

Pembangunan *Biodiversity Informatics* Genetika Tumbuhan Obat Berbasis Ontologi

Bayu Santoso(G64134017)*, Yeni Herdiyeni

Abstrak/Abstract

Indonesia memiliki lebih dari 32.000 spesies tumbuhan. Dari kumpulan spesies tersebut terdapat tumbuhan obat di dalamnya. Tidak kurang dari 2039 spesies tumbuhan obat berasal dari hutan Indonesia. Saat ini hutan Indonesia mengalami kerusakan dan kepunahan. Oleh karena itu, diperlukan upaya untuk melestarikan tumbuhan obat. Salah satu cara untuk melestarikan tumbuhan obat adalah dengan cara mengenali tumbuhan obat. Informasi yang dibutuhkan mengenai tumbuhan obat sulit untuk ditemukan. 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 obat. Selain itu sistem web semantik ini akan menyediakan *web service* yang memungkinkan untuk terintegrasi dengan sistem web semantik yang lain.

Kata Kunci

biodiversity informatics; tumbuhan obat; ontologi; web semantik

*Alamat Email: bayusantoso.mail@gmail.com

PENDAHULUAN

Latar Belakang

Indonesia memiliki lebih dari 32.000 spesies tumbuhan (Bappenas 2003). Dari kumpulan spesies tersebut terdapat tumbuhan obat di dalamnya. Menurut Zuhud (2008) tidak kurang dari 2039 spesies tumbuhan obat berasal dari hutan Indonesia. Tanaman obat yang beraneka ragam jenis, habitus, dan khasiatnya mempunyai peluang besar serta memberi kontribusi bagi pembangunan dan pengembangan hutan (Hamzari 2008). Saat ini hutan Indonesia mengalami kerusakan dan kepunahan (Zuhud 2008). Oleh karena itu, diperlukan upaya untuk melestarikan tumbuhan obat. Salah satu cara untuk melestarikan tumbuhan obat adalah dengan cara mengenali tumbuhan obat (Hamzari 2008). *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 Informa-

tion System (ITIS) dan Global Biodiversity Information Facility (GBIF) menyediakan informasi yang luas tentang tumbuhan. Sistem tersebut dibuat dengan menggunakan model basis data relasional. Model basis data relasional menimbulkan masalah apabila digunakan pada sistem berbasis inferensi (Laallam et al. 2013). Selain itu model basis data relasional dapat menghasilkan data yang berganda. Oleh sebab itu, dibutuhkan pemodelan data yang dapat mengatasi hal tersebut. Metode pemodelan data yang dapat menangani sistem berbasis inferensi adalah ontologi.

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 and Calabrese 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 obat 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 meng-

hasilkan data gen yang dinamis dan terkontrol (Ashburner et al. 2000). Namun, penelitian tersebut belum menghasilkan hubungan antara tumbuhan obat dengan informasi gen-nya. Pada penelitian ini akan dibuat sistem yang memanfaatkan web semantik yang digunakan untuk mengintegrasikan informasi gen dengan tumbuhan obat. *Resource Description Framework* (RDF) akan diterapkan pada sistem ini untuk mengatasi masalah integrasi dengan data tumbuhan obat. RDF merupakan standar untuk merepresentasikan data yang berbentuk grafik dan membagikan dengan manusia dan mesin.

Berdasarkan latar belakang di atas penelitian ini akan mengembangkan sistem web semantik yang memberikan informasi gen tumbuhan obat. Selain itu sistem web semantik ini akan menyediakan *web service* yang memungkinkan untuk terintegrasi dengan sistem web semantik yang lain.

Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menerapkan teknologi web semantik untuk menggunakan ontologi gen (*gene ontology*).
2. Mengintegrasikan sistem ontologi gen dengan data tumbuhan obat.
3. Mengintegrasikan sistem ontologi gen dengan sistem yang lain, yaitu ontologi tumbuhan (*plant ontology*) dan ontologi lingkungan (*environment ontology*).

Ruang Lingkup

Ruang lingkup penelitian adalah:

1. Ontologi yang digunakan dalam penelitian ini berasal dari situs geneontology.org.
2. Data korpus yang digunakan terbatas hanya tumbuhan obat.
3. Sistem ontologi gen yang dibangun diintegrasikan dengan ontologi tanaman dan ontologi lingkungan.

Manfaat

Hasil pengembangan sistem ontologi gen ini diharapkan dapat membantu memberikan informasi mengenai gen dari tumbuhan obat.

TINJAUAN PUSTAKA

Biodiversity Information

Biodiversity atau keanekaragaman hayati mengacu pada upaya untuk mempelajari kehidupan di bumi. Keanekaragaman yang dipelajari meliputi bidang molekul, organisme

dan ekologi (Sarkar 2007). *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).

Tumbuhan Obat

Sampai tahun 2001 Laboratorium Konservasi Tumbuhan, Fakultas Kehutanan IPB telah mendata dari berbagai laporan penelitian dan literatur tidak kurang dari 2039 spesies tumbuhan obat yang berasal dari hutan Indonesia. Berdasarkan keanekaragamannya, tumbuhan obat dapat dikelompokkan berdasarkan familinya, formasi hutan, habitus, dan bagian tumbuhan yang dapat digunakan sebagai obat (Zuhud 2008). Bagian-bagian tumbuhan yang dapat digunakan sebagai obat sangat bervariasi mulai dari daun muda atau pucuk, buah, batang, kulit, getah, umbi dan akar (Hamzari 2008).

