



Tuyen Le <ttle@iastate.edu>

Decision on Manuscript MS CPENG-1952 - [EMID:3a05df47b3684161]

1 message

Journal of Computing in Civil Engineering <em@editorialmanager.com>

Tue, Jul 19, 2016 at 1:43 PM

Reply-To: Journal of Computing in Civil Engineering <editorial.coordinator@asce.kwfc.com>

To: ttle@iastate.edu

You are being carbon copied ("cc:d") on an e-mail "To" "H. David Jeong" djeong@iastate.edu
CC: ttle@iastate.edu

Ref.: Ms. No. CPENG-1952

InfraLex: A digital dictionary for unambiguous data sharing in the civil infrastructure domain

Tuyen Le; H. David Jeong, Ph.D.

Dear Dr. Jeong,

Your Technical Paper, listed above, has completed a review for publication in ASCE's Journal of Computing in Civil Engineering. The editor has requested that a revised manuscript be prepared based on the reviewers' evaluations (shown at the end of this email) and submitted for re-review by 09/02/2016.

When preparing the revised manuscript in accordance with the reviewers' concerns and suggestions, be sure to address the following corrections, if necessary:

1. Please remove the figures from your manuscript file and upload them separately (one figure per file) in TIFF, EPS or PDF format. If uploading PDF figure files, please check to make sure the fonts are embedded (see <http://ascelibrary.org/userimages/ContentEditor/1413213918402/ASCE%20PDF%20Figure%20Tool.zip> for instructions for PDF figures). Also, please make sure to reference the figure number in each file name.
2. Please remove the table caption list from your manuscript. Tables have titles and do not need to be listed separately. Additionally, please move the Figure Caption List to be before the tables at the end of the manuscript.
3. Please print out, sign, scan and upload a copy of our Copyright Transfer Agreement which can be found at: <http://ascelibrary.org/userimages/ContentEditor/1412622184009/ASCE%20Authorship%20originality%20and%20CTA%20form.pdf>

Please submit the revised manuscript and a detailed response to the reviewers' criticisms by logging onto the Editorial Management system at <http://jrnccpeng.edmgr.com/> and clicking on the "Submissions Needing Revision" link.

For your convenience, there is a calendar entry item attached that works with electronic calendars in the iCalendar format (e.g. Outlook, iCal, Google). To use, click to open the attachment, and then save it to your calendar. Be advised that the editor may request further revision or decline your revised version if all of the reviewers' comments have not been adequately addressed.

Comments from the Editor and Reviewers can be found below.

We look forward to receiving your revised manuscript.

Sincerely,

Monica Leigh
Editorial Coordinator

Reviewers' Questions & Answers:
Reviewer's Responses to Questions

This manuscript was submitted as a Technical Paper. Does the reviewer think this is the appropriate article type? To see descriptions of the article types, Click Here.

Reviewer #1: Yes. The author is using the correct article type.

Reviewer #2: Yes. The author is using the correct article type.

Reviewers' comments:

Associate Editor: The manuscript has received two reviews. Both reviewers see the potential for publication, but have significant concerns (especially about the novelty and contribution). The authors are, thus, requested to revise the manuscript based on the reviews and resubmit it for another round of review. Please respond point-by-point to each review comment. In responding to a comment, please note any changes made in the manuscript and refer to the respective line numbers. When revising the manuscript, please better highlight your contribution, better justify the advancement in the body of knowledge (especially in comparison to recently published papers), and better describe the implications of the work and how successful the project is.

Reviewer #1: The paper is tackling an interesting and important issue in regards to the development of a lexicon for civil infrastructure.

There are a number of small grammatical errors in the paper and it would benefit from a final English language check (e.g., use of 'to the same' rather than 'of the same').

On page 3 you state that research to address the issue of terminology inconsistency has been very limited, but you also point to major decade long efforts such as bSDD. I think you are more concerned with research into automated approaches, and this should be made explicit here.

On page 4 you present no evidence that the growth of terms is exponential - and it is hard to imagine that this could be the case.

On page 6 you could present more explanation as to why reliance on digital dictionaries is becoming a bottleneck.

On page 7 it would be useful to present some information on the performance of the techniques and system that you mention here.

On page 8 isn't it more important whether efforts like bSDD are complete rather than how long they take to create?

On page 9 how do we know that the coverage of this corpora is sufficient and complete? How about terms from corpora in other English language speaking countries? Or corpora from countries which use languages other than English? Need to scope your work better here.

On page 10 you should discuss the implication of removing tables and equations from your corpus.

On page 14 you should explain what a 'one-hot vector' is.

In the discussion and conclusions there is no consideration of what is good enough in a lexicon. Is your 81% precision sufficient? Is a F-measure of 65% sufficient?

It would be useful to have the final lexicon available online somewhere so that readers of the paper can access and assess the results of this work.

Reviewer #2: The manuscript introduces the development of InfraLex, a lexicon intended for enabling data integration. The manuscript is interesting as it intended to automate the process of lexicon generation using NLP techniques. The reviewer found the manuscript to have a merit for potential publication in the Journal of Computing in Civil Engineering. However, the following comments should be addressed prior to the acceptance of the publication.

1. The authors might want to revise the manuscript title to better reflect the scope of research. Particularly, the text corpus came from only roadway design guidelines and does not constitute the entirety of civil infrastructure. While the manuscript places an equal emphasis on the approach as well as on the research product (i.e. InfraLex), the manuscript title should be more carefully crafted into something more accurate.

2. Several statements within INTRODUCTION need to be supported.

- 2.1 Page 2, Lines 38-40, "The major cost was time spent...a useful format."
- 2.2 Page 2, Lines 48-49, "Polysemy and synonymy are two ...data sources."
- 2.3 Page 3, Lines 61-63, "However, research to address...very limited."

3. The statement citing the cost of interoperability issue on Page 2 (between lines 35 - 38) does not necessarily apply to the target of the paper, civil infrastructure (or more precisely roadway design). The citation from NIST was about the capital facilities, which typically refer to buildings and industrial facilities and not horizontal construction.

4. The authors seemed to suggest that very few research looked into the terminology inconsistency issues in construction whereas a recent paper published in JCCE attempted to address this particular issue in the transportation sector. The paper is entitled "Ontology for Querying Heterogeneous Data Sources in Freight Transportation" and the authors are advised to review it in order to highlight their research uniqueness and contribution.

5. The authors used a single source of citation from arXiv to support their claim that Skip-Gram model outperforms other methods like LSA and so was adopted in the reported research. This is a critical research design decision and should be more carefully evaluated (and supported). arXiv does not require peer reviews and its publications are only moderated to check against obvious policy violations.

6. How do the results from the step 'Noun phrase detection' look like? Porter stemming (or any stemming algorithm) is known to over or under stem terms at times. How was over or under stemming managed when the authors were generating their bag of noun phrases? Why wouldn't the suffix in the word "undivided" in Table 6 be removed after the authors employed the Porter stemming algorithm?

7. A manual evaluation process was still employed to remove inadequate or meaningless terms from the term candidate list. Could this process become the road block when the authors need to scale up their research?

8. How were the precision rates in Figure 3 determined? These rates seemed to provide critical information for the manual evaluation upon the term candidates and it is unclear how these precision rates were obtained prior to the manual evaluation.

9. Results in Table 7 are not sensitive to the different parameter settings. Table 7 can be removed from the manuscript.

10. The body of knowledge does not lie within the method and was more on the research product. This is a relatively weak contribution.