



**PERTUKARAN DATA ANTARA RSTUDIO DAN MONGODB
PADA SISTEM INFORMASI GEOGRAFIS UNTUK KASUS
PERTANIAN INDONESIA**

LALITYA NOOR AHSANA



**ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2015**

PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi berjudul Pertukaran Data antara RStudio dan MongoDB pada Sistem Informasi Geografis untuk Kasus Pertanian Indonesia adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Desember 2015

Lalitya Noor Ahsana
NIM G64110044

ABSTRAK

LALITYA NOOR AHSANA. Pertukaran Data antara RStudio dan MongoDB pada Sistem Informasi Geografis untuk Kasus Pertanian Indonesia. Dibimbing oleh HARI AGUNG ADRIANTO

Penggunaan Internet mengalami perkembangan yang sangat pesat beberapa tahun ke belakang. Salah satu layanan yang disediakan oleh internet adalah media sosial. Di Indonesia media sosial sangat digemari, salah satu nya adalah Twitter, dimana salah satu bahasan yang dibicarakan adalah mengenai pertanian. Pemetaan bahasan pertanian Indonesia dirasa perlu sehingga dibangunlah Sistem Informasi Geografis Pertanian Indonesia. Penelitian ini bertujuan untuk membangun *database* sebagai pendukung pembangunan Sistem Informasi Geografis Pertanian Indonesia menggunakan data dari Twitter. Data yang besar membuat penggunaan *database* relasional menjadi tidak *feasible* sehingga digunakanlah *database* NoSQL berbasis dokumen yaitu MongoDB. Tahapan pembuatan *database* meliputi enam tahapan yaitu *database planning*, *system definition*, *requirement collection and analysis*, *database design*, *DBMS selection*, dan *implementation*. Hasil dari penelitian ini berupa *database* untuk data Twitter menggunakan MongoDB yang kemudian diterapkan ke dalam Sistem Informasi Geografis Pertanian Indonesia beserta pertukaran data antara RStudio dan MongoDB.

Kata kunci: *database*, media sosial, MongoDB, RStudio, Twitter

ABSTRACT

LALITYA NOOR AHSANA. RStudio and MongoDB Data Exchange on Geographic Information System for Agricultural Issue in Indonesia. Supervised by HARI AGUNG ADRIANTO

Internet has grown massively over few years. Social media is one of services provided by internet. In Indonesia, media social are very popular and the most popular one is Twitter, which often discussing various topics as agriculture is being one of them. Since mapping of the Indonesian agricultural discussion is necessary so Indonesian Agriculture Geographic Information System will be built to accommodate that purpose. This research goal is to build a database for Indonesian Agriculture Geographic Information System using data from Twitter. Since there are voluminous data retrieved, a relational database is not feasible to cope with, so this research will use NoSQL database document-based called MongoDB. Database development include six steps, there are database planning, system definition, requirement collection and analysis, database design, DBMS selection, and implementation. The final product of this research is a database for Twitter data using MongoDB dedicated for Indonesian Agriculture Geographic Information System and data exchange between RStudio and MongoDB.

Keywords: *database*, MongoDB, RStudio, social media, Twitter

**PERTUKARAN DATA ANTARA RSTUDIO DAN MONGODB
PADA SISTEM INFORMASI GEOGRAFIS UNTUK KASUS
PERTANIAN INDONESIA**

LALITYA NOOR AHSANA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer
pada
Departemen Ilmu Komputer

**ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2015**

Penguji:

1. Husnul Khotimah, SKomp MKom
2. Rina Trisminingsih, SKom MT

Judul Skripsi: Pertukaran Data antara RStudio dan MongoDB pada Sistem
Informasi Geografis untuk Kasus Pertanian Indonesia

Nama : Lalitya Noor Ahsana

NIM : G64110044

Disetujui oleh

Hari Agung Adrianto, SKom MSi
Pembimbing

Diketahui oleh

Dr Ir Agus Buono, MSi MKom
Ketua Departemen

Tanggal Lulus:

PASTIKAN DI BALIK LEMBAR PENGESAHAN ADALAH HALAMAN KOSONG

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Judul yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2015 ini ialah Pertukaran Data antara RStudio dan MongoDB pada Sistem Informasi Geografis untuk Kasus Pertanian Indonesia.

