 sentences
Background: 16.5%
Objective: 8.9% Methods: 23.5% Results: 35.1%
Conclusions: 16.0%

Two datasets to test the strategy

- **The JCDL dataset**
  - Contains 9 keystone citation context sentences collected by the authors YF and JS for (Fu & Schneider, 2020).
  - All contain methods keystone citations.

- **The Willoughby-Hoye dataset**
  - A collection of 99 citation context sentences citing the Willoughby-Hoye protocol (Willoughby et al., 2014, Willoughby et al., 2020) downloaded from scite.ai.
  - This paper was chosen because it contains a code glitch and was a subject of our JCDL paper.

Machine Learning Models

- **Representation**: "Bag-of-words"
- **Feature selection**: Information gain (Yang & Pederson, 1997)
- **Classification algorithm**: Support vector machine
- **Internal evaluation**: metrics averaged through 10-fold cross-validation.

Class	No. of Features	Accuracy	Precision	Recall	F1
Background	100	0.858	0.671	0.278	0.392
Objective	100	0.934	0.826	0.339	0.477
Methods	700	0.865	0.820	0.542	0.652
Results	800	0.814	0.835	0.585	0.688
Conclusions	100	0.858	0.684	0.216	0.327

Classification Result of the JCDL dataset

Keystone citation context sentences	Annotation from (Fu & Schneider, 2020)	Classification
Therefore, we assessed two synaptic markers in the perforant pathway terminal zone of rTgTauEC mice: synapsin-I, a marker of synaptic vesicles, and PSD-95, a postsynaptic marker that has been reported to decrease early in neurodegeneration (Zhao et al., 2006).	Material	Methods
The evaluation of Boltzmann-averaged 13C and 1H magnetic shielding tensors and isotropic chemical shifts from density functional theory (DFT) followed Hoye's protocol[25] adapted as follows	Methods	Methods
Therefore, we turned to a protocol that relies on density functional theory-based computations of 1H and 13C NMR chemical shifts and the use of statistical tools to assign the experimental data to the correct isomer of a compound[28].	Methods	Methods
To resolve this ambiguity, we conducted NMR prediction calculations (Figure 1 B)[13,14].	Methods	Methods

