INSTRUCTIONS TO REFEREES

For a contribution to be acceptable for publication in a journal, it must comprise novel material not previously published in a technical journal. The novelty will usually lie in original results, methods, observations, concepts, or applications, but may also reside in syntheses of/or new insights into previously reported research. In a regular paper, the title, abstract, introduction, and summary should be sufficiently informative to make the contributions of the paper clear to the broadest possible audience, and to place them in context with the related work. In addition to these fundamental requirements, acceptance for publication depends on a number of important criteria relating to reader interest, technical content, and presentation. To assist the referee in addressing these criteria, the Review form includes a short answer OVERVIEW (Section III) as well as an open form for DETAILED COMMENTS (Section IV). The principal intent of the Overview is to raise the kind of questions that should be addressed in assessing the paper. In other words, the Overview provides a list of the criteria referred to above and, in this sense, serves as a part of these instructions. In addition, the short answers to these questions provide a uniform synopsis of the review for both the editor and the author(s).

The essential part of the evaluation, however, is the information contained in the reviewer's Detailed Comments (Section IV). Please try to provide one or more pages of comments in this section. At minimum, please provide one half-page of comments. It is hoped that these comments will be guided by the responses indicated in the Overview, with emphasis placed on points that substantiate the recommendation to the editor. A recommendation to accept for publication, whether with no changes or with minor revisions, should be reserved for manuscripts that describe novel work and satisfy the readership, content, and presentation criteria indicated in the Overview.

If major revisions are recommended, the referee should point these out as specifically as possible and should differentiate changes regarded as optional from those judged as mandatory. If the revisions required are extensive, it is perhaps best to reject the paper and recommend preparation of a "new", heavily revised manuscript for resubmission to the journal. If a paper is rejected mainly on the basis of reader interest, the reviewer may wish to recommend submission to a more appropriate journal or conference. Papers with little or no salvageable material should be rejected outright and discouraged from later submission.

ABOUT THE JOURNAL:

This journal’s ambition is to stimulate the exchange of ideas and interaction between these two related fields of interest: Data Engineering and Knowledge Engineering. The journal reaches a world-wide audience of researchers, designers, managers and users. The major aim of the journal is to identify, investigate and analyze the underlying principles in the design and effective use of these systems. The journal achieves this aim by publishing original research results, technical advances and news items concerning data engineering, knowledge engineering, and the interface of these two fields.

The journal covers the following topics:
2. Architectures of database, expert, or knowledge-based systems: New architectures for database / knowledge base / expert systems, design and implementation techniques, languages and user interfaces, distributed architectures.
3. Construction of data/knowledge bases: Data / knowledge base design methodologies and tools, data/knowledge acquisition methods, integrity/security/maintenance issues.
4. Applications, case studies, and management issues: Data administration issues, knowledge engineering practice, office and engineering applications.
5. Tools for specifying and developing Data and Knowledge Bases using tools based on Linguistics or Human Machine Interface principles.
I. SUMMARY AND RECOMMENDATION (TO BE WITHHELD FROM AUTHOR)

Summary of Evaluation
[ ] Excellent
[ ] Good
[x] Fair
[ ] Poor

Recommendation
[ ] Accept without changes
[x] Accept if certain minor changes are made (see Section IV)
[ ] Author should prepare a major revision (see Section IV) for another round of review
[ ] Reject

If the paper is rejected, the author(s) should
[ ] Prepare a major revision and resubmit it as a new paper
[x] Submit it to another journal or conference
[ ] Regard it as not publishable

Contribution: (Please put marks in all columns!)

<table>
<thead>
<tr>
<th>THEORY</th>
<th>APPLICATIONS</th>
<th>OVERALL QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Significant</td>
<td>Excellent</td>
</tr>
<tr>
<td></td>
<td>Tutorial</td>
<td>Good</td>
</tr>
<tr>
<td>X</td>
<td>Possible</td>
<td>Fair</td>
</tr>
<tr>
<td>None</td>
<td>Minor</td>
<td>Poor</td>
</tr>
</tbody>
</table>

II. COMMENTS TO BE WITHHELD FROM AUTHOR

III. OVERVIEW

A) Reader Interest

1. Is the paper of current interest to a reasonable segment of JOURNAL readership?
   [x] Yes
   [ ] Perhaps
   [ ] No

2. Relative to the current level of reader interest in the paper, how is this interest likely to change during the next five years?
   [x] Growing interest
   [ ] Relatively little change
   [ ] Diminishing interest

3. Within its particular field of specialization, is the topic of the paper considered important?
   [x] Yes, definitely
   [ ] Moderately so
   [ ] Not really
B) Content

1. Is the paper technically sound?
   [x] Yes
   [ ] Appears to be, but didn't check completely
   [ ] Only partially
   [ ] No

2. How would you describe the technical depth of the paper?
   [ ] Expert level
   [x] Appropriate for someone working in the field
   [ ] Suitable for the non-specialist
   [ ] Superficial

3. Does the paper make a tangible contribution to the state-of-the-art in its field?
   [ ] Yes, definitely
   [x] To a limited extent
   [ ] No

4. Is the bibliography adequate?
   [x] Yes
   [ ] Yes, after certain additions and/or deletions are made (see Section IV)
   [ ] No