Penulis menyadari bahwa selama mengerjakan tugas akhir ini mengalami berbagai kendala. Namun, atas berkat kerja sama dan bimbingan dari berbagai pihak dan atas berkat rahmat Allah *subhanahu wa ta'ala* kendala yang dihadapi dapat diselesaikan. Untuk itu penulis ucapkan terima kasih kepada:

- 1 Bapak Hari Agung Adrianto, SKom MSi selaku pembimbing yang telah sabar, tulus, dan ikhlas meluangkan waktu, tenaga, dan pikiran dalam memberikan bimbingan, motivasi, arahan, dan saran yang bermanfaat bagi penulis.
- 2 Ibu Husnul Khotimah, SKomp MKom dan Ibu Rina Trisminingsih, SKom MT selaku penguji atas segala saran dan masukan yang diberikan.
- 3 Ihda Husnayain, Ryan Budiman Denatari, Firdaus Saptahadi Pratama, dan Muhammad Nur Husein atas kerjasama dan bantuannya selama mengerjakan tugas akhir ini.
- 4 Rekan-rekan Ilmu Komputer IPB angkatan 48.
- 5 Segenap staf dan pengajar Departemen Ilmu Komputer IPB.
- 6 Ayahanda Agus Fahri Husein, ibunda Eli Rosidah dan adik saya tersayang Fahriyya Noor Ahsana serta seluruh keluarga besar atas segala dukungan yang diberikan.

Semoga karya ilmiah ini dapat bermanfaat.

Bogor, Desember 2015

Lalitya Noor Ahsana

DAFTAR ISI

DAFTAR TABEL	vi
DAFTAR GAMBAR	vi
PENDAHULUAN	1
Latar Belakang	1
Perumusan Masalah	2
Tujuan Penelitian	2
Manfaat Penelitian	2
Ruang Lingkup Penelitian	2
TINJAUAN PUSTAKA	2
NoSQL	2
MongoDB	5
JSON	6
BSON	6
METODE	6
Data Penelitian	6
Tahapan Penelitian	6
Lingkungan Pengembangan	10
HASIL DAN PEMBAHASAN	11
<i>Database Planning</i>	11
<i>System Definition</i>	11
<i>Requirement Collection and Analysis</i>	12
<i>Database Design</i>	14
<i>DBMS Selection</i>	16
<i>Implementation</i>	17
SIMPULAN DAN SARAN	19
Simpulan	19
Saran	19
DAFTAR PUSTAKA	20

DAFTAR TABEL

1	Perbandingan <i>relational databases</i> dan NoSQL <i>database</i> (Mohamed <i>et al.</i> 2014)	3
2	Perbedaan istilah pada <i>relational database</i> dan <i>non-relational database</i> (MongoDB 2015)	3
3	Perbandingan jenis <i>database</i> NoSQL (Catell 2010)	4
4	Perbandingan antara CouchDB dan MongoDB (Cattell 2010)	5
5	Atribut pada data Twitter	12

DAFTAR GAMBAR

1	Tahapan pengembangan SIG Pertanian Indonesia	7
2	<i>Database system development lifecycle</i> (mengadopsi dari Connolly dan Begg 2015)	8
3	<i>Embedded Data Models</i>	9
4	<i>Normalized Data Models</i>	9
5	Batasan sistem	11
6	TDF data konten URL	13
7	TDM konten URL	13
8	Label data	14
9	Alur transaksi data dari RStudio ke <i>database</i> NoSQL	14
10	Struktur <i>database</i>	15
11	Fungsi <i>connection</i> antara RStudio dan MongoDB	17
12	Fungsi <i>insert</i> ke <i>database</i>	18
13	Fungsi <i>retrive collection</i> TDMURL	18
14	Hasil <i>retrieve</i> TDMURL	19

font kenapa
jadi bukan
times lagi

PENDAHULUAN

Latar Belakang

Penggunaan internet di Indonesia mengalami perkembangan pesat. Pada penelitian yang dilakukan oleh lembaga riset pasar e-Marketer seperti dikutip dalam *website* Kemenkominfo (2014) menyebutkan bahwa populasi *netter* tanah air mencapai 83.7 juta orang pada 2014, terbesar keenam di dunia. Jumlah tersebut diperkirakan akan terus bertambah sampai 40% pada tahun 2018.

