



Visualisasi dan Pembobotan Jaringan Sitasi Makalah Ilmiah Dengan Algoritma Pagerank

Devy Rizky Andriyana, Agus Komarudin, Ridwan Ilyas*

Fakultas Sains dan Informatika, Jurusan Informatika, Universitas Jenderal Achmad Yani
Jl. Terusan Jenderal Sudirman, Cimahi, Jawa Barat, 40513

Email: ¹devyrizkyandriyana@gmail.com, ¹agus.komarudin@lecture.unjani.ac.id, ²ilyas@lecture.unjani.ac.id

Abstrak

Makalah merupakan karya tulis yang memuat pemikiran tentang suatu masalah atau topik tertentu yang ditulis secara sistematis disertai dengan analisis yang logis. Banyaknya makalah dapat dimanfaatkan untuk visualisasi berdasarkan sitasi makalah ilmiah tersebut. Untuk memecahkan masalah tersebut digunakan visualisasi 3D, visual tree collapse, visual tree, visual indented tree, visual network dan visual pagerank yang berfungsi menampilkan sitasi dari masing-masing makalah ilmiah. Visualisasi data didefinisikan sebagai berbagai jenis cara untuk membuat gambar, diagram atau animasi dengan tujuan untuk mengkomunikasikan sebuah pesan atau informasi. Makalah ilmiah seringkali banyak ditemukan di internet maupun di perpustakaan untuk berbagai macam judul makalah ilmiah, sitasi atau rujukan dapat ditemukan di setiap makalah ilmiah dan dapat diperoleh dengan mudah, namun untuk menampilkan seluruh sitasi pada makalah ilmiah dalam bentuk visualisasi tidak dapat dilakukan dengan mudah. Pembobotan dilakukan menggunakan algoritma PageRank untuk menentukan setiap bobot makalah ilmiah berdasarkan sitasi. Visualisasi jaringan sitasi makalah ilmiah berupa graf, dengan node merepresentasikan makalah peneliti dan edge merepresentasikan hubungan makalah ilmiah peneliti dengan makalah ilmiah yang lain berdasarkan sitasi makalah ilmiah. Penyajiannya menampilkan sitasi makalah ilmiah berdasarkan ukuran node dengan menunjukkan banyaknya sitasi makalah ilmiah tersebut. dapat disimpulkan bahwa semua proses pada sistem telah berjalan sesuai dengan rancangan, dan secara fungsional sistem visualisasi dan pembobotan jaringan sitasi makalah ilmiah telah sesuai dengan perancangan. Hasil penelitian berupa perangkat lunak visualisasi dan pembobotan makalah ilmiah telah berhasil diimplementasikan. Hasil yang diperoleh dari 20 naracoba, perangkat lunak ini menghasilkan ketertarikan pengguna secara visual 61,75%, kesesuaian data yang ditampilkan visualisasi 61,33% dan 54,66% membantu memudahkan pencarian makalah ilmiah dalam jumlah banyak.

Kata Kunci: Makalah Ilmiah, Network Visualization, Pagerank, Visualisasi, Sitasi.

1. PENDAHULUAN

Makalah merupakan karya tulis yang memuat pemikiran tentang suatu masalah atau topik tertentu yang ditulis secara sistematis dengan analisis yang logis dan objektif. Di setiap perguruan tinggi, makalah ditulis untuk memenuhi tugas terstruktur yang diberikan oleh pengajar. Tulisan ilmiah merupakan sebuah sarana komunikasi dalam bentuk tertulis yang sering dilakukan oleh para sarjana atau ilmuwan untuk merealisasikan ide-idenya, sehingga dapat dimengerti oleh semua orang. Dengan banyaknya makalah ilmiah menyulitkan bagi setiap orang yang ingin mencari referensi untuk makalah penelitiannya. Visualisasi dapat membantu memecahkan masalah ini.

Visualisasi data didefinisikan sebagai berbagai jenis cara untuk membuat gambar, diagram atau animasi dengan tujuan untuk mengkomunikasikan sebuah pesan atau informasi[1]. Terdapat berbagai banyak cara untuk memvisualisasikan setiap data menggunakan *Network Visualization* untuk membuat gambaran visual berdasarkan tabel *node* dan tabel *edge* yang akan digunakan sebagai data visualisasi[2], nilai probabilitas pada tiap makalah digunakan untuk menentukan kategori makalah[3].

Terdapat banyak penelitian mengenai visualisasi data yang terdapat pada penelitian sebelumnya terdapat dataset yang diterapkan pada *network visualization* untuk memvisualisasikan data pengidap penyakit AIDS yang memudahkan untuk mengetahui banyaknya pengidap penyakit AIDS[4]. Sitasi, berfungsi sebagai hyperlink, interlink publikasi ilmiah dan bentuk grafik ilmiah. Analisis kutipan membangun cara yang berguna untuk menganalisis dan memberi peringkat dokumen, penulis, atau jurnal[5].

