



PROSIDING

SEMIRATA 2014

Bidang MIPA BKS-PTN-Barat

"Integrasi sains MIPA untuk mengatasi masalah pangan, energi, kesehatan, reklamasi, dan lingkungan"

IPB International Convention Center dan Kampus IPB Baranangsiang, 9-11 Mei 2014

BUKU 6

MATEMATIKA, FISIKA, KIMIA, BIOLOGI, STATISTIKA, KOMPUTER, STEM, GEOFISIKA DAN METEOROLOGI

Diterbitkan oleh: Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Pertanian Bogor



ISBN 978-602-70491-0-9

VISUALISASI SILSILAH VARIETAS KEDELAI MENGGUNAKAN ALGORITME *DIRECTED ACYCLIC GRAPH*

VISUALIZATION OF SOYBEAN VARIETIES PEDIGREE USING DIRECTED ACYCLIC GRAPH ALGORITHM

Ayu Riza Bestary¹, Firman Ardiansyah², Desta Wirnas³

¹Departemen Ilmu Komputer, Fakultas MIPA, Institut Pertanian Bogor, Bogor
ayu.bestary10@apps.ipb.ac.id, Jl. Meranti wing 20, Lv. V, IPB Darmaga, Bogor 16680 Telepon:
(0251) 8625584

²Departemen Ilmu Komputer, Fakultas MIPA, Institut Pertanian Bogor, Bogor

³Departemen Agronomi dan Hortikultura, Fakultas Pertanian, Institut Pertanian Bogor

ABSTRACT

The pedigree data of soybean varieties in Indonesia is still difficult to be traced. Soybean varieties genealogy is useful in recognizing the lineage of soybean varieties that can simplify the process of determining what varieties will be used for further soybean breeding. Graph-based approach to information visualization is used to represent the pedigree varieties. 121 varieties of soybean are represented by nodes and the relationship between varieties are represented by edges. A variety that produces many offspring is represented by size of node, while the plant breeding methods are represented by colors. This research applies Directed Acyclic Graph algorithm to clarify the form of the graph visualization by forming the edges to always point downwards. The visualization was implemented in a web based application where users can interact and explore the pedigree data to gain knowledge about soybean varieties pedigree.

Keywords: Information visualization, Soybean varieties pedigree, Directed Acyclic Graph.

ABSTRAK

Data silsilah varietas kedelai hasil pemuliaan di Indonesia saat ini masih sulit untuk ditelusuri. Penelusuran silsilah varietas kedelai berguna untuk mengetahui garis keturunan varietas kedelai sehingga dapat mempermudah proses penentuan varietas apa saja yang akan digunakan untuk kegiatan pemuliaan tanaman kedelai selanjutnya. Pendekatan visualisasi informasi berbasis graf digunakan untuk merepresentasikan silsilah varietas. Terdapat 121 varietas kedelai yang direpresentasikan dengan *nodes* dan keterhubungan setiap varietas direpresentasikan dengan *edges*. Sebuah varietas yang menghasilkan banyak keturunan direpresentasikan dengan besarnya ukuran *node* sedangkan metode pemuliaan tanaman yang digunakan untuk menghasilkan varietas direpresentasikan melalui warna yang berbeda. Algoritme *Directed Acyclic Graph* digunakan untuk memperjelas bentuk graf visualisasi yang dihasilkan dengan cara membentuk *edge* yang selalu mengarah ke bawah. Visualisasi diimplementasikan dalam aplikasi berbasis web sehingga pengguna dapat berinteraksi dan mengeksplorasi data silsilah tersebut untuk mendapatkan pengetahuan mengenai silsilah varietas kedelai.

Katakunci : Visualisasi Informasi, Silsilah Varietas Kedelai, *Directed Acyclic Graph*

PENDAHULUAN

Kedelai (*Glycine max* L.) merupakan salah satu komoditas pangan penting di Indonesia setelah padi dan jagung. Kebutuhan kedelai terus meningkat seiring dengan pertumbuhan jumlah penduduk dan kebutuhan bahan industri olahan pangan seperti

tahu, tempe, kecap, susu kedelai, tauco, *snack*, dan sebagainya [1]. Untuk memenuhi kebutuhan kedelai dalam negeri, perlu adanya peningkatan produktivitas kedelai di Indonesia. Dalam upaya meningkatkan produktivitas usaha tani kedelai di Indonesia sangat diperlukan ketersediaan varietas unggul dan benihnya yang bermutu tinggi, di samping penggunaan teknik-teknik budidaya lainnya [9].