Salah satu layanan internet yang cukup digemari di Indonesia adalah media sosial. Media sosial merupakan layanan berbasis *web* yang memungkinkan individu untuk membangun profil dalam sistem yang terbatas, berhubungan dengan pengguna lain untuk berbagi informasi, dan melihat profil pengguna lain di dalam sistem tersebut (Boyd dan Ellison 2007). Adler dan Rodman (2006) menyatakan komunikasi secara *online* dianggap lebih murah, cepat, dan mudah. Komunikasi dengan menggunakan media sosial juga tidak terikat ruang dan waktu (*real time*) sehingga memudahkan siapapun dalam berinteraksi, berkomunikasi, dan memperoleh informasi. Beberapa media sosial yang populer adalah Facebook, Twitter, Path, Google+ dan LinkedIn. Pengguna Facebook di Indonesia merupakan terbesar keempat di dunia, sedangkan Twitter menempati posisi lima di dunia (Kemenkominfo 2013).

Topik yang dibicarakan melalui media sosial sangat beragam, termasuk tentang pertanian (Rhoades dan Hall 2007). Semakin banyaknya orang yang terlibat dalam sebuah topik di media sosial berarti banyak pula orang yang mengetahui dan menanggapi permasalahan terkait. Oleh karena itu dirasa perlu untuk mengembangkan suatu sistem yang dapat memetakan bahasan topik di media sosial terutama mengenai pertanian di Indonesia dalam bentuk Sistem Informasi Geografis. Sistem yang akan dibangun adalah Sistem Informasi Geografis Pertanian Indonesia.

Pembangunan Sistem Informasi Geografis Pertanian Indonesia terdiri dari lima modul. Modul pertama akuisisi data dari Twitter. Modul kedua melakukan praproses data. Modul ketiga adalah membangun *database* dan pertukaran data antara aplikasi pengolahan data dan *database*. Modul keempat melakukan *clustering*. Modul kelima adalah visualisasi.

Pembangunan sistem yang dikerjakan mengalami perpindahan data yang cepat antar modul karena proses pengolahan data yang meliputi akuisisi, praproses data, clustering dan visualisasi dilakukan secara terpisah oleh Pratama (2015), Husnayain (2015), Denatari (2015), Husein (2015) menggunakan aplikasi RStudio. Sehingga pada pembangunan SIG Pertanian Indonesia, penelitian ini dikhususkan untuk membangun *database* dan proses pertukaran data antara aplikasi pengolahan data RStudio untuk mempermudah pertukaran data antar modul. Tahapan pembuatan *database* mengadopsi dari Connolly dan Begg (2015) meliputi enam tahapan yaitu *database planning*, *system definition*, *requirement collection and analysis*, *database design*, *DBMS selection*, dan *implementation*.

Data yang digunakan pada penelitian ini adalah data dari media sosial Twitter. Data dari Twitter merupakan data yang besar dan terdistribusi, sehingga mengakibatkan penggunaan *database* relasional kurang dapat diandalkan,

sehingga digunakan NoSQL *database*. Salah satu NoSQL *database* adalah MongoDB (Chodorow dan Dirolf 2010). Ramalho (2011) menyatakan dalam penelitiannya bahwa MongoDB optimal untuk data dengan perubahan cepat. MongoDB sendiri memiliki struktur penyimpanan BSON yang memperbolehkan *update* pada *fields* tertentu pada dokumen. Oleh karena itu penelitian ini menggunakan MongoDB karena lebih sesuai dengan karakteristik data media sosial yang mengalami perubahan dengan cepat juga didukung dengan bahasa pemrograman R untuk memudahkan proses pengolahan data.

