# Annotation Instructions for Panasonic

Project Title: Extracting Causal Relations from Electronics & Supply Chain News

## Introduction

There are two parts to this annotation effort.

* **Task 1: Cause-Effect Span Annotation:**

Annotate 10 articles (more the better) with Cause-Effect spans

* **Task 2: Span Clustering Annotation:**

For each span identified from Task 1, assign an ideal topic label that would be used for grouping of spans in the end.

## Purpose

We wish to annotate a small subset of the data so as to obtain a gold dataset. This gold dataset will be a benchmark for us to verify our model performance. For each task, the following illustrates how we can use the annotations to construct evaluation metrics:

* **Task 1: Cause-Effect Span Annotation:**

We will check the performance of our Cause-Effect Span Detection model using Precision, Recall, F1 and Accuracy metrics against the gold dataset.

* **Task 2: Span Clustering Annotation:**

We will check the clustering discrepancies between our clustering method and the gold dataset using metrics like Normalized Mutual Information (NMI), Variational Inference (VI), Adjusted Rand Index (ARI), etc. [To be decided]

## Instructions

### Task 1: Cause-Effect Span Annotation:

1. Using the “MarketIntelligenceReport.csv” file, refer to the “Translated” column for the article data. Select 10 (more the better) rows to annotate. Note down the ROW NUMBER.
2. For each article, annotate each sentence independently. Note down the SENTENCE NUMBER.
3. For each sentence, identify the Cause and Effect spans and:
   1. Mark them using <CAUSE> and <EFFECT> tags.

<EFFECT>Prior to Ford, two more American firms- General Motors and Harley Davidson have given up on manufacturing their vehicles in India</EFFECT>, <CAUSE>citing a lack of demand for their products</CAUSE>.

Add to it <CAUSE>the rising cost of fuel</CAUSE>, and <EFFECT>driving a car in India becomes a truly elite experience</EFFECT>.

* 1. Create a new line for each relation if there are multiple Cause-Effect spans

<EFFECT>Many battery makers are ramping up production</EFFECT> <CAUSE>to meet soaring worldwide demand</CAUSE> because car makers are accelerating their shift to electric vehicles.

Many battery makers are ramping up production to meet <EFFECT>soaring worldwide demand</EFFECT> because <CAUSE>car makers are accelerating their shift to electric vehicles</CAUSE>.

* 1. Save in Excel (.xlsx/ .csv) with column headers: “row\_id”, “sent\_id”, “annotations”. Refer to “annotation\_example\_format.xlsx” and screenshot below for example:

Text

Description automatically generated with medium confidence

If you need guidelines on what is a Cause or Effect, you may want to refer to [the annotation guidelines for the Causal News Corpus](https://github.com/tanfiona/CausalNewsCorpus/blob/master/documentation/Event_Causality_Manual_for_CASE_2022_Public.pdf). The TLDR summary is a Cause is anything that leads to an Effect. We can construct questions like “Why did the Effect happen?” and check if the Cause does answer this question or not. If it does, it is likely to be a valid causal relation.

### Task 2: Span Clustering Annotation:

1. Obtain all Cause and Effect spans from Task 1 (You can use Python to extract all words between the markers, if you need help coding let me know) and store in a table with column headers: ‘span’, ‘ce\_label’, ‘row\_id’, and ‘sent\_id’
2. Create a ‘topic’ column and annotate with any words you deem fit. Note that for the examples below, Rows 4 & 9 (both have the topic “high demand”) would be grouped together in the final graph.

Graphical user interface, text, application

Description automatically generated

The topic can be anything you decide, but keep in mind that the ultimate goal is to create a useful graph. Hence, too detailed == sparse graph with few node overlaps to make conclusions from. Conversely, too general == dense graph, but nodes might be too general to be informative.