

Pembangunan *Biodiversity Informatics* Genetika Tumbuhan Berbasis Ontologi

Bayu Santoso(G64134017)*, Yeni Herdiyeni

Abstrak/*Abstract*

Indonesia memiliki lebih dari 32.000 spesies tumbuhan. Saat ini hutan Indonesia mengalami kerusakan dan kepunahan. Oleh karena itu, diperlukan upaya untuk melestarikan tumbuhan. Salah satu cara untuk melestarikan tumbuhan adalah dengan cara mengenali spesies tumbuhan tersebut. Berdasarkan hal tersebut maka muncul bidang baru dalam pengumpulan informasi tumbuhan yang bernama *biodiversity informatics*. Metode pemodelan data yang dapat menangani sistem berbasis inferensi adalah ontologi. Ontologi dapat diterapkan pada web semantik. Penelitian ini akan mengembangkan sistem web semantik yang memberikan informasi genetika tumbuhan. Selain itu juga merumuskan masalah bagaimana melakukan inferensi pengetahuan dari sistem web semantik gen tumbuhan dengan sistem web semantik yang lain.

Kata Kunci

biodiversity informatics; tumbuhan; ontologi; web semantik

***Alamat Email:** bayusantoso.mail@gmail.com

PENDAHULUAN

Latar Belakang

Indonesia memiliki lebih dari 32.000 spesies tumbuhan (Bappenas 2003). Saat ini hutan Indonesia mengalami kerusakan dan kepunahan (Zuhud 2008). Oleh karena itu, diperlukan upaya untuk melestarikan tumbuhan. Salah satu cara untuk melestarikan tumbuhan adalah dengan cara mengenali tumbuhan tersebut. Biodiversity Informatics merupakan upaya untuk membuat sumber informasi keanekaragaman hayati global tersedia dalam format digital yang efisien, dan untuk mengembangkan alat yang efektif dalam menganalisis dan memahami data tersebut (Gilman et al. 2009). Informasi yang dapat diperoleh dari *biodiversity informatics* adalah informasi mengenai taksonomi, gambar tumbuhan, lingkungan, dan DNA tumbuhan.

Implementasi dari *biodiversity informatics* sudah menghasilkan beberapa sistem yang menyediakan informasi mengenai tumbuhan. Integrated Taxonomic Information System (ITIS) dan Global Biodiversity Information Facility (GBIF) menyediakan informasi yang luas tentang tumbuhan. Proses identifikasi dan pengelolaan informasi keanekaragaman hayati tersebut memerlukan sistem yang terpadu dan holistik dengan menggunakan IPTEKS komputer yang berkembang pesat saat ini (Herdiyeni et al. 2013). Upaya pemanfaatan IPTEKS yang telah dilakukan seperti diantaranya pembangunan sistem IPB

Biodiversity Informatics (IPBiotics) untuk pengelolaan informasi keanekaragaman hayati sumber daya alam Indonesia. Sistem BI tersebut berguna meningkatkan pengelolaan pengetahuan (knowledge management), eksplorasi, analisis, sintesis dan interpretasi data keanekaragaman hayati mulai dari level genomik, level spesies sampai dengan level ekosistem (Herdiyeni et al. 2013). Pada pengembangan selanjutnya, sistem IPBiotics didesain agar dapat melakukan inferensi pengetahuan. Sistem yang ada saat ini masih menggunakan model basis data relasional. Permasalahannya adalah model basis data relasional kurang sesuai diterapkan pada sistem berbasis inferensi (Laallam et al. 2013).

Ontologi adalah metode yang digunakan untuk merepresentasikan ide, fakta dan lain sebagainya, yang digunakan untuk mendefinisikan hubungan dan klasifikasi dari pengetahuan tertentu (Jepsen 2009). Ontologi dapat menentukan kelas, hubungan, fungsi dan objek lain (DiLecce V 2008). Selain itu, model ontologi lebih sesuai diterapkan pada web semantik dibandingkan dengan model basis data relasional (Laallam et al. 2013).

Penelitian dengan menggunakan ontologi mengenai tumbuhan sudah banyak dilakukan, seperti penelitian tentang ontologi yang digunakan untuk menganalisis hubungan tumbuhan obat dengan istilah medis yang standar (Vadivu G 2012). Penelitian yang terkait dengan ontologi gen juga sudah pernah dilakukan untuk menghasilkan

data gen yang dinamis dan terkontrol (Ashburner et al. 2000) dan pemodelan ontologi tumbuhan obat menggunakan pengetahuan etnobotani (Sanjaya 2014). Namun penelitian tersebut berfokus pada pemodelan ontologi. Berdasarkan latar belakang di atas penelitian ini akan membangun sistem web semantik yang memanfaatkan ontologi yang sudah ada untuk mengelola informasi tumbuhan.

Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Membangun sistem *biodiversity informatics* tumbuhan menggunakan ontologi gen (*gene ontology*).
2. Menerapkan sistem inferensi pengetahuan pada ontologi gen untuk mengembalikan informasi berupa *molecular function*, *biological processes* dan *cellular components* yang terdapat pada tumbuhan.

Ruang Lingkup

Ruang lingkup penelitian adalah:

1. Ontologi yang digunakan dalam penelitian ini berasal dari situs geneontology.org.
2. Membangun *biodiversity informatics* pada level genetik.

Manfaat

Manfaat yang diinginkan dari penelitian ini adalah membantu proses dokumentasi data dan pengetahuan keanekaragaman hayati tumbuhan. Dengan memanfaatkan ontologi, memungkinkan sistem untuk melakukan inferensi dan mengembalikan informasi detail tumbuhan mengenai *molecular function*, *biological processes* dan *cellular components* tumbuhan. Dengan informasi tersebut, diharapkan proses dokumentasi keanekaragaman hayati tumbuhan berjalan lebih efektif.

TINJAUAN PUSTAKA

Biodiversity Information

Biodiversity atau keanekaragaman hayati mengacu pada upaya untuk mempelajari kehidupan di bumi. Keanekaragaman yang dipelajari meliputi bidang molekuler, organisme dan ekologi. *Biodiversity information* (BI) merupakan upaya untuk membuat sumber informasi keanekaragaman hayati global tersedia dalam format digital yang efisien, dan untuk mengembangkan alat yang efektif dalam menganalisis dan memahami data tersebut (Gilman et al. 2009). Salah satu tantangan utama

dalam BI adalah untuk memberikan data pengetahuan mengenai keanekaragaman hayati dengan cepat, dengan pengetahuan tersebut dapat dibangun sistem keanekaragaman hayati yang terpadu (Gilman et al. 2009).

Ontologi

Ontologi adalah suatu metode yang digunakan untuk merepresentasikan pengetahuan yang mendefinisikan hubungan dan klasifikasi dari pengetahuan tertentu (Jepsen 2009). Seperti pemrograman berorientasi objek, ontologi juga menggunakan kelas dan *instance* dalam merepresentasikan pengetahuan. Namun, ontologi dengan pemrograman berorientasi objek berbeda dalam implementasinya. Pada pemrograman berorientasi objek *instance* merupakan contoh nyata dari kelas tersebut. Pada ontologi, *instance* bermakna anggota suatu domain tertentu (Jepsen 2009).

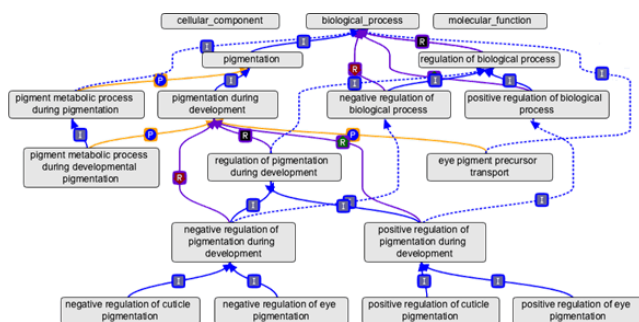
Gene Ontology (GO)

Gene ontology (GO) dikembangkan oleh Gene Ontology Consortium untuk memberikan informasi mengenai gen yang lebih dinamis, terkontrol dan memiliki aturan yang sama. Dibuat dari tahun 1998 GO pada awalnya mengkolaborasi tiga buah *database* organisme yaitu Fly Base (*Drosophila*), *Saccharomyces Genome Database* (SGD) dan *Mouse Genome Database* (MGD). Semenjak itu GO terus berkembang dengan bergabungnya penelitian-penelitian mengenai ontologi gen dari lembaga lain. Saat ini GO sudah menampung data-data mengenai gen tumbuhan, hewan dan mikroba.

GO memiliki tiga komponen ontologi yaitu *Molecular Function*, *Biological Processes* dan *Cellular Components* (Ashburner et al. 2000). *Molecular Function* merupakan aktivitas yang terjadi pada level molekuler seperti aktivitas katalis atau aktivitas binding. *Biological processes* menggambarkan serangkaian kegiatan dengan tujuan tertentu yang melibatkan proses *molecular function*. *Cellular components* menggambarkan komponen-komponen sel yang merupakan bagian dari objek yang lebih besar sebagai contoh retikulum endoplasma atau nukleus. Selain itu *cellular components* juga memberikan informasi *gene-product*, sebagai contoh *ribosome* dan *proteasome*. Gambar 1 merupakan sebagian kecil dari term yang ada di bawah *biological processes*.

