

Interpretasi Teks Menggunakan Ontologi Berbasis Additive Fuzzy Spectral Clustering = Text interpretation using ontology based additive fuzzy spectral clustering

Ionia Veritawati, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20480665&lokasi=lokal>

Abstrak

Saat ini, data dalam bentuk teks semakin berlimpah pada berbagai domain dan media, baik media cetak maupun online. Penambahan kumpulan dokumen teks ini menyebabkan kemudahan akses suatu informasi atau pengetahuan yang ada pada teks semakin berkurang. Selain itu, informasi atau pengetahuan yang ada tersebut semakin sulit untuk diinterpretasi dan dipahami secara menyeluruh. Untuk itu diperlukan suatu cara untuk membantu mempermudah pemahaman suatu data teks. Hal ini dengan melakukan penggalian pengetahuan pada data teks yang melimpah melalui pemrosesan data yang tidak terstruktur (*text mining*), dengan mengembangkan metode interpretasi berbasis ontologi pada teks untuk memperoleh pengetahuan baru sebagai *state of the art*.

Dalam penelitian ini, dikembangkan beberapa teknik/metode. Pertama adalah pengembangan teknik *preprocessing* pada data teks (korpus) serta *key phrase extraction* menggunakan AST (*Annotated Suffix Tree*) untuk memperoleh *key phrase* (frasa kunci) dan frekuensi kemunculan. Kedua adalah pengembangan pemodelan ontologi sebagai basis pengetahuan pada suatu domain berupa relasi antar *key phrase* menggunakan *clustering* dan *Bayesian Network*. Ketiga adalah pengembangan metode *sparse clustering* pada data *sparse*, yaitu *is-FADDIS* (*iterative scaling Additive Fuzzy Spectral Clustering*) untuk proses pemilahan data teks, yang merupakan pengembangan dari metode *clustering FADDIS* (*Additive Fuzzy Spectral Clustering*) serta keempat adalah pengembangan metode *matching* dan *correlating* terhadap ontologi, sebagai teknik yang digunakan saat interpretasi teks.

Secara terintegrasi, pembangunan ontologi dari teks, dengan domain berita, dilakukan diawal dengan tahapan ekstraksi *key phrase*, *clustering* (*is-FADDIS*, opsional) dan *structure learning* untuk membentuk ontologi-tree. *Key phrase* sebagai konsep, menjadi *node* pada ontologi tersebut, yang menjadi basis pengetahuan domain. Tahapan berikutnya adalah melakukan interpretasi teks pada suatu teks input yang terdiri dari satu *key phrase* atau satu *cluster* menggunakan ontologi tersebut untuk mendapatkan pengetahuan baru. Interpretasi dilakukan dengan ontologi berasal dari teks dengan dua domain dan satu domain. Hasil interpretasi teks menggunakan ontologi berbasis *Additive Fuzzy Spectral Clustering* (*is-FADDIS*) ini dievaluasi menggunakan usulan *score relevansi*.

Pada teks input dengan satu *key phrase* sejumlah lima input yang diinterpretasi, hasilnya adalah 40% relevan, 40% kurang relevan dan 20% tidak relevan. Pada teks input satu *cluster* sejumlah dua input yang diinterpretasi, hasilnya adalah relevan. Nilai *score relevansi* yang relevan, secara empiris adalah lebih 0,3 dari skala 1, dan *score relevansi* yang didapat, ada yang mencapai 0,33. Dengan perbandingan hasil interpretasi melalui variasi teknik pada pembangunan ontologi,

didapatkan, penggunaan ontologi berbasis is-FADDIS untuk interpretasi teks, relatif pada penelitian ini belum memberikan hasil optimal. Dalam penggunaan teknik-teknik yang dikembangkan, metode ini memberikan keluaran interpretasi teks yang dapat membantu untuk mengolah informasi teks dalam jumlah tidak terlalu besar tetapi cepat.

.....Currently, the data in the form of text more abundant on various domains and media, both print and online media. The addition of this text document causes the ease of access to any information or knowledge contained in the text is reduced. In addition, the existing information or knowledge is increasingly difficult to interpret and understand comprehensively. For that background, the purpose of the research is to extract knowledge on abundant text data through the processing of unstructured data (text mining), by developing ontology-based interpretation method on text to gain a new knowledge as state of the art.

In this research, some technique/method were developed. The first is the development of preprocessing techniques on text data (corpus) and key phrase extraction using AST (Annotated Suffix Tree) to obtain key phrase and frequency of occurrence. The second is the development of ontology modeling as a knowledge base on a domain in the form of relationships between key phrases using Bayesian Network. The third is the development of sparse clustering method in sparse data, namely is-FADDIS (iterative scaling-Additive Fuzzy Spectral Clustering) for text grouping process, which is the addition of FADDIS clustering method (Additive Fuzzy Spectral Clustering) and the fourth is the development of matching and correlating method as a technique used at interpreting the text entered using ontology.

In an integrated manner, the ontology development of the text, with news domains, is done by processes include key phrase extraction, clustering (is-FADDIS, optional) and structure learning to form ontology-tree. Key phrase as a concept, being the node on the ontology, which becomes the domain knowledge base. The next step is to interpret the text on an input text consisting of a key phrase or a cluster using the ontology to gain new knowledge. Interpretation done with ontology comes from text with two domains and one domain. Text interpretation results using Fuzzy Spectral Clustering (is-FADDIS) based ontology is evaluated using relevancy scores.

In the input text with one key phrase a total of five interpreted inputs, the result is 40% relevant, 40% less relevant and 20% irrelevant. In one-cluster input text a number of two inputs are interpreted, the result is relevant. Relevant relevance score score, empirically more than 0.3 of scale 1, and score relevance obtained, some reaching 0.33. By comparing the results of interpretation through the variation of techniques on ontology development, it was found, the use of FADDIS-based ontology for textual interpretation, relative to this research has not provided optimal results. In the use of developed techniques, this method provides textual interpretation output that can help to process text information in quantities not too large but fastly.