

Supplemental Material

Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content

Alan Lundgard and Arvind Satyanarayan

IEEE Transactions on Visualization and Computer Graphics (TVCG), Special Issue on the 2021 Visualization Conference (VIS)

corpus_sentences_labeled.json

Our corpus, consisting of all visualization descriptions (582 total) and labeled sentences (2,147 total), organized as follows.

- **vis id number**
A unique four-digit id number.
 - **descriptions**
A list of each author id and the description they wrote for a given visualization.
 - **sentences**
A list of each sentence in the description, labeled according to our four-level model of semantic content.

corpus_summary_and_evaluation.pdf

A typeset document containing the following.

- **Corpus Summary**
 - Corpus descriptive statistics.
 - Corpus fingerprint visualization.
- **Evaluation Design**
 - Examples of the rank-choice interfaces.
 - All questions shown in the rank-choice evaluation.
- **Evaluation Questionnaire**
 - Demographic questions.
 - Visualization questions.

/evaluation

A folder containing data and code from the rank-choice evaluation.

- **calculate_stats.py**
Code for calculating the statistics reported in the paper.
- **blind_rankings_numeric.csv**
Rank-choice data from blind readers, in numerical form.
- **blind_rankings_text.csv**
Rank-choice data from blind readers, textual form.
- **sighted_rankings_numeric.csv**
Rank-choice data from sighted readers, in numerical form.
- **sighted_rankings_text.csv**
Rank-choice data from sighted readers, in textual form.

/visualizations

A folder containing the visualizations (50 total) used for gathering the corpus, each file named according to the following.

- **vis id number**
A unique four-digit number.
- **chart type**
Bar, line, or scatter.
- **difficulty**
Easy, medium, or hard.
- **topic**
Academic, business, or journalism.