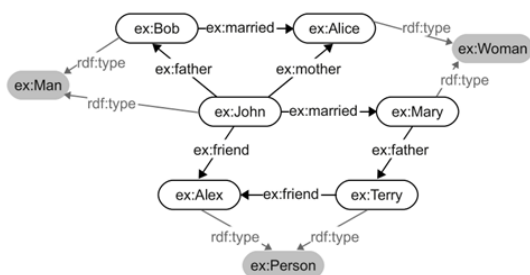
Resource Description Framework (RDF)

eXtensible Markup Language (XML) adalah suatu bahasa yang direpresentasikan dengan *metadata*. XML merepresentasikan informasi agar dapat dengan mudah



Gambar 1. Term di bawah *biological processes*

diakses oleh mesin. XML tidak menyediakan tag-tag tertentu, tetapi penggunaannya dapat mendefinisikan tag secara mandiri. Namun, XML tidak dapat menunjukkan arti dari suatu data. RDF merupakan suatu bentuk dari data model. RDF mengadopsi sintaks-sintaks yang dimiliki oleh XML. Menurut Antoniou dan Hermalen (2008), RDF memiliki konsep dasar RDF yaitu *resource*, *properties*, dan *statement*. *Resource* adalah objek atau tentang sesuatu yang ingin diungkapkan. Beberapa contoh *resource* seperti judul buku, pengarang, penerbit, orang dan sebagainya. Setiap *resource* memiliki *Universal Resource Identifier* (URI). *Properties* adalah deskripsi hubungan antar *resource*, misalnya properti yang menghubungkan antara buku dengan penulisnya adalah “ditulis oleh”. *Statement* adalah bentuk literal dari *resource*, *properties* dan nilainya. Menurut Raimbault (2010), statement disebut juga dengan triple model.



Gambar 2. Contoh Model RDF

Gambar 2 menunjukkan contoh model RDF yang direpresentasikan dalam bentuk *graph*. Contoh *statement* yang dapat dilihat dari Gambar 2 adalah John menikah dengan Mary.

Web Ontology Language (OWL)

Web Ontology Language (OWL) adalah bahasa yang digunakan untuk membaca dan mendefinisikan kelas dan properti model RDF (Segaran et al. 2009). OWL merupakan bahasa yang kompleks. Oleh karena itu, OWL

dibagi menjadi tiga sub bahasa berdasarkan tingkat kompleksitasnya, yaitu OWL-Lite, OWL DL dan OWL Full. OWL-Lite merupakan bentuk sub bahasa OWL yang paling sederhana. OWL DL memiliki fitur yang dapat digunakan untuk memaksimalkan penggunaan sistem *reasoning* dan dapat menjamin kelengkapan informasi yang ingin didapat. OWL Full memberikan fitur yang membebaskan dalam mendeskripsikan sintaks RDF, tetapi tidak menjamin kelengkapan informasi yang didapat (He dan An 2011). OWL memiliki struktur data yang terdiri dari *namespace* dan *ontology head*. *Namespace* digunakan untuk mengelompokkan *identifier*. *Ontology head* digunakan untuk mengumpulkan informasi metadata OWL yang digunakan (He and An 2011).

Simple Protocol and RDF Query Language (SPARQL)

Simple Protocol and RDF Query Language (SPARQL) adalah bahasa *query* yang digunakan untuk melakukan query data terhadap model RDF. Sama halnya dengan query pada SQL, SPARQL juga digunakan untuk melakukan query data-data yang diinginkan. Pembedanya, SQL melakukan *query* pada *database* yang terdiri dari satu atau beberapa tabel, sedangkan SPARQL melakukan *query* pada data RDF yang berupa triple model (Segaran et al. Segaran et al. 2009). Berikut contoh sintaks SPARQL:

```
PREFIX fb:<http://rdf.freebase.com/ns/>
SELECT ?who ?film
WHERE {
  ?film fb:film.film.directed_by ?who
  ?film fb:film.film.starring ?who
}
```

Query dari SPARQL hampir sama seperti pada bahasa *query* yang lain. Seperti sintaks SELECT, WHERE dan lainnya. Sintaks PREFIX digunakan untuk menyingkat alamat sumber dari data ontologi.

