**Relation Extraction review**

# Objective

This summary is created to understand state-of-the-art techniques in relation extraction in research.

# Summary of Relations Extraction

1. The goal of few-shot learning is to classify new data having seen only a few training examples, which can be common to relation extraction
2. Relation classification is to determine the correct relation between two entities in a given sentence [1].
3. Dataset for relation extraction
   1. 2018 FewRel: A Larget-Scale supervised few-shot relation classification dataset [1]
      1. Open-sourced dataset in few-shot relation extraction.
      2. Dataset is here: <http://www.zhuhao.me/fewrel/>
      3. Each relation contains 700 examples, there are in total 100 relations. 64, 16, and 20 relations for train, validation, and test.
   2. 2010 SemEval-2010
      1. 10-20 relation types

# Covid-19 relations

Here are some relations that we want to track related to covid-19 based on the tasks suggested here: <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks>

1. Factors: find the factors related to cause/spread of covid-19 (whether will lead to covid-19), especially for these factors:
   1. Effectiveness of case isolation/isolation of exposed individuals (i.e. quarantine), community contact reduction, inter/inner travel restriction, school distancing, workplace distancing, a multifactorial strategy prevent secondary transmission
   2. Seasonality of transmission
   3. How does temperature and humidity affect the transmission of 2019-nCoV?
   4. Significant changes in transmissibility in changing seasons?
   5. Effectiveness of personal protective equipment (PPE)
2. Risk factors. Factors that increase risk of disease. Specifically, the relation of covid to these risk factors 🡨 Not sure if we can say “hypertension increases chance of covid-19”. Instead, we can definitely say “hypertension increases the severity of covid-19”. May be here we are finding if certain (risk) factors worsen the **outcome** of covid-19 patient or worsen the **progression** of covid-19 patient
   1. Hypertension, Diabetes, Male gender, Heart Disease, COPD, Smoking Status, Age, Cerebrovascular disease, Cardio- and cerebrovascular disease, Cancer, Respiratory system diseases, Chronic kidney disease, Chronic respiratory diseases, Drinking, Overweight or obese, Chronic liver disease
   2. Examples:
      1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7202837/>

“The two-way interaction between COVID-19 and diabetes mellitus sets up a vicious cycle wherein COVID-19 leads to worsening of dysglycemia and diabetes mellitus”

* + - 1. Covid-19 **has relationship** with diabetes.
      2. Covid-19 **worsen** diabetes
      3. Does not indicate diabetes is a risk factor; but indicate covid-19 impact diabetes.
    1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7218375/>

“Existing studies show that the cardiovascular disease increased the incidence and severity of coronavirus infection. “

* + - 1. Covid-19 **has relationship** with cardiovascular disease
      2. Cardiovascular disease **increases** the risk of coronavirus infection
      3. Cardiovascular disease i**ncreases** the severity of coronavirus infection
    1. <https://watermark.silverchair.com/ciaa414.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAoUwggKBBgkqhkiG9w0BBwagggJyMIICbgIBADCCAmcGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMYr5ZcR-runIikFNfAgEQgIICOCuBb4G-FpLkx4VsXEs3o3YW8WJDLL2BGuRutja2pWIpqETnqmWoV8zPRyYRuwf0gDhc-3Inw_WynkwQtUlHp4WGvUn13rpjC7-ZUmLBB-MOMwDivHOIoYijsjnLuyGno0JmUlDdyuUaGb47KUi-zZUo-QEG1sjL47t5Jpq6N52v7xqiSWkcsuSj4KfXHO0YIY9e2dw_sWnIOoJrWZAGERj-o__5Y9GvAqSUKfGf3bpOxc_v8My2IFwpK6y9RG2e5V4RHbj3cF1hRI7LMl0dzjFt2wG9iO61rLbc4tiEg57CvpydettoVyWGZHXxouU0cVQGEFehv8CCOEX9FluHvI75oWpVehzN4z8KD3ImN_5Yq_CIiYq2fFGIMRR8BINQAyYZT__baa82Iqgzao5GDbpJIonqbStsmeq6eHTjSALcEjE9cirePwSO_NGafg4fppUSEv3X_zrjgNstNbeDus766p9UM8Ec7Qq_JwUCTJyoWTaBQSuj1owOKm8ogYitGDmgn0VVhwoeffD3Oh4HSrW2HnH8E1ToCpM0tlTrP9KQYQQly7je4LOaiJqNEqwLItVJ7ZSwock4_doqxBU2M56Ubyv5juMGb212X_1ALA7YUesXufJD2nIenyEFslekUM2-sVjaK42ztLHB6j0W7hVsrvx6D55xT6KcY_BObHSldIfJ2WVHTn169sAYzHwi9_EzCcfumIwock_SDO9-DfFC2UZqM4MvaMWQn2xMKH3LY-qrIRgNYXQ>

1. Efficacy of therapeutics and interventions
   1. Examples:
      1. <https://www.sciencedirect.com/science/article/pii/S1201971220301326>

“Increasing eosinophils may be an indicator of COVID-19 improvement. The COVID-19 patients may benefit from sustained lopinavir use.”

* + - 1. lopinavir **may be efficient** for treating covid-19.
      2. OpenNRE: “eosinophils” has part (P527) “covid-19”

1. Patient characteristics of COVID-19:
   1. Symptoms
   2. Incubation period across different age groups
   3. Proportion of patients who were asymptomatic
2. Factors related to severity outcome of covid-19
   1. <https://www.nature.com/articles/s41392-020-0148-4?fbclid=IwAR0oNmxwN2RaKh4N7xSNizHOJCBwTfWwRoZLV942Hq9OreoT0-E4wmrtFC8>

“Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study”

* + 1. Lymphopenia is **predictive** of COVID-19 severity.
  1. <https://pubs.rsna.org/doi/pdf/10.1148/radiol.2020201433>

“Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia”

* + 1. “Well-aerated lung” is predictive of covid-19 outcome

# Relation Extraction Approach

1. Look like among the 100 relations that is defined in OpenNRE, only “has\_part” (P527) may be useful. The first approach is to see if the risk factor and the covid-19 has\_part.
2. The second approach is that we expand the relation tailor towards our need, and then train few-shot classifiers.
3. We can visualize the clustering of the paper, pretty grouping similar papers together. Each of these clusters probably represent some keywords and we can show that in the interface. This probably can be achieved by topic modeling