PageRank adalah sebuah algoritma yang berfungsi menentukan situs laman mana yang lebih penting/populer. *Pagerank* merupakan salah satu fitur utama mesin pencari google. Cara kerja *pagerank* adalah sebuah situs akan semakin populer jika semakin banyak situs lain yang mengarah ke situsnya, dengan asumsi isi/content situs tersebut lebih berguna dari isi/content situs lain. Peringkat halaman dihitung dengan skala 1-10[6]. Model yang mendasari *PageRank* sangat terkait dengan indeks kutipan yang digunakan dalam literatur ilmiah untuk mengevaluasi pentingnya publikasi[7].

2. METODOLOGI PENELITIAN

Metode penelitian terdiri dari lima tahap yaitu, pengambilan dan pengumpulan data, ekstraksi dokumen dan perhitungan bobot, perancangan perangkat lunak dan pembuatan perangkat lunak, pengujian dan evaluasi, pelaporan dan publikasi ilmiah.

Tahap pertama yaitu pengambilan data dilakukan dengan mengumpulkan data dari kumpulan repository makalah ilmiah kemudian dipilah menjadi beberapa komponen. Tahap kedua ekstraksi dokumen dan perhitungan bobot dilakukan dengan memilah kumpulan makalah ilmiah menjadi beberapa komponen yaitu judul, abstrak, peneliti, tahun dan referensi, kemudian perhitungan bobot pada setiap makalah ilmiah dengan menggunakan algoritma *PageRank*, Tahap ketiga visualisasi yang diterapkan yaitu tiga dimensi, tree dan visualisasi hasil pembobotan *PageRank*, Tahap keempat perancangan perangkat lunak dan pembuatan sistem, Tahap kelima pengujian dan evaluasi.

2.1 Pengumpulan Data

Tahap pertama pengambilan data pada visualisasi makalah ilmiah menjelaskan proses data dikumpulkan dari repository makalah ilmiah. Makalah ilmiah yang diambil berupa satu *file pdf* dan kemudian dipilah menjadi beberapa komponen.

2.2 Ekstraksi Dokumen dan Perhitungan Bobot

Pada tahap ekstraksi dokumen data yang sudah diambil dilakukan *preprocessing* yang merupakan tahap dimana data yang akan digunakan dipilah menjadi beberapa komponen yaitu judul, abstrak, peneliti, tahun, dan referensi. Tahap selanjutnya adalah Pembobotan pada setiap makalah ilmiah dilakukan menggunakan algoritma *PageRank* untuk menentukan bobot dari tiap makalah ilmiah, terdapat beberapa tahap untuk mendapatkan nilai bobot, pada proses *PageRank* ini akan membentuk distribusi probabilitas, sehingga rata-rata nilai *PageRank* untuk seluruh sitasi adalah 1. Rumusan algoritma *PageRank* sebagai berikut:

$$PR(A) = (1 - d) + d \left(\frac{PR(T1)}{C(T1)} + \dots + \frac{PR(T1)}{C(T1)} \right) \quad (2.1)$$

dimana,

- PR(Tn) = Nilai PageRank dari halaman Tn.
C(Tn) = Jumlah link yang keluar dari halaman Tn.
d = Damping Factor

2.3 Tahap Visualisasi

Pada tahap ini hasil dari pra proses akan dilakukan tahap visualisasi atau perubahan ke dalam bentuk graf yang terdiri dari visualisasi 3D, visualisasi tree collapse, visualisasi tree, visualisasi indented tree, visualisas network dan visualisasi pagerank.

3. ANALISA DAN PEMBAHASAN

Analisis data masukan pada visualisasi makalah ilmiah merupakan kumpulan data makalah ilmiah dari repository makalah ilmiah, kemudian dipilah menjadi beberapa bagian saja seperti judul, abstrak, peneliti, tahun, dan referensi. Contoh data makalah ilmiah yg sudah dipilah seperti pada Tabel 1.

Tabel 1. Konten Makalah Ilmiah

Judul	Evaluating Word Embeddings Using a Representative Suite of Practical Tasks
Abstrak	Word embeddings are increasingly used in natural language understanding tasks requiring sophisticated semantic information. However, the quality of new embedding methods is usually evaluated based on simple word similarity benchmarks. We propose evaluating word embeddings in vivo by evaluating them on a suite of popular downstream tasks. To ensure the ease of use of the evaluation, we take care to find a good point in the tradeoff space between (1) creating a thorough evaluation – i.e., we evaluate on a diverse set of tasks; and (2) ensuring an easy and fast evaluation – by using simple models with few tuned hyperparameters. This allows us to release this evaluation as a standardized script and online evaluation, available at http://veceval.com/ .
Peneliti	Neha Nayak, Gabor Angeli, dan Christopher D. Manning
Tahun	2016
Referensi	Ini berisi rujukan pada makalah ilmiah