Perumusan Masalah

Berdasarkan latar belakang, maka perumusan masalah dalam penelitian ini adalah bagaimana membangun *database* sebagai sarana penyimpanan data agar mempermudah pertukaran data antara RStudio dan MongoDB.

Tujuan Penelitian

Tujuan penelitian ini adalah membangun *database* dari media sosial Twitter menggunakan MongoDB dan melakukan proses pertukaran data antara RStudio dan MongoDB

Manfaat Penelitian

Hasil penelitian adalah *database* dalam bentuk MongoDB. *Database* diintegrasikan dengan RStudio mempermudah proses transportasi dan transformasi data. *Database* kemudian diterapkan pada Sistem Informasi Geografis Pertanian Indonesia untuk dapat mempermudah pemetaan topik yang menjadi tren di media sosial menyangkut pertanian Indonesia.

Ruang Lingkup Penelitian

Ruang lingkup pada penelitian ini adalah membangun *database* menggunakan MongoDB dari data spasial dan teks Twitter di Indonesia beserta pertukaran data dari RStudio dan MongoDB dengan *package* ‘rmongodb’ untuk kemudian diterapkan pada Sistem Informasi Geografis Pertanian Indonesia.

TINJAUAN PUSTAKA

NoSQL

NoSQL merupakan *database non-relational* yang tidak memiliki ketergantungan pada tabel/*key* model. Kelebihan dari NoSQL adalah dapat menjamin BASE (*Basically Available, Soft State, and Eventual consistency*) tapi tidak menjamin ACID (*Atomicity, Consistency, Isolation and Durability*) (Nayak *et al.* 2013).

Tabel 1 Perbandingan *relational databases* dan *NoSQL database* (Mohamed *et al.* 2014)

Kategori	<i>Relational database</i>	<i>NoSQL database</i>
<i>Authentication</i>	Memiliki mekanisme autentifikasi dan pilihan penggunaannya	Tidak menyediakan mekanisme autentifikasi. Harus menggunakan metode eksternal.
<i>Data Integrity</i>	Transaksi integrasi data terjamin	Data integrasi tidak selalu dapat dicapai
<i>Confidentiality</i>	Dapat dicapai karena digunakan teknik enkripsi	Tidak dapat dicapai karena tidak adanya proses enkripsi
<i>Auditing</i>	Menyediakan mekanisme audit sehingga mengizinkan penulisan ke dalam <i>database</i> syslog atau <i>file xml</i>	Tidak menyediakan mekanisme audit.
<i>Client communication</i>	Menyediakan mekanisme keamanan komunikasi antar <i>client</i> menggunakan enkripsi dan protokol SSL	<i>Database</i> NoSQL tidak menyediakan mekanisme keamanan komunikasi antar <i>client</i>

Dari Tabel 1 banyak dijelaskan masih banyak kekurangan yang dimiliki oleh *NoSQL database*. Walaupun demikian Chodorow dan Dirolf (2010) dalam penelitiannya mengungkapkan bahwa *database* NoSQL merupakan solusi untuk *big data* terutama data dari media sosial yang merupakan data yang besar dan memiliki perubahan cepat. *Relational database* dan *non-relational database* juga memiliki perbedaan dalam istilah, perbedaan istilah pada dapat dilihat pada Tabel 2.

Tabel 2 Perbedaan istilah pada *relational database* dan *non-relational database* (MongoDB 2015)

<i>Sql terms/concepts</i>	<i>Mongodb terms/concepts</i>
<i>Table</i>	<i>Collection</i>
<i>Row document</i>	<i>BSON document</i>
<i>Column</i>	<i>Field</i>
<i>Index</i>	<i>Index</i>
<i>Table join</i>	<i>Embedded document dan linking</i>
<i>Primary key</i> (User yang menentukan kolom yang menjadi <i>primary key</i>)	<i>Primary key</i> (<i>Primary key</i> diatur sebagai <i>field _id</i> secara otomatis terbentuk ketika memasukan data)
<i>Aggregation (Group by)</i>	<i>Aggregation pipeling</i>

Database NoSQL memiliki empat jenis bentuk *database* yaitu *key-value*, *document-oriented*, *column based*, dan *graph oriented*. Perbandingan pengertian, penggunaan dan contoh dari keempat jenis *database* dapat dilihat pada Tabel 3.

