Appendix O Prompts

O.1 Sharding

We show the prompts for the sharding process below, using Math as an example task. Double-bracketed terms are placeholders that get replaced with the actual data. Other tasks share the same outline with different exemplars and rules to enforce stable outputs. We refer the readers to the GitHub repository for the exact prompts on other tasks.

Segmentation

```
You are a given a fully specified instruction, and your task is to segment the instruction into a
units of information that each reveal a single piece of information of the instruction.
You must output a list of segments in the following JSON format:
[
    {"segment": "[exact excerpt from the instruction]"},
    {"segment": "[exact excerpt from the instruction]"},
    ...
]

Rules:
- [Non-overlapping] The segments must be non-overlapping and cover the entire instruction. You can
optionally leave some gaps for non-essential portions of the original instruction (delimiters,
headers, etc.)
- [Minimalistic] You should split the information in the segments to as small as possible. If you
have a compound expression (X and Y), you should split it into two segments. Each segment should
represent a unit of information.

Example Query:
What are the names and locations of the stadiums that had concerts that occurred in both 2014 and
2015?

Output:
{"segments": [
    {"segment": "names and locations"},
    {"segment": "stadiums"},
    {"segment": "concerts"},
    {"segment": "in both 2014"},
    {"segment": "and 2015"}
]}

Now complete the task for the following fully specified instruction:

[[INSTRUCTION]]
```

Rephrasing

```
You are given segments of a fully specified instruction, and your task is to: (1) choose one that
will be the initial shard of a multi-step query, and then (2) rephrase each segment into a
conversational version that are provided to the system in a follow-up turn of the conversation.

Your output should be a JSON object in the following format:
{
    "initial_segment": "[exact excerpt from the instruction]",
    "initial_shard": "conversational version of the initial segment",
    "shards": [
    {"segment": "[exact excerpt from the instruction]", "shard": "conversational version of the
    segment taking the rest of the instruction into account"}
    ]
}

Example:

Full Query:
What are the names and locations of the stadiums that had concerts that occurred in both 2014 and
2015?

Segments:
[
    {"segment": "names and locations"},
    {"segment": "stadiums"},
    {"segment": "concerts"},
    {"segment": "in both 2014"},
    {"segment": "and 2015"}
]

Output:
{
    "initial_segment": "stadiums",
    "initial_shard": "popular stadiums",
    "shards": [
        {"segment": "concerts", "shard": "the stadiums should have concerts during a period"},
        {"segment": "in both 2014", "shard": "the concerts should have occurred in 2014 in the
        stadiums"},
        {"segment": "and 2015", "shard": "the concerts should have also occurred in 2015 in the same
        stadiums"},
        {"segment": "names and locations", "shard": "for the stadiums, returned both the name and
        location"}
    ]
}

Rules:
- [Transform each segment] Make sure each segment is included either as the initial shard or in the
rest of the shards. Do not forget any segments.
- [Short initial shard] Make the initial shard short, not a full sentence, similar to how users use a
search engine like Google.
- [Order of shards] Order the shards in order of importance, from most to least important to the
initial shard. You do not need to keep the order the segments that are provided in.

Now complete the task for the following fully specified instruction and segments:

Fully Specified Instruction:
[[QUESTION]]

Segments:
[[SEGMENTS]]
```

Verification

```
You are given an instruction that fully specifies a problem, and a list of shards. Your task is to
decide whether all the information from the full instruction is captured by the shards.

If not, you should output the information unit from the instruction that is not captured by the
shards.

Example 1:

Instruction:
What are the names and locations of the stadiums that had concerts that occurred in both 2014 and
2015?

Shards:
{"initial_segment": "stadiums", "initial_shard": "I'm looking for active stadiums", "shards":
[{"segment": "concerts", "shard": "the stadiums should have concerts during a period"}, {"segment":
"in both 2014 and 2015", "shard": "the concerts should have occurred in both 2014 and 2015"},
{"segment": "names and locations", "shard": "for the stadiums, returned both the name and
location"}]}

Output:
{"converage": "complete"}


Example 2:
Instruction:
Which Asian countries have a population that is larger than any country in Africa?

Shards:
{"initial_shard": "I'm interested in learning about countries in Asia", "shards": [{"shard":
"consider the population size of these Asian countries"}, {"shard": "the population should be
compared in size"}, {"shard": "specifically, compare to the population of African countries"}]}

Output:
{"coverage": "incomplete", "missing_segment": "the shards do not specify that the population of the
Asian countries should be *larger* than the population of any African countries"}



You must output in JSON format as shown in the examples above.
Now complete the task for the following fully specified instruction and shards:

Instruction:
[[QUERY]]

Shards:
[[SHARDS]]
```