Perhitungan pagerank menggunakan tujuh buah makalah ilmiah untuk dijadikan contoh dengan nilai awal pagerank setiap node PR(Tn)=10, dan nilai damping factor d=0.85 dan kemudian di visualisasikan. Seperti berikut :

$$\begin{aligned} 1. \quad PR(A) &= (1-0.85) + 0.85 (10/37+10/1+10/1+10/1+10/1) \\ &= (0.15) + 0.85(10/30) \end{aligned}$$

$$PR(A) = 34.38$$

$$\begin{aligned} 2. \quad PR(B) &= (1-0.85) + 0.85(10/30) \\ &= (0.15) + 0.85 (10/30) \end{aligned}$$

$$PR(B) = 0.43$$

$$\begin{aligned} 3. \quad PR(B) &= (1-0.85) + 0.85(10/51) \\ &= (0.15) + 0.85 (10/51) \end{aligned}$$

$$PR(B) = 0.32$$

$$\begin{aligned} 4. \quad PR(C) &= (1-0.85) + 0.85(10/30+10/1) \\ &= (0.15) + 0.85 (10/30+10/1) \end{aligned}$$



$$\begin{aligned} \text{PR(C)} &= 8.93 \\ 5. \text{ PR(D)} &= (1-0.85) + 0.85(10/32) \\ &= (0.15) + 0.85 (10/32) \\ \text{PR(D)} &= 0.41 \\ 6. \text{ PR(D)} &= (1-0.85) + 0.85(10/21+10/1) \\ &= (0.15) + 0.85 (10/32+10/1) \\ \text{PR(D)} &= 9.05 \\ 7. \text{ PR(D)} &= (1-0.85) + 0.85(10/46) \\ &= (0.15) + 0.85 (10/46) \\ \text{PR(D)} &= 0.33 \end{aligned}$$

4. KESIMPULAN

Sistem Visualisasi dan Pembobotan Jaringan Sitasi Makalah Ilmiah merupakan sistem yang dapat mempermudah dalam memvisualisasikan dan melihat setiap makalah ilmiah lengkap beserta sitasi dari masing-masing makalah ilmiah. Berdasarkan analisis sistem berjalan dan analisis fungsional sistem yang telah dilakukan sebelumnya, maka dapat disimpulkan bahwa sistem ini dapat membantu dalam menangani proses pembobotan untuk setiap makalah ilmiah beserta sitasi dari masing-masing makalah ilmiah, guna memudahkan mahasiswa atau orang yang hendak mencari penelitian–penelitian yang sudah dilakukan atau yang sudah dikerjakan sebagai referensi, tidak memakan waktu yang lama dalam mencari dokumen-dokumen yang diinginkan karena sudah tersedia dalam bentuk visualisasi. Hasil penelitian berupa perangkat lunak visualisasi dan pembobotan makalah ilmiah telah berhasil diimplementasikan. Hasil yang diperoleh dari 20 naracoba, perangkat lunak ini menghasilkan ketertarikan pengguna secara visual 61,75%, kesesuaian data yang ditampilkan visualisasi 61,33% dan 54,66% membantu memudahkan pencarian makalah ilmiah dalam jumlah banyak..

REFERENCES

- [1] T. Ernawati, “Analisis Visualisasi Data Keamanan Jaringan,” *Jurnal Teknologi*, vol. 5, no. 1, pp. 53–61, 2012.
- [2] A. Arfina and H. Khotimah, “Visualisasi Co-Authorship Peneliti IPB Menggunakan Metode Chi Visualization of Co-Authorship between IPB ’ s Researcher Using Chi ’ s Method,” *Ilmu Komput. Agri-Informatika*, vol. 5, pp. 31–39, 2017.
- [3] J. N. Apriliana, Natalis Ransi, “Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode Naïve Bayes Classifier,” *Semant. Vol.3, No.2, Jul-Des 2017*, vol. 3, no. 2, pp. 329–343, 2017.
- [4] N. A. Abdul Rahim, S. Sulaiman, S. Z. Mohd Hasim, and N. B. Ahmad, “Visualization And Centrality Measurement Of Social Network Analysis,” *J. Teknol.*, vol. 78, pp. 65–74, 2016.
- [5] H. Small, “Co-citation in the Scientific literature: A New Measure of the Relationship Between Two Documents,” vol. 24, no. 4, pp. 28–31, 1973.
- [6] Sujito and D. Wahyuningsih, “Analisis popularitas laman pemerintah daerah di wilayah malang raya berbasis perangkat pemerinkatan dalam jaringan,” vol. 5, pp. 13–19, 2015.
- [7] M. Bianchini, M. Gori, and F. Scarselli, “Inside PageRank,” vol. 5, no. 1, pp. 92–128, 2005.