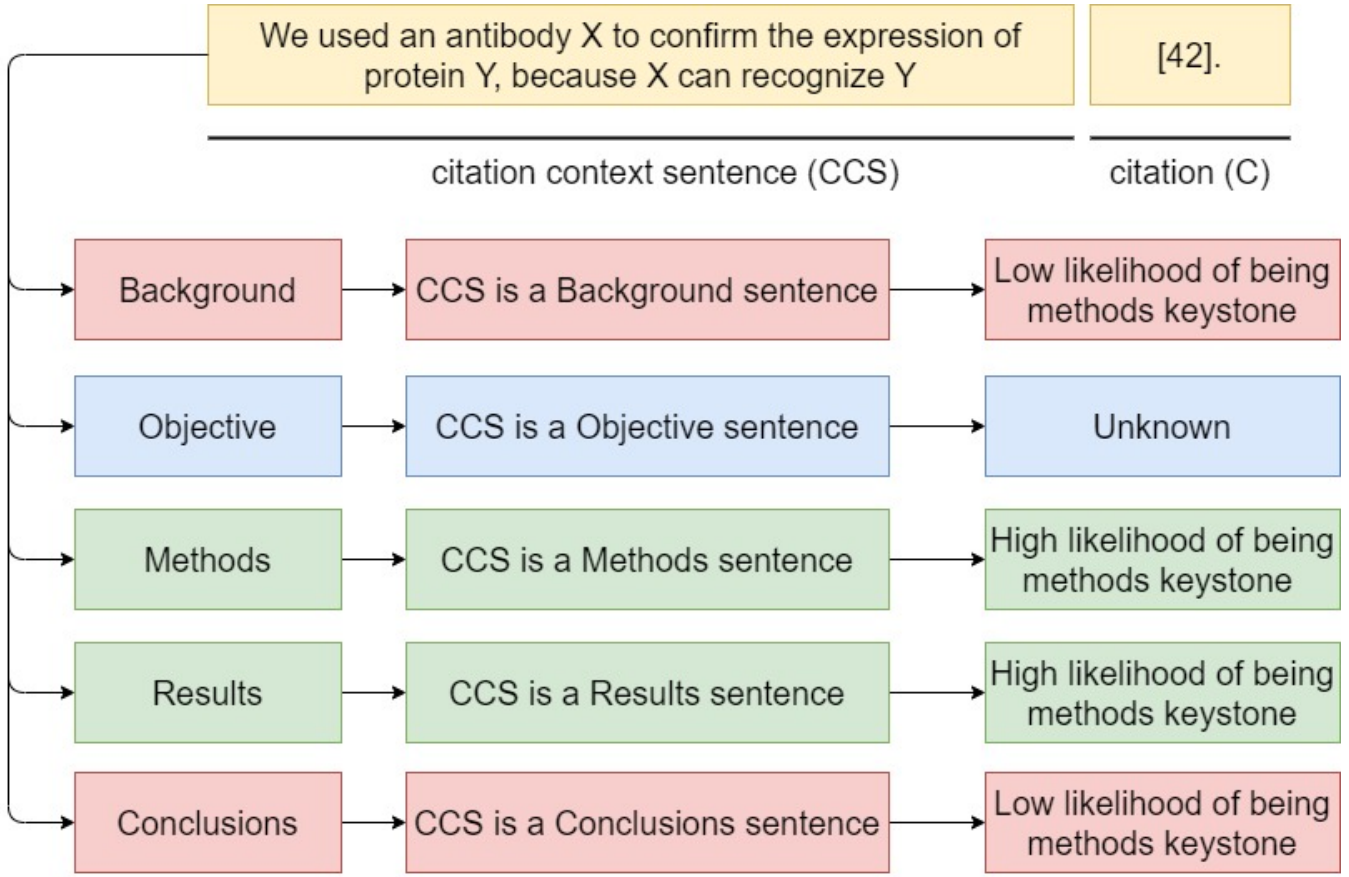
Keystone citation context sentences	Annotation from (Fu & Schneider, 2021)	Classification results
Immunohistochemistry using the 5A6 antibody (courtesy of Dr.G.V. Johnson, University of Rochester), a monoclonal antibody raised against the longest form of recombinant human tau which recognizes an epitope between amino acids 19 and 46 (Johnson et al., 1997), confirmed strong expression of tau protein in superficial layers of the MEC and parasubiculum in rTgTauEC mice at 3 months of age compared to a control brain (Figure 1D).	Material	Results
	Methods-Results hybrid	
We took advantage of a mouse line in which expression of a tet transactivator transgene is under control of the neuropsin gene promoter (Yasuda and Mayford, 2006).	Material	No hit
	Imperfect Methods classifier	
This line was crossed with the Tg(tetO tauP301L)4510 line that only expresses human tau carrying the P301L frontotemporal dementia mutation in the presence of a tet transactivator (Santacruz et al., 2005).	Material	No hit
In AD, early hallmarks include the loss of synapses, and comparison of AD patients to age-matched control individuals showed that the density of synapses correlated strongly with cognitive impairment, suggesting that loss of connections is associated with the progression of the disease (DeKosky and Scheff, 1990; Scheff and Price, 2006; Terry et al., 1991).	Methods	No hit
	The sentence containing the citation alone is not enough.	

Classification Result of the Willoughby-Hoye Dataset

Class	No. of instances where Willoughby-Hoye protocol is a methods keystone citation	Total No. of Instances	Percentage
Background	1	10	10%
Objective	0	0	-
Methods	21	22	95%
Results	10	10	100%
Conclusions	0	2	0%
No hit	19	56	21%
Total[a]	50	99	50%

[a]One instance was classified as both Methods and Results, and therefore the total number is 99, not 100.

The Revised Strategy



Future Work

- Fine tuning current classifiers
- Experimenting with multiclass classifiers and neural networks.
- Evaluating the revised strategy with more data.
- Constructing a larger gold standard citation context sentence dataset, using the current classifiers as screen tools.
- Using the gold standard dataset to train classifiers for real-world applications.

Thank you!

contact: fu5@illinois.edu

Thank you!

contact: fu5@illinois.edu

References

Bar-Ilan, J., & Halevi, G. (2017). Post retraction citations in context: A case study. *Scientometrics*, *113*(1), 547–565. https://doi.org/10.1007/s11192-017-2242-0

Bolboacă, S. D., Buhai, D.-V., Aluaș, M., & Bulboacă, A. E. (2019). Post retraction citations among manuscripts reporting a radiology-imaging diagnostic method. *PLOS ONE*, *14*(6), e0217918. https://doi.org/10.1371/journal.pone.0217918

Budd, J M, Sievert, M., Schultz, T. R., & Scoville, C. (1999). Effects of article retraction on citation and practice in medicine. *Bulletin of the Medical Library Association*, *87*(4), 437–443.