5. To what extent is material in the paper likely to be used by other researchers and practitioners?
   [ ] Large
   [x] Average
   [ ] Small

C) Presentation

1. Is the abstract an appropriate digest of the work presented?
   [x] Yes
   [ ] No

2. Does the introduction clearly state the background and motivation in terms understandable to the nonspecialist?
   [x] Yes
   [ ] Probably
   [ ] No

3. How would you rate the overall organization of the paper?
   [x] Satisfactory
   [ ] Could be improved
   [ ] Poor

4. Relative to its technical content, is the length of the paper appropriate?
   [x] Yes
   [ ] No, it should be lengthened
   [ ] No, it should be shortened

5. Is the English satisfactory?
   [x] Yes
   [ ] No
IV. DETAILED COMMENTS (TO BE RETURNED TO AUTHORS) (Please try to provide two or more pages of comments in this section. At minimum, please provide two pages of comments. THIS SECTION MUST BE SUBMITTED ELECTRONICALLY ONE HOUR BEFORE THE DEADLINE, THE EMAIL NEEDS TO BE SENT TO ANGRYK at CS.MONTANA.EDU)

• Identify your view of the major contributions (or potential thereof) of the paper.

• Specify the major reason(s) for acceptance/resubmit/reject.

• Itemize specific revisions needed/suggested.

Detailed Comments

Major Positives

• The major contribution of this research towards its particular specialization of study is: this is a comprehensive study of using Latent Semantic Indexing and Random Projection in Semantic Text Indexing. The authors use variety of experiments to show the readers the results of using different type of text indexing techniques. If someone is working on Semantic Text Indexing, he can very well use this study to decide on the type of indexing he wants to use on a particular dataset in hand.

• The approach used by the authors seemed pretty unique since, the related works never tried to use so many Semantic Text Indexing scheme in a single study. This can be regarded as a comprehensive study of the few semantic text indexing mentioned over here, since the authors used a total of 12 combinations of a few indexing schemes.

• Since the dataset used is a very large dataset, thus the results of this study can be considered pretty stable. There would not be a huge difference in results if the future researchers use a dataset even larger than this, since the dataset for this study is already scaled to a high degree.
• Based on their results the authors clearly specify in their study, the indexing techniques that will not be scaling well.

• The Background review is quite comprehensive. The authors have provided a few referenced works which the future researchers can reference. Since the authors have provided the summary of a few well-known semantic text indexing techniques, which have been researched in the past, the readers can choose a referenced work according to his liking (based on the summary of the work).

• Before the Experimental sections the authors’ introduction of “Statistics and Semantic Comparison” section sets the readers’ expectations straight about what comparisons they are expecting and what are the techniques that would be employed. Moreover, the uniformity with which the authors performed the experiments is appreciable.

• In Table 5, where the authors show the corresponding matrices from RP-Original, etc., the authors clearly states that the results of this study are not what are expected. The reasons are stated in brief and the future researchers who are using this technique for Semantic Text Indexing should take extra care into using these techniques in their research because they certainly are not ensuring good results under the authors’ experimental setup.

• The mathematical proofs although very few are quite sound.

• The authors provide a very brief but crisp summary of their study at the end. The summary would be really helpful for the researchers if the researchers are still undecided on the technique they need to use for their Semantic Text Indexing.

• The organization and the clarity of language in this paper is highly appreciable. Moreover the adherence to the format is also appreciable.

Major Drawbacks

• Although the authors have used a total of 12 combinations, but the number of disparate indexing schemes used in this study seems to be pretty small. It would have a much better comprehensive study if the authors would have introduced a few more text indexing schemes (for example:- ontology based text indexing schemes).

• The study has no novelty of idea attached to it.

• Although RP was projected to be an efficient technique throughout the paper, the results of Table 5 where the RP does not perform well, should have received a bit of more study and attention from the authors. If the reason for the bad performance is a mistake on the authors’ part or due to the inherent nature of the dataset, the future researchers who are referencing this might be misled into not using this technique entirely for their research.
• The extent of the sparseness of the matrix is never conveyed to the users.

• The authors have used the concept of TF-IDF for all their experiments, but in the “background study”, the authors have provided a very little reference about the concept.

• Moreover a comprehensive justification of using TF-IDF is not provided in the study.

Revisions

• In the quantile – quantile plots of Figures 9 & 10, the sudden anomalous behavior of the tail should have received a bit of attention from the authors in the paper.

• There are a few typing errors in the paper. For an example, in the second paragraph of the Results section, the authors have used “Fig. 3Fig. 4”, which I think is a case of typing mistake.

• In the graphs used, the authors have not provided the legends to the graphs, which would have ensured easier readability on the readers’ part.

• Since the authors have marked the highest and lowest difference in their “Normalized Intra- and Inter-Group Differences Between Means- Original Data Set”, it would be really helpful if they do the same for all the tables presented thereafter for each comparison to ensure better readability.

• The authors made a hypothesis in the Introduction section that “LSI will enhance measures of semantic relationships, and RP will preserve such relationships” which is proved in the course of the study. In the summary (or in the introduction section itself) the authors should justify this hypothesis in brief.