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)

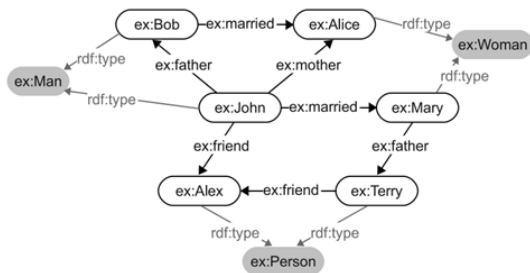
Gene ontology (GO) dikembangkan oleh Gene Ontology Consortium untuk memberikan informasi mengenai gen yang lebih dinamis, terkontrol dan memiliki aturan yang sama. GO memiliki tiga komponen ontologi yaitu *Biological Processes*, *Cellular Components* dan *Molecular Function*. *Biological processes* mengacu gen atau *gene product* yang berkontribusi pada suatu objek biologis. *Molecular function* mengacu pada aktifitas biokimia yang terjadi pada *gene product*. *Cellular components* adalah tempat sel aktif yang berada di dalam *gene product* (Ashburner et al. 2000).

Web Semantik

Web semantik adalah suatu *framework* yang memungkinkan data dan pengetahuan untuk dipublikasikan, dibagi dan menggunakan kembali pengetahuan yang terdapat pada internet (Ding et al. 2005). Web semantik adalah suatu *graph* yang besar yang tersusun dari *graph-graph* yang lebih kecil. (T et al. 2009).

Resource Description Framework (RDF)

eXtensible Markup Language (XML) adalah suatu bahasa yang direpresentasikan dengan *metadata*. XML merepresentasikan informasi agar dapat dengan mudah diakses oleh mesin. XML tidak menyediakan tag-tag tertentu, tetapi penggunaanya 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 Hermelen (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 1. Contoh Model RDF

Gambar 1 menunjukkan contoh model RDF yang direpresentasikan dalam bentuk *graph*. Contoh *statement* yang dapat dilihat dari Gambar 1 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 (T 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. T 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-ODM 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).

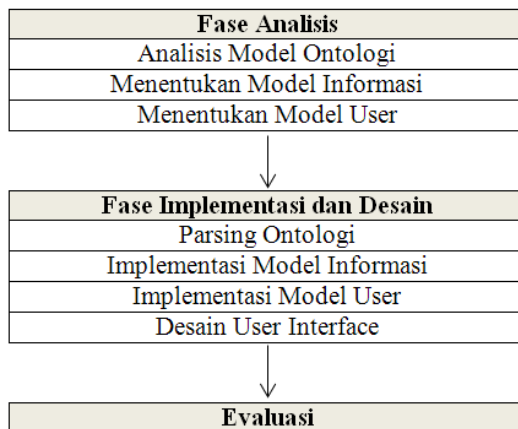
METODE PENELITIAN

Data Penelitian

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

Tahapan Penelitian

Tahapan-tahapan yang dilakukan pada penelitian ini mengacu pada metode pengembangan aplikasi SW-ODM. Tahapan pengembangan aplikasi SW-ODM dapat dilihat pada Gambar 2.



Gambar 2. 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 model dari 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. Pada aktivitas menentukan model user akan didefinisikan diagram use case, class diagram, activity diagram dan sequence diagram.

Fase Implementasi dan Desain

Pada fase implementasi dan desain akan diawali dengan parsing ontologi. Hasil dari parsing ontologi dapat digunakan untuk template ontologi, sehingga dari template tersebut mesin dapat memahami model dari ontologi gen. Pada tahapan implementasi model akan dihasilkan fungsi-fungsi dari sistem ontologi gen yang berupa web. Pada tahapan implementasi user akan menghasilkan fungsi yang dapat digunakan oleh pengguna agar dapat berinteraksi dengan sistem. Tahap desain *user interface* akan menghasilkan halaman desain halaman yang diakses pengguna yang dapat memudahkan pengguna dalam menggunakan sistem.

Evaluasi

Fase evaluasi akan dilakukan pengujian dari sistem yang sudah dibuat. Pengujian yang dilakukan meliputi kecepatan proses pencarian dari sistem yang dibuat, kelengkapan data yang dihasilkan dari proses pencarian, dan kecocokan hasil pencarian dengan kata kunci pencarian.

Lingkungan Pengembangan

Pembangunan sistem ontologi gen tanaman obat 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 Python
- Lingkungan pengembangan (IDE) Visual Studio 2013
- Protégé 4.3.0 sebagai pemodelan ontologi

Rencana Jadwal Penelitian

DAFTAR PUSTAKA

Antoniou, G and F Hermalen (2008). *A Semantic Web Primer, Second edition*. Cambridge (GB): The MIT Pr.

Tabel 1. Rencana Jadwal Penelitian

Kegiatan	Mei		Juni				Juli				Agustus				September				Oktober				November	
	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2
Penyusunan Proposal Skripsi																								
Kolokium																								
Pengumpulan Data																								
Fase Analisis																								
Fase Implementasi data																								
Fase Evaluasi																								
Penulisan Skripsi																								
Seminar																								
Sidang skripsi																								
Perbaikan laporan penelitian																								

- 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 and M Calabrese, eds. (2008). *LOF: Identifying Density-Based Local Outliers*. Taxonomies and Ontologies in Web Semantic Applications: the New Emerging Semantic Lexicon-Based Model. Vol. CIMCAVadivu G, Hopper SW, ed. (2012). *Ontology mapping of indian medicinal plants with standardized medical terms*. Journal of Computer Science. 8(9):1576-1584.
- Ding, L, T Finin, A Joshi, Y Peng, R Pan, and P Reddivari, 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.
- 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.
- T, Segaran, Evans C, and Taylor J (2009). *Programming the semantic web*. New York(US): O'Reilly.
- Zuhud, EAM (2008). *Potensi hutan tropika indonesia sebagai penyangga bahan obat alam untuk kesehatan bangsa*. Bogor(ID). Fakultas Kehutanan IPB.