Tabel 3 Perbandingan jenis *database* NoSQL (Catell 2010)

	<i>Key-value</i>	<i>Document-oriented</i>	<i>Column-based</i>	<i>Graph-oriented</i>
Pengertian dan Penggunaan	Menyimpan nilai terindeks berdasarkan <i>keys</i> memiliki performa cukup baik tapi sulit untuk melakukan <i>query</i> dan implementasi. Digunakan untuk aplikasi dengan satu jenis <i>object</i> dan pencarian satu atribut saja	Penyimpanan dan pengorganisasian data sebagai koleksi dokumen, bukan sebagai tabel terstruktur dengan bidang <i>uniformsized</i> untuk setiap <i>record</i> . <i>Fields</i> dengan panjang berapapun dapat ditambahkan ke dokumen. Dokumen terindeks dan tersedia mekanisme <i>query</i> sederhana Digunakan untuk penyimpanan <i>multiple object</i> dan pencarian berdasarkan <i>multiple fields</i> .	Penyimpanan <i>record</i> yang dapat dipartisi secara vertikal dan horizontal	Digunakan untuk menyimpan <i>object</i> dan keterhubungannya dalam bentuk <i>node</i> dan <i>edge</i> pada <i>graph</i> . Cocok untuk data yang memiliki hirarki.
Contoh	Amazon Dynamo dan Project Voldemort	CouchDB, MongoDB	Cassandra, BigTable	Neo4j

Dari Tabel 3 dijelaskan bahwa *database* NoSQL berbasis *key-value* biasa digunakan untuk penyimpanan satu jenis *object* saja, *database* NoSQL berbasis *document* digunakan penyimpanan *multiple object*, *database* NoSQL berbasis

column biasa digunakan untuk data yang dapat dipartisi sedangkan *graph-oriented* biasa digunakan untuk untuk memperlihatkan keterhubungan antar *object* pada *database*.

MongoDB

MongoDB merupakan penyimpanan dokumen data terdistribusi berbasis NoSQL (Not Only SQL) dengan tipe *Document Store Databases*. MongoDB mengkhususkan untuk menyediakan penyimpanan untuk data yang besar dan *high performance*. MongoDB dapat digunakan untuk menggantikan *database* relasional biasa. Kelebihan dari MongoDB adalah pada *query language* yang mirip dengan *object-oriented query language*, mendukung *full index*, *dynamic queries*, kaya akan bahasa *query*, mendukung pemrosesan *cloud* dengan data fragmentasi, penyimpanan berbasis JSON maupun BSON, dan dapat digunakan bersama dengan C, C++, C#, Erlang, Haskell, Java, JavaScript, Perl, PHP, Python, Ruby, dan Scala (Prajapati 2013).

Tabel 4 Perbandingan antara CouchDB dan MongoDB (Cattell 2010)

	CouchDB	MongoDB
<i>Basic Concept</i>	<i>Document oriented</i>	<i>Document oriented</i>
<i>Indexing</i>	<i>R Tree</i>	<i>2D</i>
<i>Vector Data Types</i>	<i>Only2D</i>	<i>2Dsphere</i>
	<i>Fully</i>	<i>Fully</i>
	<i>Only</i>	<i>Only</i>
<i>Topological Functions</i>	<i>Within()</i>	<i>Within(Point)</i>
	<i>Contains()</i>	<i>Contains(Point)</i>
<i>Analysis and Metric Functions</i>	<i>Only Distance()</i>	<i>Only Distance(point)</i>
<i>Set Functions</i>	<i>Not supported</i>	<i>Only Intersection(Point)</i>
<i>Input/output format</i>	<i>Input:</i> <i>.SHP</i> <i>Output:</i> <i>.KML, .CSV</i> <i>.GeoJSON</i>	<i>Input:</i> <i>.GeoJSON</i> <i>Output:</i> <i>.GeoJSON</i>