Budd, John M., Coble, Z. C., & Anderson, K. M. (2011). *Retracted publications in biomedicine: Cause for concern*. 390. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/conferences/confsandpreconfs/national/2011/papers/retracted_publicatio.pdf

Clark, T., Ciccarese, P. N., & Goble, C. A. (2014). Micropublications: A semantic model for claims, evidence, arguments and annotations in biomedical communications. *Journal of Biomedical Semantics*, *5*(1), 28. https://doi.org/10.1186/2041-1480-5-28

De Calignon, A., Polydoro, M., Suárez-Calvet, M., William, C., Adamowicz, D. H., Kopeikina, K. J., Pitstick, R., Sahara, N., Ashe, K. H., & Carlson, G. A. (2012). Propagation of tau pathology in a model of early Alzheimer's disease. *Neuron*, *73*(4), 685–697. https://doi.org/10.1016/j.neuron.2011.11.033

References

Fu, Y., & Schneider, J. (2020). Towards Knowledge Maintenance in Scientific Digital Libraries with the Keystone Framework. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 217–226. https://doi.org/10.1145/3383583.3398514

Gilbert, G. N. (1977). Referencing as Persuasion. *Social Studies of Science*, *7*(1), 113–122.

Hamilton, D. G. (2019). Continued Citation of Retracted Radiation Oncology Literature—Do We Have a Problem? *International Journal of Radiation Oncology, Biology, Physics*, *103*(5), 1036–1042. https://doi.org/10.1016/j.ijrobp.2018.11.014

Kochan, C. A., & Budd, J. M. (1992). The persistence of fraud in the literature: The Darsee case. *Journal of the American Society for Information Science*, *43*(7), 488–493. https://doi.org/10.1002/(SICI)1097-4571(199208)43:7<488::AID-ASI3>3.0.CO;2-7

Schneider, J., Ye, D., Hill, A. M., & Whitehorn, A. S. (2020). Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. *Scientometrics*, *125*(3), 2877–2913. https://doi.org/10.1007/s11192-020-03631-1

Small, H. G. (1978). Cited documents as concept symbols. *Social Studies of Science*, *8*(3), 327–340. https://doi.org/10.1177/030631277800800305

Willoughby, P. H., Jansma, M. J., & Hoye, T. R. (2014). A guide to small-molecule structure assignment through computation of (1 H and 13 C) NMR chemical shifts. *Nature Protocols*, *9*(3), 643–660. https://doi.org/10.1038/nprot.2014.042

Yang, Y., & Pedersen, J. (1997). A Comparative Study on Feature Selection in Text Categorization. *ICML*. https://dl.acm.org/doi/10.5555/645526.657137