Cara mendapatkan varietas unggul adalah dengan pemuliaan tanaman, baik melalui seleksi maupun melalui persilangan-persilangan dengan varietas sebelumnya [11]. Oleh karena itu varietas-varietas baru tersebut merupakan keturunan dari varietas sebelumnya sehingga menghasilkan silsilah keturunan varietas kedelai. Data keturunan varietas kedelai yang tersedia masih sulit untuk ditelusuri. Penelusuran silsilah varietas kedelai seringkali dilakukan untuk mengetahui varietas induk manakah yang menurunkan sifat pada varietas kedelai tertentu dan untuk mencari informasi apakah dua varietas yang berbeda berasal dari keturunan yang sama.

Data keturunan varietas kedelai ini dapat dimodelkan dengan visualisasi informasi menggunakan struktur berbasis silsilah. Visualisasi data silsilah memungkinkan pemulia dan peneliti untuk membuat keputusan yang lebih efisien berkaitan dengan persilangan tanaman dalam rangka mengembangkan varietas tanaman baru [12]. Penelitian ini merepresentasikan data silsilah varietas kedelai dengan visualisasi berbentuk graf. Algoritme *Directed Acyclic Graph* diterapkan untuk menghasilkan graf berarah di mana tidak ada *edge* yang dimulai dan diakhiri pada *node* yang sama [10]. *Directed Acyclic Graph* memperjelas bentuk graf dengan membentuk *edge* yang selalu mengarah ke bawah.

Tujuan dan Manfaat Penelitian

Tujuan penelitian ini adalah memodelkan silsilah varietas kedelai guna memberikan kemudahan penelusuran dan pencarian informasi mengenai keturunan varietas unggul kedelai di Indonesia. Hasil penelitian ini diharapkan dapat dimanfaatkan oleh para peneliti dan pemulia kedelai dalam mencari informasi terkait keturunan hasil seleksi dan persilangan varietas kedelai untuk keperluan mengembangkan dan memuliakan varietas kedelai di Indonesia.

Ruang Lingkup Penelitian

Ruang lingkup penelitian ini adalah: (1) Visualisasi diterapkan untuk memodelkan silsilah varietas hasil pemuliaan kedelai di Indonesia, (2) Pengguna dapat melakukan pencarian berupa nama varietas kedelai, (3) Aplikasi yang dihasilkan masih bersifat statis, belum dapat menangani penambahan data ke dalam *database*. (4) Data utama yang digunakan adalah data varietas unggul kedelai yang digunakan oleh Kumalasari [8] dan Balitkabi [2].