Dari perbandingan pada Tabel 4 bahwa kemampuan MongoDB dan CouchDB tidak jauh berbeda. Ramalho (2011) menyatakan dalam penelitiannya bahwa MongoDB optimal untuk data dengan perubahan cepat. MongoDB sendiri memiliki struktur penyimpanan BSON yang memperbolehkan *update* pada *fields* tertentu pada dokumen. CouchDB merupakan *database* yang *fault-tolerant* keuntungannya datanya selalu konsisten tetapi proses penyimpanan dan pengolahan data menjadi sangat lambat.

JSON

JSON merupakan singkatan dari *Javascript Object Notation* adalah *script* yang merepresentasikan suatu struktur data sederhana dan *array* asosiatif, atau dengan kata lain, JSON adalah suatu format pengiriman data yang berbasis javascript yang dibuat sedemikian rupa hingga mempermudah dibaca manusia (MongoDB 2015). Format JSON dalam 1 *record* ditandai dengan kurung kurawal, selanjutnya ditulis nama *field* dan diikuti dengan titik dua , baru diikuti dengan nilai dari kolom tersebut. Secara umum format JSON sebagai berikut:

```
{[nama Field] : [Nilai Field],  
[nama Field2] : [Nilai Field2];}
```

BSON

BSON merupakan singkatan Binary JSON adalah suatu pengkodean binary dari dokumen berbentuk JSON. Seperti JSON, BSON mendukung *embedding documents* dan *array*. BSON dapat mengandung tipe data selain data dengan spesifikasi yang sesuai dengan JSON. BSON memiliki tiga karakteristik yaitu *lightweight* (ukuran yang kecil), *transversable* (mudah dalam perpindahan data) dan *efficient* (proses *decoding* dan *encoding* cepat di berbagai bahasa pemrograman) (MongoDB 2015).

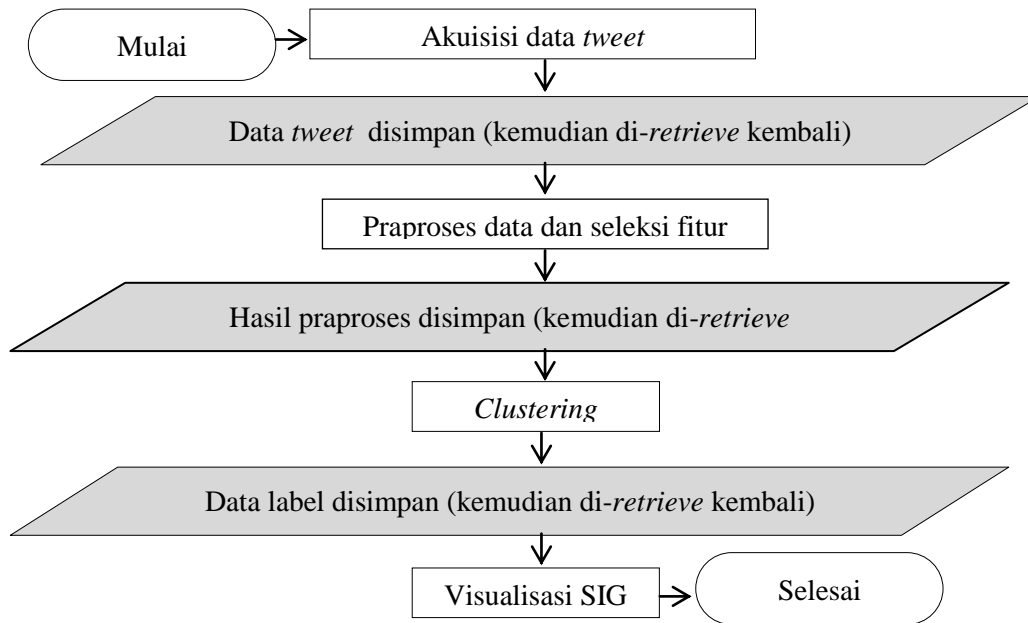
METODE

Data Penelitian

Data yang digunakan dalam penelitian ini adalah data Twitter menyangkut pertanian di Indonesia yang diambil dari penelitian yang dilakukan oleh Pratama (2015), Husnayain (2015) dan Denatari (2015).

