

# The opponent's evaluation of Ekaterina Garanina's master's thesis

Thesis author: Ekaterina Garanina

Thesis title: Table-to-Text Generation via Logical Forms

Opponent's name: Jiří Hana

Affiliation of the opponent: Charles University Prague, MFF, ÚFAL

The thesis introduces a system for converting the information contained in tables into text.

- The text is **well-structured and easy to understand**, with language that is nearly flawless.
- The research and its implementation in Python constitute a **robust piece of scientific work**.

However, there are some **important issues**:

A. The primary weakness lies in the third chapter, the one dedicated to **related work**. It is both **narrow and shallow**.

Firstly, the author should provide a more comprehensive overview of how their work fits into general Natural Language Generation (NLG) research, reviewing important papers/books such as Reiter & Dale 2000 and Gatt & Krahmer 2017. And by that, I mean really exploring what previous researchers have done and its relevance (or lack thereof) to this thesis. Currently, the thesis often simply mentions a relevant work without any detail. For example, Reiter & Dale 2000 would definitely warrant more space than the two lines on page 5.

Secondly, even when focusing on table-to-text generation, the author intentionally ignores any research that does not use the Logic2Text and LogicNLG datasets.

Thirdly, the papers included are covered too briefly, lacking an in-depth discussion beyond trivial mentions of aspects found in the abstracts.

Moreover, the whole chapter is organized around papers. I would hope to find a "pivot" of this: text organized around problems, referencing papers addressing them.

Section 4.2 on the Logic2Text dataset is a good example of how related work should be described.

B. Often, terms, datasets, and models are **not appropriately defined/referenced** where they are used. Definitions should be provided or referenced in the same chapter without requiring readers to search through previous or subsequent chapters. For instance, the term "logical" in the title of the thesis is mentioned on pages 3 and 5 but indirectly defined only on page 8. In section 5.2, PLOG is mentioned without any reference. And so on.

C. While the **code is elegant and well-designed**, it lacks consideration for others trying to understand it, including the author in the future.

- 1) There are **nearly no comments**. I would expect at least basic docstrings for all public functions.
- 2) Typing hints and at least some unit tests would also help.

D. On top of page 26, you claim "Since we target both validity and variability, ..." Maybe I missed it, but I do not recall seeing any argument as to **why variability is your focus**.

Minor issues:

- On page 14 below Figure 4.1, point 1 claims that references are not formulated in natural language. Yet, "England National Rugby Union Team played 2 times at Twickenham, London" is in natural language (assuming the typo "time" x "times" is not relevant).
- It would have been nice to include hyperlinks in the pdf.

Despite these drawbacks, the thesis presents solid research, and I **fully recommend it for defense**.