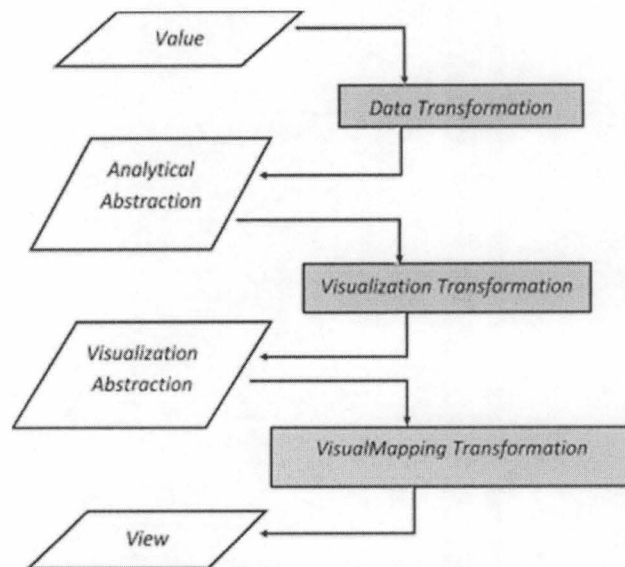
METODE PENELITIAN

Penelitian ini menggunakan kerangka metode untuk membangun aplikasi visualisasi informasi yang diperkenalkan oleh Chi [5]. Metode ini menggunakan operator *value* (data mentah) yang akan ditransformasikan menggunakan empat tahap utama, yaitu *Value*, *Analytical Abstraction*, *Visualization Abstraction*, dan *View*. Di antara masing-masing tahapan tersebut terdapat 3 proses utama, yaitu *Data Transformation*, *Visualization Transformation*, dan *Visual Mapping Transformation*. Tahapan penelitian dapat dilihat pada Gambar 1.

Value

Tahap *Value* merupakan tahap pengumpulan data. Data yang digunakan adalah

data varietas unggul kedelai yang digunakan oleh Kumalasari [8] dengan beberapa tambahan data sebagai pelengkap dari Balitkabi [2], BB Biogen [3], dan varietas baru yang dilepas pada tahun 2013 [6]. Data yang diperoleh sudah dalam bentuk *database* dengan puluhan atribut terkait deskripsi varietas unggul kedelai.



Gambar 11 Metode penelitian [5]

Data Transformation

Data Transformation merupakan proses transformasi tahap *Value* menjadi tahap *Analytical Abstraction*. Pada proses ini data diolah dengan praproses data agar diperoleh data yang sudah memperoleh informasi. Terdapat 2 tahap praproses data yang dilakukan, yaitu pemangkasan atribut dan pembersihan data.

Pemangkasan Atribut

Atribut yang dibutuhkan dalam penelitian ini hanya atribut yang memiliki keterkaitan dengan silsilah varietas kedelai berupa nama varietas yang menjadi tetua dan anakan, serta metode pemuliaan tanaman yang digunakan untuk menghasilkan varietas.

Pembersihan Data

Salah satu tujuan pembersihan data (*data cleaning*) adalah untuk mengatasi *missing values* atau nilai data yang hilang [7]. Pilihan untuk mengatasi *missing values* dapat dilakukan dengan dua cara, yaitu dengan mengisi nilai data yang kosong, atau dengan menghapus *instance* yang memiliki nilai data yang kosong. Kedua cara tersebut diterapkan pada data yang akan digunakan. *Missing values* yang masih memungkinkan untuk dilengkapi akan diisi dengan data-data tambahan dari Balitkabi [2]. Sedangkan *missing values* yang tidak dapat dilengkapi akan dihapus.

Analytical Abstraction

Tahap ini merupakan hasil dari praproses data yang belum dipetakan namun mencakup semua informasi dari data mentah yang akan divisualisasikan [5]. Pada tahap ini diperoleh 5 atribut hasil praproses data yang relevan dengan informasi yang dibutuhkan untuk visualisasi.

Visualization Transformation

Visualization Transformation adalah proses mengambil nilai-nilai dari *Analytical*

Abstraction untuk diproses lebih lanjut membentuk *Visualization Abstraction* [5]. Proses utama pada tahap ini adalah perancangan visualisasi dan pengolahan data yang akan dipetakan ke dalam visualisasi.

Perancangan Visualisasi

Perancangan visualisasi pewarnaan dan bentuk graf yang akan ditampilkan. Visualisasi dirancang agar dapat merepresentasikan silsilah varietas kedelai dan mampu menangani proses pelacakan keturunan varietas kedelai.

Pengolahan Data yang Akan Dipetakan

Pengolahan data dilakukan agar data dapat disesuaikan untuk menghasilkan visualisasi seperti yang sudah dirancang. Pada proses ini dilakukan pencarian induk varietas kedelai, sehingga dapat ditemukan keterhubungan antara tetua varietas dengan keturunannya.

Visualization Abstraction

Visualization Abstraction merupakan proses pemetaan data menjadi suatu visualisasi dengan menggunakan minimal satu teknik visualisasi [5]. Pada proses ini dihasilkan tabel *nodes* yang berisi informasi nama varietas kedelai beserta induknya dan tabel *edges* yang mengandung informasi keterhubungan antar varietas.

