

PENGEMBANGAN SISTEM *DATA-TO-TEXT* (D2T) UNTUK MEMBANGKITKAN BERITA PADA DATA *UNSPECIFIC*

Oleh

Muhammad Ridwan — just.muhammadridwan@gmail.com

1403407

ABSTRAK

Sistem *Data-to-Text* menjadi salah satu pilihan untuk menerjemahkan data *non-linguistik* kedalam bentuk tekstual. Namun seiring dengan perkembangan teknologi, beragamnya bidang dari suatu data dan beragamnya pengguna menjadi salah satu fokus yang harus diperhatikan dalam pengembangan sistem *Data-to-Text*. Penelitian ini bertujuan untuk mengembangkan sistem *Data-to-Text* dengan masukan berupa data *unspecific*, sebagai solusi agar sistem *Data-to-Text* dapat menerima masukan berupa data dari bidang atau domain apapun, baik data tersebut memiliki identitas berupa informasi *header*, tipe data, *rule* ataupun tidak. Maka digunakan pendekatan *Fuzzy Rule* untuk menginterpretasikan data *unspecific* tersebut. Selain itu digunakan beberapa algoritma *Machine Learning* seperti *Gradient Descent*, dan analisis lainnya seperti *Exponential Smoothing*, *Knuth-Morris-Pratt*, *Statistical tools* dan *Pearson Correlation Coefficient*. Sistem yang dikembangkan dapat menghasilkan informasi berupa ringkasan data, informasi data terkini dan informasi prediksi. Pengembangan sistem dilakukan dalam bahasa pemrograman R dengan memanfaatkan beberapa *packages* yang tersedia. Eksperimen dilakukan dengan mengukur tingkat *Readability* dari berita yang dibangkitkan, *Computation Time*, dan membandingkan hasil dengan penelitian terkait. Hasil eksperimen menunjukkan bahwa informasi yang dihasilkan terbukti merepresentasikan data yang diberikan dan dapat dipahami oleh tingkat siswa pada tingkat sekolah dasar sekalipun, serta waktu komputasi cukup baik. Sistem ini mampu menghasilkan informasi berdasarkan data meteorologi, data klimatologi, data keuangan, dan data *time series* lainnya.

Kata Kunci— *Data-to-Text; Natural Language Processing, Natural Language Generation; Machine Learning; General purpose; Unspecific Corpora; Fuzzy Rule-based; Crisp Rule-based; Time-series Analysis; Exponential Smoothing; Linear Model; Gradient Descent; Kunth-morris-pratt; Pearson Correlation Coefficient*

DEVELOPMENT OF DATA-TO-TEXT (D2T) SYSTEMS TO GENERATE NEWS BASED ON UNSPECIFIC DATA

Arranged by

Muhammad Ridwan — just.muhammadridwan@gmail.com

1403407

ABSTRACT

The Data-to-Text system is an option for translating non-linguistic data into textual form. But along with the development of technology, the diverse fields of data and the variety of users have become one of the focuses that must be considered in the development of Data-to-Text systems. This study aims to develop a Data-to-Text system with input in the form of unspecific data, as a solution so that the Data-to-Text system can receive input in the form of data from any field or domain, both data has identity in the form of header information, data types, rules or not. Then the Fuzzy Rule approach is used to interpret the unspecific data. In addition, several Machine Learning algorithms such as Gradient Descent were used, and other analyzes such as Exponential Smoothing, Knuth-Morris-Pratt, Statistical tools and Pearson Correlation Coefficient. The system developed can produce information in the form of summary data, current data information and predictive information. The development of the system was written using the R programming language by utilizing several available packages. Experiments are carried out by measuring the level of Readability of the news generated, Computation Time, and comparing the results with related research. The experimental results show that the information produced is proven to represent the data provided and can be understood by the level of students at the elementary school level though, and computing time is quite good. This system is able to produce information based on meteorological data, climatological data, financial data, and other time series data.

Keywords — *Data-to-Text; Natural Language Processing, Natural Language Generation; Machine Learning; General purpose; Unspecific Corpora; Fuzzy Rule-based; Crisp Rule-based; Time-series Analysis; Exponential Smoothing; Linear Model; Gradient Descent; Kunth-morris-pratt; Pearson Correlation Coefficient*