Tahapan Penelitian

Penelitian ini merupakan bagian dari pengembangan Sistem Informasi Geografis Pertanian Indonesia. Tahapan-tahapan yang dilakukan pada pengembangan sistem dapat dilihat pada Gambar 1.



Gambar 1 Tahapan pengembangan SIG Pertanian Indonesia

Gambar 1 merupakan tahapan dari pengembangan Sistem Informasi Geografis Pertanian Indonesia. Ada lima tahapan pengembangan, yaitu, akuisisi data, praproses data, pertukaran data, *clustering*, dan visualisasi.

1 Akuisisi Data

Mendapatkan *tweets* mengenai kasus pertanian di Indonesia, dan membangun aplikasi akuisisi data Twitter berbasis web yang dapat mengategorikan data berdasarkan *timeline*, jumlah kata kunci, waktu, lokasi, dan gabungan ketiganya.

2 Praproses Data

Menerapkan metode *text mining* yang terdiri atas normalisasi teks, dan seleksi fitur hasil praproses disajikan dalam bentuk *term document matrix* (TDM)

3 Pertukaran Data

Melakukan proses penyimpanan data RStudio ke MongoDB, dan *retrieve* data dari MongoDB ke RStudio

4 *Clustering*

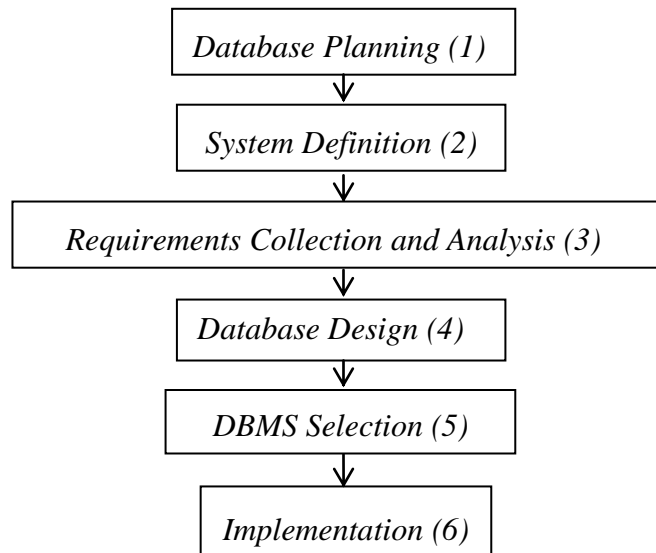
Proses *clustering* data teks Twitter untuk kasus pertanian di Indonesia sehingga dapat digunakan menganalisis kasus pertanian di Indonesia

5 Visualisasi

Menampilkan hasil *clustering* dalam bentuk *wordcloud* dan *maps*

Penelitian ini dikhususkan untuk menyimpan data pada *database*. Berdasarkan Conolly dan Begg (2015) tahapan *database system development lifecycle* memiliki 13 tahapan yaitu *database planning*, *system definition*, *requirements collection and analysis*, *database design*, *DBMS selection*, *application design*, *prototyping (optional)*, *implementation*, *data conversion and loading*, *testing*, dan *operational maintenance*.

Pada penelitian ini akan mengadopsi enam tahapan awal yaitu *database planning*, *system definition*, *requirement collection and analysis*, *database design*, *DBMS selection*, dan *implementation*. Tahapan penelitian dapat dilihat di Gambar 2.