Visual Mapping Transformation

Visual Mapping Transformation adalah proses pengambilan informasi yang telah berbentuk format visual dan merepresentasikannya dalam bentuk tampilan grafis [5]. Pada proses ini diterapkan algoritme *Directed Acyclic Graph* menggunakan perangkat lunak Gephi versi 0.8.2. Gephi merupakan *open source software* untuk eksplorasi dan manipulasi *network* [4]. Pada proses ini juga diterapkan pelabelan *nodes* berupa nama varietas kedelai dan pewarnaan *nodes* berdasarkan metode pemuliaan tanaman.

View

View merupakan hasil akhir *Visualization Mapping* dimana *user* sudah dapat melihat dan menginterpretasikan grafis visual yang ditampilkan di layar. Pada tahap ini pengguna dapat melakukan interaksi terhadap visualisasi seperti memperbesar atau memperkecil ukuran graf, menghilangkan bagian *nodes* dan *edges* yang tidak berhubungan dengan penelusuran, dan melakukan pencarian *nodes*. Tahap ini dilakukan implementasi fail graf dengan format GEXF (*Graph Exchange XML Format*) ke dalam aplikasi berbasis web yang dibangun menggunakan *Javascript*, *HTML*, dan *CSS*. Perangkat lunak yang digunakan untuk membuat aplikasi web ini adalah Notepad++, WampServer, dan Google Chrome *browser*.

HASIL DAN PEMBAHASAN

Value

Penelitian ini menggunakan data varietas unggul kedelai yang digunakan oleh Kumalasari [8] yang terdiri dari 58 *records* varietas unggul kedelai dan 36 atribut. Data tersebut kemudian dilengkapi dengan data yang dimiliki oleh Badan Penelitian Tanaman Aneka Kacang dan Umbi berupa buku Deskripsi Varietas Unggul Kedelai 1918–2012 [2]. Selain itu ditambahkan juga 2 data varietas baru yang dilepas pada tahun 2013 [6]. Sedangkan untuk mengetahui negara asal varietas digunakan data milik Balai Besar Litbang Bioteknologi & Sumber Daya Genetik Pertanian berupa Buku Katalog Plasma Nutfah Tanaman Pangan 2010-BB-Biogen [3]. Setelah melalui pelengkapan data, didapatkan hasil sebanyak 123 *records* varietas.

Data Transformation

Pemangkasan Atribut

Database yang diperoleh memiliki 36 atribut dan hanya 5 atribut yang berkaitan dengan silsilah kedelai. Sehingga dari 36 atribut tersebut hanya digunakan 5 atribut, yaitu 'id kedelai', 'nama kedelai', 'induk1', 'induk2', dan 'metode pemuliaan'. Atribut 'induk1', 'induk2', dan 'metode pemuliaan' merupakan atribut yang dipecah secara manual dari satu atribut di dalam *database* yaitu 'asal varietas'.

Pembersihan Data

Dari 123 *records* di dalam *database*, terdapat 29 *records* yang memiliki nilai yang kosong atau *missing values*. *Missing values* tersebut diatasi dengan melengkapi data menggunakan data Balitkabi [2]. Pada *database* juga ditemukan 2 *records* varietas yang tertulis berasal dari keturunan 16 tetua. 16 tetua yang dimaksud oleh data tersebut tidak memungkinkan untuk ditelusuri sehingga harus dihapus dari *database*.

Analytical Abstraction

Pada tahap *Analytical Abstraction* telah didapatkan hasil dari proses *Data Transformation*, yaitu berupa data yang sudah melalui praproses data dan pembersihan data. Data yang dihasilkan terdiri dari 121 *records* berupa varietas kedelai dan 7 *records* berupa negara asal kedelai yang diambil dari data BB Biogen [3].