Semantic Web Object Oriented Design Methodology (SW-OODM)

Object Oriented Design Methodology (OODM) adalah metode yang digunakan untuk mengembangkan aplikasi web. Metode ini berguna untuk menangkap kebutuhan pengguna. Dengan adanya metode ini aktifitas pengembangan dapat dikontrol dengan baik (Farooq et al. 2010). SW-OODM merupakan metode pengembangan aplikasi web semantik yang merupakan pengembangan dari OODM. SW-OODM meliputi aktivitas yang

berguna untuk memformulasikan dan mendeskripsikan pengetahuan, sehingga aplikasi yang dibuat dapat dimengerti baik oleh mesin maupun oleh manusia (Farooq et al. 2010). Aktivitas yang terdapat di metode SW-OODM yaitu meliputi analysis phase dan design phase. Tahapan-tahapan yang berada dalam *analysis phase* dikelompokkan ke dalam empat model: *preliminary web-ontology model*, *information model*, *user model*, dan *operation model*. Sedangkan pada design phase dikelompokkan menjadi enam unit: *namely, building component model, building navigation model, building operation partitioning model, web ontology model* dan *semantic web-page design* (Farooq et al. 2010).

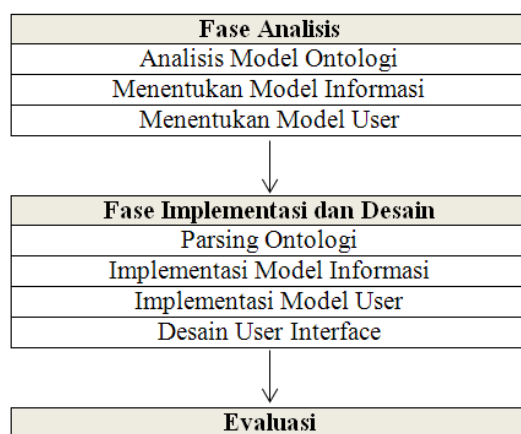
METODE PENELITIAN

Data Penelitian

Data yang digunakan pada penelitian ini adalah data ontologi gen yang berasal dari situs geneontology.org.

Tahapan Penelitian

Tahapan-tahapan yang dilakukan pada penelitian ini mengacu pada metode pengembangan aplikasi SW-OODM. Tahapan pengembangan aplikasi SW-OODM dapat dilihat pada Gambar 3.



Gambar 3. Tahapan Penelitian

Fase Analisa

Pada fase analisa terdapat 3 aktivitas yang akan dilakukan, yaitu analisis model ontologi, menentukan model informasi, menentukan model user. Aktivitas analisis model ontologi akan diidentifikasi domain yang terdapat pada ontologi gen. Hasil identifikasi model ontologi digambarkan dengan bentuk graph, sehingga bentuk dari ontologi gen akan dapat lebih mudah dipelajari.

Aktivitas menentukan model informasi akan memanfaatkan hasil dari analisis model ontologi untuk menentukan kelas, atribut dan keterkaitan antar objek yang ada di dalam ontologi. Hal ini perlu dilakukan untuk menjadi acuan dalam pembuatan query dan membuat query yang tepat dengan SPARQL untuk mengambil informasi yang terdapat di dalam ontologi dan juga melakukan inferensi pengetahuan tumbuhan yang ada di dalam ontologi.

Pada aktivitas menentukan model user akan dianalisis kebutuhan dari user yang akan menggunakan sistem yaitu berupa kelompok user yang mengakses sistem, rancangan kelas dari sistem, alur aktivitas yang dilakukan user dan skenario alur akses sistem dari user. Hasil dari tahapan ini berupa definisi diagram use case, class diagram, activity diagram dan sequence diagram.

Fase Implementasi dan Desain

Pada fase implementasi dan desain akan diawali dengan parsing ontologi. Parsing ontologi memetakan ontologi gen menjadi triple yang berupa subject, predicate dan object. Setelah ontologi dilakukan parsing dan menghasilkan triple, hasil ini yang akan dilakukan query dengan menggunakan SPARQL.

Pada tahapan implementasi model informasi mencakup pembuatan query SPARQL untuk mengembalikan informasi yang terdapat dalam ontologi, yaitu informasi gene-product, cellular location dan sequence. Kemudian pada tahapan implementasi model akan dibuat fungsi-fungsi yang akan memanfaatkan query yang telah dibuat pada tahap implementasi model sehingga informasi dari sistem ontologi gen dapat diakses dengan menggunakan web.

Pada tahapan implementasi user akan menghasilkan fungsi-fungsi yang akan menangani input yang diberikan oleh user, melakukan pengambilan informasi berdasarkan input yang diterima, dan memberikan output yang sesuai dengan input yang sudah diberikan. Tahap desain user interface akan menghasilkan halaman yang akan diakses oleh user. Halaman yang dibuat berupa halaman input dan halaman output yang dapat dilihat oleh user.