Appendix 1: the Micropublication Ontology

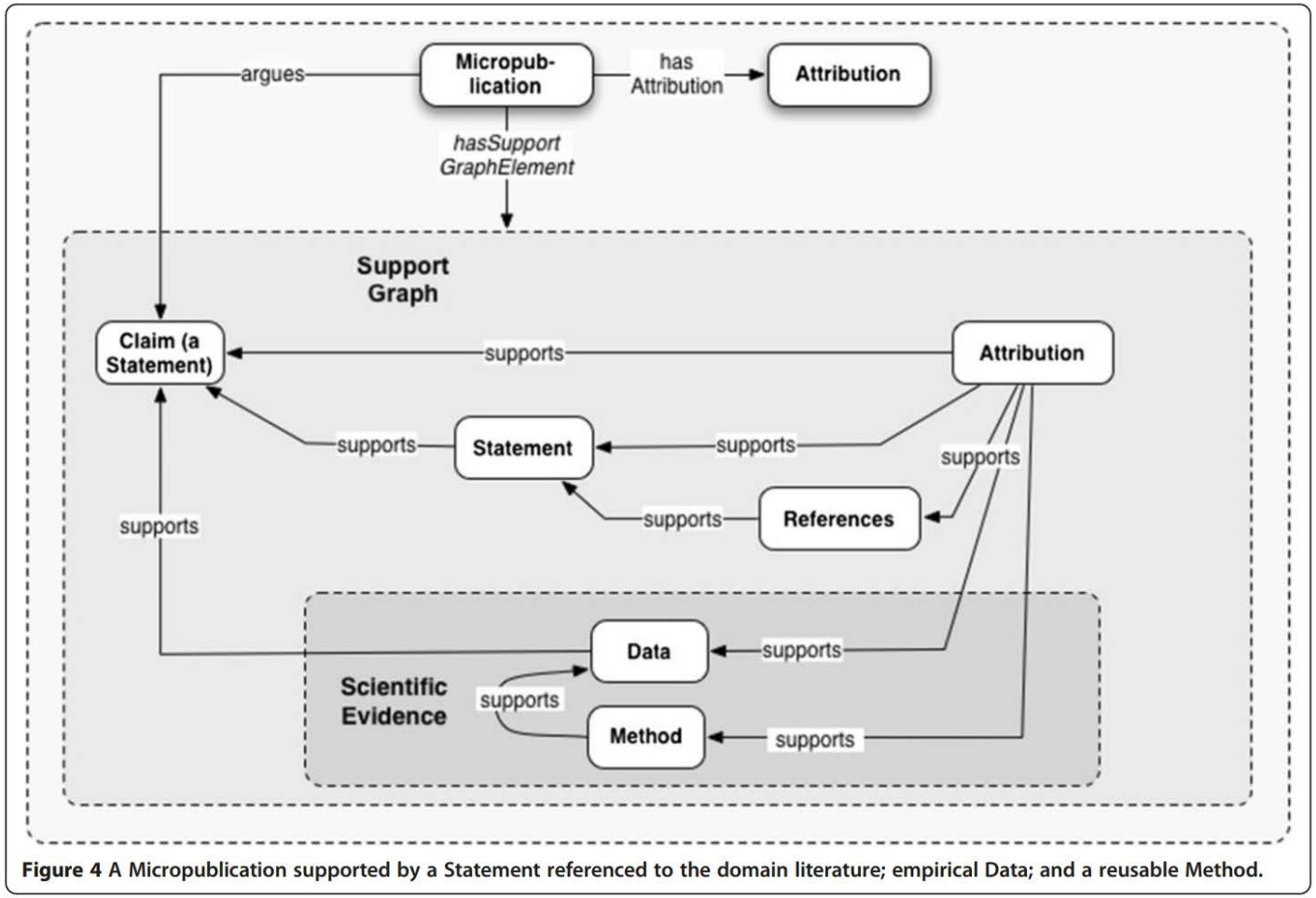


Figure 4 A Micropublication supported by a Statement referenced to the domain literature; empirical Data; and a reusable Method.

- "Our model is based on understanding scientific publications as arguments, which present a narrative of experiments or observations, the data obtained, and a reasoned interpretation ("finding") of the data's meaning.
- Grounded on Toulmin model (Claim, data, warrants, qualifier, rebuttal, and backing).