Visualization Transformation

Perancangan Visualisasi

Visualisasi ditampilkan dalam bentuk graf. Graf dapat digunakan untuk memvisualisasikan hubungan antara dua objek. Selain itu, graf dapat menyediakan pemodelan yang tepat untuk kompleksitas data program pemuliaan tanaman [12]. Varietas kedelai dan negara asal varietas divisualisasikan dalam bentuk *nodes*, sedangkan keterhubungan antara dua varietas divisualisasikan dalam bentuk *edges* (garis yang menghubungkan dua *nodes*). Bentuk visual lain yang ditentukan pada tahap ini, yaitu:

- 1 Ukuran *nodes* menunjukkan besarnya bobot *nodes* yang ditentukan oleh seberapa banyak *nodes* lain yang terhubung dengan *nodes* tersebut. Jadi semakin besar ukuran suatu *nodes* maka bobotnya semakin besar dan semakin banyak *nodes* lain yang terhubung dengannya.
- 2 Pewarnaan *nodes* ditentukan berdasarkan metode pemuliaan tanaman yang dilakukan untuk menghasilkan varietas kedelai tersebut.

Pengolahan Data yang Akan Dipetakan

Proses selanjutnya adalah pengolahan data yang akan dipetakan pada proses *Visual Mapping Transformation*. Proses ini dilakukan dengan membuat tabel *nodes* dan *edges*. Pada tabel *nodes*, nama varietas kedelai dan negara asal varietas kedelai diberi id untuk direpresentasikan sebagai *nodes*. Setelah itu dilakukan penghitungan bobot *nodes* berdasarkan banyaknya keterhubungan suatu *node* dengan *node* lain. Bobot tersebut disimpan dalam atribut *degree*. Tabel *nodes* mengikutsertakan atribut metode untuk digunakan dalam pewarnaan *nodes*. Tabel *edges* dibuat untuk menunjukkan keterhubungan antar *nodes*. Keterhubungan ini ditunjukkan dengan menghubungkan *nodes* asal (*source*) dengan *nodes* tujuan (*target*). *Source* berisi id varietas kedelai yang menjadi tetua, sedangkan *target* berisi id varietas yang menjadi keturunannya. Selain itu pada tabel *edges* terdapat atribut *type* yang menunjukkan bahwa graf yang akan dihasilkan adalah graf berarah.

Visualization Abstraction

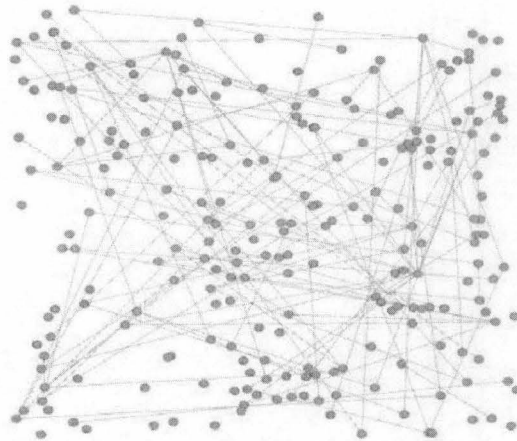
Pada tahap *Visualization Abstraction* telah didapatkan tabel *nodes* dan tabel

edges. Terdapat 129 *records* yang dihasilkan pada tabel *nodes* yang terdiri dari 121 varietas kedelai, 7 negara asal varietas kedelai, dan 1 *nodes* berlabel 'Unknown' untuk varietas yang tidak diketahui negara asalnya. Pada tabel *edges* terdapat 157 *records* yang menghubungkan setiap *nodes*.

Visual Mapping Transformation

Data yang didapatkan dari tahap *Visualization Abstraction* telah siap untuk dipetakan ke dalam bentuk visualisasi. Representasi informasi dilakukan dengan beberapa proses, yaitu:

- 1 Tabel *nodes* dan *edges* diimpor ke dalam *software* Gephi dan kemudian dihasilkan graf berarah dengan *random layout*. Graf yang dihasilkan dengan *random layout* masih sulit untuk dipahami dan dieksplorasi. Hasil graf dengan *random layout* dapat dilihat pada Gambar 2.



Gambar 12 Graf dengan *random layout*.

