

Ekstraksi Informasi Numerik dari Teks Berbahasa Indonesia = Numerical Information Extraction from Indonesian Text

Muhammad Fairuzi Teguh, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920534238&lokasi=lokal>

Abstrak

Open Information Extraction (Open IE) merupakan topik yang telah diteliti cukup lama terutama pada bahasa Inggris. Pada pekerjaan open information extraction, dikembangkan banyak model baik yang melakukan ekstraksi relasi umum maupun yang berfokus pada relasi khusus, misalnya relasi numerik. Pada bahasa Indonesia, telah dikembangkan beberapa model open information extraction namun belum ada model yang berfokus pada relasi numerik. Padahal, sangat banyak informasi yang disajikan dalam bentuk numerik sehingga informasi tersebut belum dapat dimanfaatkan dengan baik. Karena itu, pada riset ini kami berfokus pada ekstraksi relasi numerik pada teks berbahasa Indonesia. Tantangan dalam pekerjaan ini adalah banyaknya relasi yang mungkin dari ekspresi numerik serta kategori ekspresi numerik yang beragam yaitu kardinal, kuantitas, persen, dan uang. Selain itu, banyak juga ekspresi numerik yang tidak menyatakan sebuah relasi numerik. Kontribusi riset ini adalah model ekstraksi atribut numerik dari teks berbahasa Indonesia. Walaupun pekerjaan ini adalah pilot task, model kami memperoleh hasil yang cukup baik dengan precision score 61.06%.

.....Open Information Extraction (Open IE) is a widely studied topic, especially in English. In open information extraction research, many models have been developed for general relation extraction or for specific relations, for instance, numerical relations. In Indonesian language, there have been some works on open information extraction models but none is focusing on numerical relations. With so much information given in numerical expression, it is so unfortunate if that information can not be used. Thus, in this research, we focus on numerical relation extraction in Indonesian texts. The challenge in this work is a huge number of relations that can be produced from numerical expressions as well as several categories of numerical expressions: cardinal, quantity, percent, and money. Furthermore, many numerical expressions do not express any numerical relation. Our contribution is a numerical relation extraction model from Indonesian texts. While this work is a pilot task, our model obtained a good result with precision score of 61.06%.