Evaluating ParsCit's Citation Context Extraction Functionality

Xiaokai Cui, Jodi Schneider
School of Information Sciences, University of Illinois at Urbana-Champaign

ILLINOIS
School of Information Sciences

Motivation

What is citation context?

Citation context refers to the sentences or words which state the content from the cited paper. In the figure, the citation context is the yellow highlighted part. The red "[11]" is a **citation anchor** to the 11th paper in the reference section.

> Hence, it may or may not be a "good" one. There has also been extensive research on automating application rewriting due to database evolution, e.g., [11]. This work focuses on semantic-preserving schema changes. As we noted earlier, DBAs will generally not make such changes because of application risk. Hence, our work focuses on changes that do not necessarily preserve application semantics.

Sample Citation Context[3]

Why do people extract citation context?

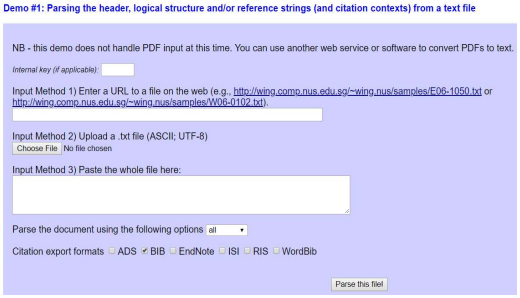
- ❑ To better evaluate the value of citations
- ❑ To reveal manipulation
- ❑ To categorize citations
- ❑ To do sentiment analysis
- ❑ To cluster documents
- ❑ To recommend citations and papers
- ❑ To disambiguate authors

Why is automated citation context extraction important?

- ❑ Citation contexts are important for various applications.
- ❑ Too much human effort is needed for manual annotation.
- ❑ Automated citation context extraction could help to scale many applications, especially when large amounts of data are needed.

ParsCit

The only automated citation context parsing tool[2]



How does ParsCit work?

- ❑ **Parses reference strings to get metadata**
- ❑ **Locates citation anchors in the full text**
- ❑ **Scopes the citation context using a window**

Data

Criteria for ideal corpus	Athar's corpus
Human annotated or at least human cleaned (Minimum requirement)	√
All citation contexts annotated	√
Citing papers provided	√
In English	√
In text format	√
Not domain specific	X
Large data size	√
Publicly available	√

Sample data from Athar's corpus (our gold standard)[1]

Cited Paper No.	Citing Paper No.	Gold standard citation context
A00-1043	A00-2024	"We analyzed a set of articles and identified six major operations that can be used for editing the extracted sentences, including removing extraneous phrases from an extracted sentence, combining a reduced sentence with other sentences, syntactic transformation, substituting phrases in an extracted sentence with their paraphrases, substituting phrases with more general or specific descriptions, and reordering the extracted sentences (Jing and McKeown, 1999; Jing and McKeown, 2000)."

Methods

Input papers in Athar's corpus to ParsCit. Get ParsCit's automatically extracted citation contexts.

↓

Compare the extracted anchor locations with the gold standard. Check whether ParsCit locates all citation contexts correctly.

↓

Compare the extracted citation context with the gold standard. Check whether ParsCit scopes all citation contexts correctly.

Evaluation metrics

Precision and Recall

Evaluation of citation anchor locating
- ❑ Recall measures how many citation anchors ParsCit located, from the gold standard.
- ❑ Precision measures how many of ParsCit's citation anchor locations match the gold standard.

Evaluation of citation context scoping
- ❑ Recall measures how many tokens of the gold standard are within ParsCit's scoping window.
- ❑ Precision measures how many of tokens within ParsCit's scoping window match the gold standard.

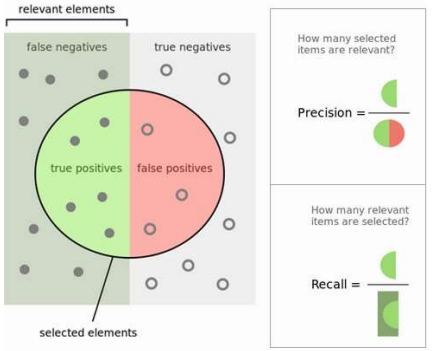


Figure from Wikipedia[4]

Sample of different scoping windows[3]

…Data Civilizer assumes the new schema will obey the dominant enterprise objective of minimizing application maintenance.
Hence, it may or may not be a "good" one. There has also been extensive research on automating application rewriting due to database evolution, e.g., [11]. This work focuses on semantic-preserving schema changes. As we noted earlier, DBAs will generally not make such changes because of application risk. Hence, our work focuses on changes that do not necessarily preserve application semantics.

…Hence, it may or may not be a "good" one. There has also been extensive research on automating application rewriting due to database evolution, e.g., [11]. This work focuses on semantic-preserving schema changes. As we noted earlier, DBAs will generally not make such changes because of application…

…minimizing application maintenance.
Hence, it may or may not be a "good" one. There has also been extensive research on automating application rewriting due to database evolution, e.g., [11]. This work focuses on semantic-preserving schema changes. As we noted earlier, DBAs will generally not make such changes because of application risk. Hence, our work focuses on changes that do not necessarily preserve application semantics.
Our research is also…

Window size	Extracted words	Gold Standard	Overlap	Precision	Recall
30–token	60	23	23	38.3%	100%
One sentence	15	23	15	100%	65.2%
One paragraph	61	23	23	37.7%	100%

Discussion

- ❑ **Recall is the most important metric here because we don't want any information to be lost during extraction.**
- ❑ **Precision is used here to make sure that the auto-extraction doesn't extract too much unwanted content.**

Future work

- ❑ **Complete the evaluation**
- ❑ **Give recommendations to ParsCit**

References

[1] Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. In Proceedings of the ACL 2011 student session (pp. 81-87). Association for Computational Linguistics.
[2] Councill, I. G., Giles, C. L., & Kan, M. Y. (2008). ParsCit: an Open-source CRF Reference String Parsing Package. In LREC (Vol. 8, pp. 661-667).
[3] Stonebraker, M., Deng, D., & Brodie, M. L. (2016). Database decay and how to avoid it. In BigData (pp. 7-16).
[4] Wikipedia. (n.d.) Precision and recall. https://en.wikipedia.org/wiki/Precision_and_recall