- 2 Graf diperbaiki dengan menggunakan algoritme *Directed Acyclic Graph* (DAG) agar pola dari silsilah varietas kedelai terlihat.
- 3 Pewarnaan *nodes* varietas pada graf dilakukan berdasarkan metode pemuliaan tanaman. Pada data terdapat 5 metode pemuliaan tanaman, yaitu persilangan, seleksi massa/galur murni, mutasi, introduksi dan *unknown*. Sedangkan *nodes* dengan label negara akan diberikan warna tersendiri.
- 4 Penentuan ukuran *nodes* dilakukan dengan pembobotan di mana semakin banyak *edge* yang terhubung pada suatu *node*, maka ukuran *node* tersebut akan semakin besar. *Range* bobot *nodes* yang dihasilkan sebesar 1 sampai 18.
- 5 Proses terakhir adalah pemberian label pada graf. Label dapat diterapkan pada *nodes* maupun *edges*. Penelitian ini hanya memberikan label pada *nodes* saja dan tidak memberikan perhatian khusus pada *edges*. Graf yang telah diperbaiki dan diberi label ditunjukkan pada Gambar 3.

PUSTAKA

- [1] [Balitbang Deptan] Badan Penelitian dan Pengembangan Departemen Pertanian. 2005b. *Prospek Agribisnis dan Arah Pengembangan Agribisnis Kedelai*. Jakarta (ID): Balitbang Deptan.
- [2] [Balitkabi] Badan Penelitian Tanaman Aneka Kacang dan Umbi. 2012. Deskripsi Varietas Unggul Kedelai 1908-2012 [Internet]. Jakarta (ID): Balitkabi. [diunduh 2013 Nov 26]. Tersedia pada <http://balitkabi.litbang.deptan.go.id/images/PDF/vub%20kedelai.pdf>.
- [3] kedelai.pdf.
- [4] [BB Biogen] Balai Besar Litbang Bioteknologi & Sumber Daya Genetik Pertanian. 2013. Katalog Plasma Nutfah Kedelai (*Glycine max*) [Internet]. Bogor (ID): BB Biogen. [diunduh 2013 Apr 15]. Tersedia pada <http://biogen.litbang.deptan.go.id/wp/wp-content/plugins/download-monitor/download.php?id=paspor+kedelai.pdf>.
- [5] Bastian M, Heyman S, Jacomy M. 2009. Gephi : an open source software for exploring and manipulatif Networks. *International AAAI Conference on Weblogs and Social Media*[Internet]; 2009 Mar. [tempat pertemuan tidak diketahui]. North America(OA):Third International AAAI Conference on Weblogs and Social Media.
- [6] Chi EH. 1999. A framework for information visualization spreadsheets [disertasi].Minneapolis (US): University of Minnesota.
- [7] Faisal. 2013. Gamasugen 1 dan 2, Varietas Baru Kedelai Unggul [Internet]. Jakarta (ID): Poskota News. [diunduh 2013 Mar 16]. Tersedia pada <http://poskotanews.com/2013/12/02/gamasugen-1-dan-2-varietas-baru-kedelai-unggul/>
- [8] Han J, Micheline K, Jian P. 2013. *Data mining: concepts and techniques*. Ed ke-3. Massachusetts (US): Morgan Kaufmann.
- [9] Kumalasari T. 2013. Sistem Pemilihan Varietas Unggul Kedelai (*Glycine Max L. Merril*) [Tesis]. Bogor (ID): Institut Pertanian Bogor
- [10] Nugraha A. 2008. Keragaan Galur-galur Kedelai (*Glycine max* (L.) merr.) F8 Berdaya Hasil Tinggi di KP BB Biogen, Cikeumeuh [Skripsi]. Bogor (ID): Institut Pertanian Bogor.
- [11] Papamanthou C. 2004. *Depth First Search and Directed Acyclic Graph*. Heraklion (GR): University of Crete.
- [12] Rukmana R, Yuniarsih Y. 1995. *Kedelai Budidaya dan Pasca Panen*. Jakarta (ID): Kanisius.
- [13] Shaw PD, Kenney J, Graham M, Milne I, Marshall DF. 2013. *Visualizing Genetic Transmission Patterns in Plant Pedigrees*. *Plant and Animal Genome XX Conference (January 14-18, 2012)*. Skotlandia (UK): Plant and Animal Genome.
- [14] Yi JS, Kang Y, Stasko JT, Jacko JA. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans Visualization and Computer Graphics*. 13(6):1224-1231.doi: 10.1109/TVCG.2007.70515.