Evaluasi

Fase evaluasi akan dilakukan pengujian dari sistem yang sudah dibuat. Pengujian yang dilakukan menggunakan metode black box. Metode ini akan memberikan kasus untuk dilakukan oleh sistem dengan memberi input dan menguji kesesuaian hasil yang diberikan oleh sistem.

Lingkungan Pengembangan

Pembangunan sistem ontologi gen tanaman berbasis web ini dilakukan dengan menggunakan perangkat keras dan perangkat lunak sebagai berikut:

- Prosesor Intel Core i7 4500U 1,8 GHz
- Memori 12 GB
- Hard disk 1 TB
- Sistem operasi Windows 7 Ultimate
- Bahasa pemrograman Python dengan Flask sebagai web framework
- RDFLib sebagai library yang digunakan untuk penggunaan RDF pada Phyton
- Lingkungan pengembangan (IDE) Visual Studio 2013
- Protégé 4.3.0 sebagai pemodelan ontologi

Rencana Jadwal Penelitian

Penelitian ini akan dilakukan selama 6 bulan dengan rincian kegiatan seperti tercantum pada Tabel 1.

DAFTAR PUSTAKA

- Antoniou, G and F Hermalen (2008). *A Semantic Web Primer, Second edition*. Cambridge (GB): The MIT Pr.
- Ashburner, M et al. (2000). *Gene ontology: tool for the unification of biology*. California (US): Stanford University School of Medicine.
- Bappenas (2003). *Indonesia Biodiversity and Action Plan 2003-2020*. Jakarta (ID): Bappenas.
- DiLecce V, Calabrese M, ed. (2008). *LOF: Identifying Density-Based Local Outliers*. Taxonomies and Ontologies in Web Semantic Applications: the New Emerging Semantic Lexicon-Based Model. Vol. CIMCA International Conference, IEEE. 277-283. Viena (AUT).
- Ding, L, Finin T, Joshi A, Peng Y, Pan R, and Reddipari P, eds. (2005). *Search on The Semantic Web*. Computer Society, IEEE. 38(10):62-69.
- Farooq, A, S Ahsan, and A Sha, eds. (2010). *Engineering Semantic Web Applications by Using Object-Oriented Paradigm*. Journal of Computer Science. 2(6):156-165.
- Gilman, E, N King, T Peterson, V Chavan, and A Hahn (2009). *Building The Biodiversity Data Commons The Global Biodiversity Information Facility*. Graz (AT): Graz University of Technology.
- Guralnick, R and A Hill (2008). *Biodiversity Informatics: Automated Approaches for Documenting Global Biodiversity Patterns and Processes*. Boulder (CO): Department of Ecology and Evolutionary Biology University of Colorado.
- Hamzari, ed. (2008). *Identifikasi tanaman obat-obatan yang dimanfaatkan oleh masyarakat sekitar hutan Tabo-tabo*. Jurnal Hutan dan Masyarakat. 3(2):159-167.
- He, G and L An, eds. (2011). *Ontology Language OWL Research Study*. MASS International Conference, IEEE. 1-4.
- Herdiyeni, Y, Zuhud EAM, Hikmat A, Mustari AH, Pravista DS, Mega M, Setiawan R, and Metananda AA, eds. (2013). *IPB Biodiversity Informatics (IPBi-otics) Untuk Pembangunan Berkelanjutan*. Prosiding Seminar Hasil-Hasil PPM IPB 2013. 2: 389-402.
- Jepsen, TC, ed. (2009). *Just What Is an Ontology, Anyway?* IT professional, IEEE. 11(5):22-27.
- Laallam, FZ, ML Kherfi, and SM Benslimane, eds. (2013). *Using ontologies to overcoming draw-back of database and vice versa: a survey*. CSEIJ. 3(2):1-21.
- Raimbault, T, ed. (2010). *Overviewing the RDF(S) semantic web*. CiSE International Conference, IEEE. 1-4.
- Segaran, T, Evans C, and Taylor J (2009). *Programming the semantic web*. New York(US): O'Reilly.
- Vadivu G, Hopper SW, ed. (2012). *Ontology mapping of indian medicinal plants with standardized medical terms*. Journal of Computer Science. 8(9):1576-1584.
- Zuhud, EAM (2008). *Potensi hutan tropika indonesia sebagai penyangga bahan obat alam untuk kesehatan bangsa*. Bogor(ID). Fakultas Kehutanan IPB.