Clark et al., 2014

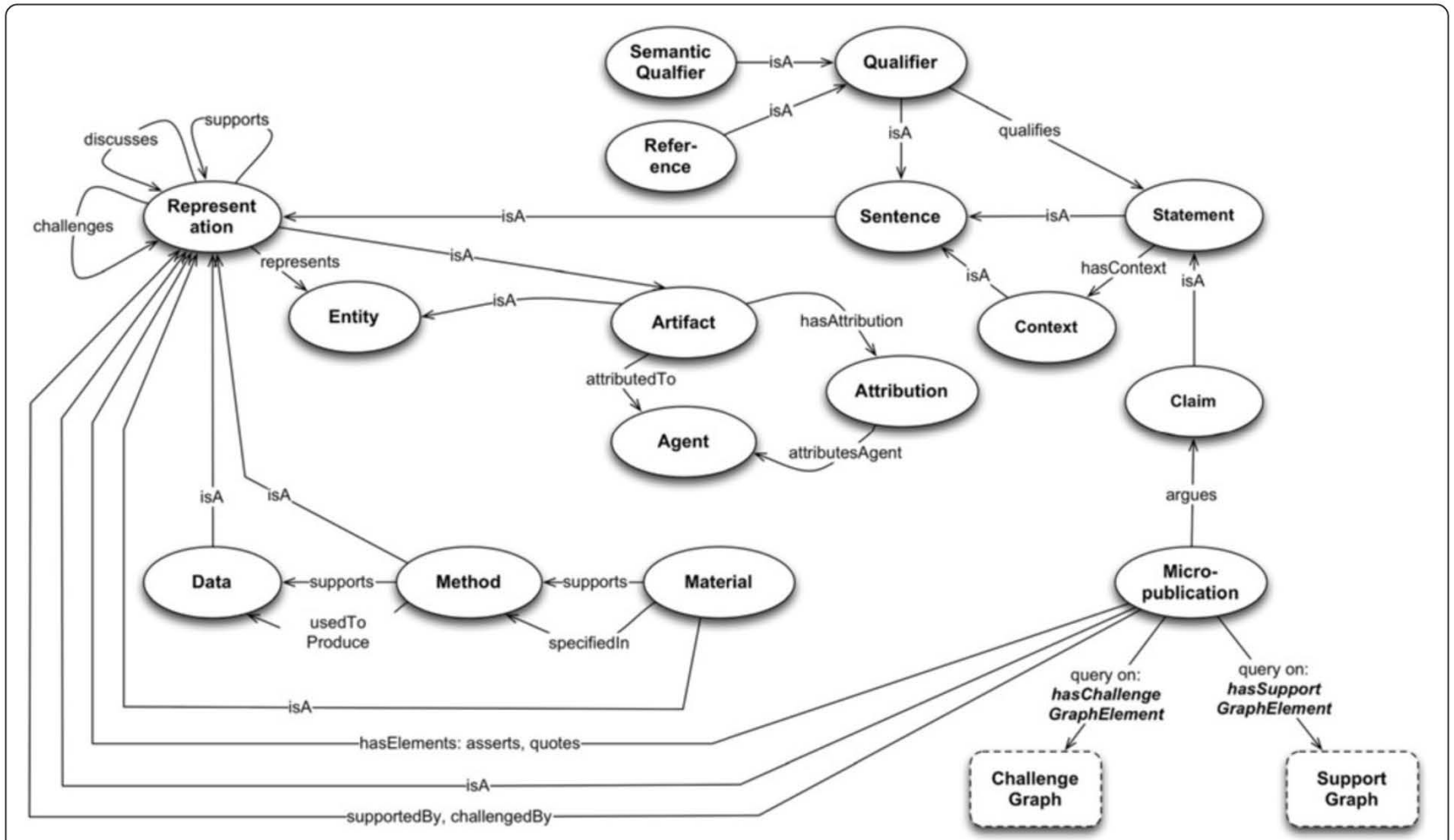


Figure 5 Major classes and relationships in the model. A *Claim* is the main *Statement* argued by a *Micropublication*. A Statement is a truth-bearing *Sentence*, which may be variously *qualifiedBy* some *Qualifier*. A Sentence is a well-formed sequence of symbols, intended to convey meaning; and is not necessarily either complete or truth-bearing. A *Micropublication hasElements* consisting of those *Representations* it *asserts* or *quotes*. A Representation *supports* or *challenges* other Representations. The supporting Representations which are *elementOf* the Micropublication will be in its *SupportGraph*; challenging elements, will be in its *ChallengeGraph*. Dashed-line boundaries indicate graphs instantiated by query.

Clark et al., 2014

Appendix 2: Top 20 Features

Background	Objective	Methods	Results	Conclusions
%	study	p	%	)
study	aim	assessed	p	(
p	investigate	included	<	%
known	evaluate	performed	=	findings
=	aimed	%	ci	suggest
<	determine	not	;	p
remains	compare	conducted	study	<
important	purpose	significant	)	effective
many	whether	evaluated	(	;
disease	aims	retrospective	0.001	future
analysis	describe	analyzed	higher	=
yet	objective	underwent	significant	needed
little	%	calculated	mean	results
ci	investigated	increased	respectively	indicate
mean	explore	measured	vs.	will
still	present	data	+/-	implications
despite	assess	more	0.05	contribute
compared	examine	higher	decreased	conclusion
patients	review	associated	clinical	ci

Samples of Methods Keystone Citations

"Therefore, we assessed two synaptic markers in the perforant pathway terminal zone of rTgTauEC mice: synapsin-I, a marker of synaptic vesicles, and PSD-95, a postsynaptic marker that has been reported to decrease early in neurodegeneration (Zhao et al., 2006)."

"The evaluation of Boltzmann-averaged 13C and 1H magnetic shielding tensors and isotropic chemical shifts from density functional theory (DFT) followed Hoye's protocol[25] adapted as follows."

"Therefore, we turned to a protocol that relies on density functional theory-based computations of 1H and 13C NMR chemical shifts and the use of statistical tools to assign the experimental data to the correct isomer of a compound[28]."