Legend:

- updated → represents typos and figure scales. Document was updated without any rev tags
- rev → moderate change. Document was updated and text is marked in orange so the changes can be quickly identified
- deferred → discussed it with Gail. We agreed to not address these revisions

Chapter 1 - Introduction

- Examiners (Gail's email + external examiner's report)
  - Add more about how the structure of the dissertation fits together Deferred
    - AM: 1.3 already has a roadmap.
- Alex
  - Figure 1.2 Not readable on paper updated
  - Figure 1.4 Not readable on paper updated
  - p10. Better explain the overall thesis goal (similar to Christina's) Deferred

Chapter 2 - Related Work

- Examiners (Gail's email)
  - Update the IR part of related work to be up to date rev
    - AM: added Gerard Salton definition about IR
    - AM: detailed some of the subfields of IR
    - AM: detailed that 2.2.2 focus on traditional IR approaches and that other sections of Chapter 2 detail NLP, ML, and DL applications of information retrieval
    - AM: Cited [113, 184, 204] which are all papers that present literature reviews about IR
- Luanne
  - Examiners' comments about IR
    - AM: see above
- Alex
  - Explain tags PRP, VP, NNP, etc updated
    - AM: I added a footnote to Penn's treebank containing the description for each acronym

Chapter 3 - Characterizing Task-relevant text

- Examiners (Gail's email)
  - Add a table of independent and dependent variables and measures. rev
    - AM: Added table 3.1 Experimental Variables
  - Make it clear that the tool was used for the study and the study was not about the tool. rev
    - AM: In 3.2.4, I added sentence describing that the tool was a means to an end
  - Clarify the preamble and the research questions (see External Examiner report attached) rev
  - AM: I rephrased the research question. I tried to better align it with the text justifying the RQ goal: "The first research question seeks to characterize text in documents relevant to a task."
    - AM: rationale was also updated
- Christina
  - The mention of "open the questions" followed by "research questions" in the next paragraph hinders comprehension, among other problems. rev
    - AM: rephrased last sentence in the 1st paragraph removing the "open questions" term
  - I missed a more careful description of the experimental design presented in Chapter 3.
    - What is/are the goal(s) of the experimental study? rev
      - AM: Added introductory sentence to 3.1 to make the study goal more explicit
- Luanne
  - You describe this study as an experiment - what were the independent and dependent variables? How were participants assigned to the conditions? Perhaps it is more of an observational user study than a true experiment deferred
    - ~~AM: I don't have any null hypotheses as in a true experiment. So I could rename "experiment" to either a "quasi-experiment" or an "observational user study" as suggested by Luanne~~
  - Wilcoxon test: Not clear what is being tested here rev
    - AM: Moved what is being tested to the start of the sentence to make it more clear (p. 43)

- Alex
    - Clarify artifact selection process, what are examples of assumptions? rev
        - AM: 3.2.2. added Table 3.4 with example of assumptions
    - p29. Explain how reading time was calculated rev
        - AM: 3.2.2. I already had a citation [87], but I moved it to a footnote and added a little bit more text
    - Figure 3.2 Not readable on paper updated
    - Section 3.3.1 (p35) How does the conclusion to “How much text in an artifact is deemed relevant to a task?” this section tie to the chapter RQs? rev
        - AM: add short paragraph tying key findings to RQ1 (p. 35)
    - Similar comment in the conclusion at “How much agreement is there between participants about the text relevant to a task?” rev
        - AM: add short paragraph tying key findings to RQ1 (p. 36)

Chapter 4 - Android Task Corpus

- Luanne
    - corpa →corpora updated
- Alex
    - 4.1 Github tasks: clarify if random selection was across all the issues in a repo rev
        - AM: slightly edited the text to state that selection happened across all the resolved issues
    - Figure 4.3 and 4.4 make them a little bit larger? updated
    - 4.3.1. Clarify if 45 hours was from one or all annotators rev
        - AM: clarified it was across all annotators

Chapter 5 - Identifying Task-relevant text

- Examiners (Gail’s email)
    - 5.2. Make it clear there is a need to train rev
        - AM: added sentence in intro of Section 5.2 to disclose which techniques need training and which do not
            - AM: added reference to Section 5.3.1 which discusses training data in detail

    - 5.2.2. add a statement that the Adam on p 67 is standard approach… rev
        - AM: added sentence that Adam optimizer and other hyper-parameters are standard across SE research that uses DL

- Alex
    - 5.2.1 Explain averaging is a common procedure for this type of function rev
        - AM: add citations to other studies that do averaging
    - 5.2.1 p66 $w_j \rightarrow w_a$ updated
    - 5.2.2. Explain DL jargons briefly (examiners' comments) rev
        - AM: added introductory sentence to explain that the paragraph details the model hyper-parameters
    - 5.3.1. Explain that the 10 experimental runs is to address randomness with BERT rev
        - AM: added sentence to clarify that the average of the 10 executions apply to BERT

Chapter 6 - Evaluating a Semantic Based …

- Examiners (Gail's email)
    - 6.2 Clarify RQ1 and RQ2 - e.g. why they measure something different rev
        - AM: added one extra paragraph in 6.2 explaining each RQ
    - 6.2.1 Clarify that the tasks were different and might not have been equivalent in difficulty. Rev
        - AM: rephrased 2nd paragraph in 6.2.1 to explain difficulty
    - Clarify the results in Chapter 6 and make consistent with intro and summary rev
        - AM: address under "others" (at the end of this doc)

- Luanne
    - 6.2.1 (p89) what was the query? Did you use the Task "title" here as well? rev
        - AM: updated footnote where inputs are detailed
    - would like to see info on the metrics in the methods deferred
        - AM: this is a style choice.
    - p91. Change from "To compute how correct a participant's solution is" to a more direct phrase: " to assess the quality of each solution?" rev
        - AM: updated text accordingly

- Alex

    - 6.1. Explain unsurprising argument (examiners' comments) rev
        - AM: updated text. Removed "unsurprising" and added a more direct explanation
    - Check consistency with Python capitalization updated
    - p83 show figure 6.2 earlier, it appears on section 6.2.3 (p85) updated

Chapter 8 - Summary

- Examiners (Gail's email)
- Alex
    - p. 116 (last bullet point) claim seems strong when compared to the discussion in Chapter 6 rev
        - AM: address under "others" (at the end of this doc)

Others

- Typos pointed by Alex/Luanne updated

- Consider making randomization and blocking more clear for all experiments rev
    - AM: Chapter 3
        - Updated introductory text in 3.2 to explicitly mention randomization and balancing
        - Added table with summary of experimental design

    - AM: Chapter 6
        - Updated introductory text in 6.2 to explicitly mention randomization and balancing
        - Updated 6.2.5 - moved text discussing balancing and randomization to a separate paragraph.

- Clarify the results in Chapter 6 and make consistent with intro and summary rev
    - AM: I had to make changes across Chapters 1, 6 and 8 to ensure consistency
    - AM: Chapter 6
        - 6.3.1 removed statement about performing worse. Rephrased it as the task has more variability
        - 6.3.4 added statement about correctness to the summary of results
        - 6.4 added statement indicating that solutions with TARTI are on average equally or more correct than solutions without it
    - AM: Chapter 8:
        - Changed last bullet point to be consistent with claim in the summary of 6.4
    - AM: Chapter 1:
        - 1.3 Added a similar statement about correctness in the last paragraph of this section
- Library/Formatting updated
    - AM: Added committee page
    - AM: Add Acknowledgements