Gambar 2 *Database system development lifecycle* (mengadopsi dari Connolly dan Begg 2015)

- 1 *Database planning*
Pengumpulan tujuan umum dari pengembangan *database* dan fungsi-fungsi apa saja yang perlu dipenuhi *database* untuk memenuhi kebutuhan.
- 2 *System definition*
Identifikasi cakupan dan batasan sistem *database*, serta mendefinisikan kebutuhan *database* dari berbagai sudut pandang pengguna.
- 3 *Requirement collection and analysis*
Proses pengumpulan dan analisis informasi mengenai organisasi untuk mendukung sistem *database*, kemudian menggunakan informasi tersebut untuk mengidentifikasi kebutuhan *database* baru. Kebutuhan untuk tiap pengguna terdiri dari deskripsi data yang digunakan/dihasilkan serta detail bagaimana suatu data digunakan/dihasilkan.
- 4 *Database design*
Tidak seperti SQL dimana skema *table* harus ditentukan sebelum insert data, pada MongoDB tidak memiliki struktur dokumen, sehingga fleksibel dalam memetakan dokumen ke dalam entitas maupun objek. *Design* dari model data mengikuti penggunaan data pada sistem (MongoDB 2015).

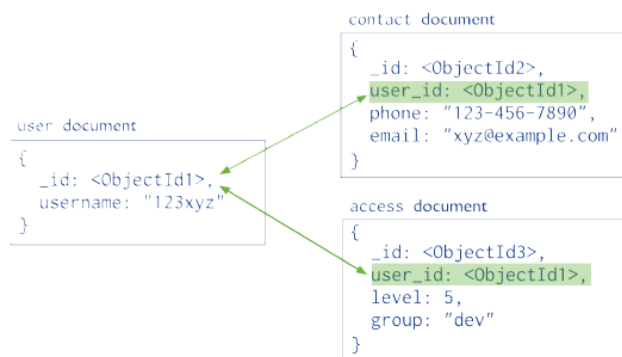
- Data Model Design
 - 1 *Embedded Data Models*



Gambar 3 *Embedded Data Models*

Pada Gambar 3, di dalam *field* “contact” memiliki *field* lain yaitu “phone” dan “email” dimana *field* tersebut dinamakan embedded sub-document. Demikian juga dengan *field* “access” yang memiliki *field* “level” dan “group” di dalamnya.

- 2 *Normalized Data Models*



Gambar 4 *Normalized Data Models*

Pada Gambar 4, terdapat tiga buah dokumen yaitu “user”, “contact” dan “access”, dimana pada dokumen “contact” dan “access” *field* “user_id” menggunakan “_id” dari dokumen “user”.

- Data Model Examples and Patterns
 - 1 *Model Relationships Between Documents Examples for modeling relationships between documents.*
 - 1.1 *Model One-to-One Relationships dengan Embedded Documents*
Menyajikan model data yang menggunakan *embedded documents* untuk mendeskripsikan *one-to-one relationships* antar data.

- 1.2 *Model One-to-Many Relationships with Embedded Documents*
Menyajikan model data yang menggunakan *embedded documents* untuk mendeskripsikan *one-to-many relationships* antar data.
- 1.3 *Model One-to-Many Relationships with Document References*
Menyajikan model data yang menggunakan *document references* untuk mendeskripsikan *one-to-many relationships* antar data.
- 2 *Model Tree Structures*
 - 2.1 *Model Tree Structures with Parent References*
Merepresentasikan model data yang mengorganisasikan dokumen dalam bentuk pohon, dengan menyimpan *references* ke node “parent” pada node “child”.
 - 2.2 *Model Tree Structures with Child References*
Merepresentasikan model data yang mengorganisasikan dokumen dalam bentuk pohon, dengan menyimpan *references* ke node “child” pada node “parent”.
- 3 *Model Specific Application Contexts Examples for models for specific application contexts.*
 - 3.1 *Model Data for Atomic Operations*
Mengilustrasikan *embedding fields* yang terkait dengan *atomic update* pada dokumen yang sama untuk memastikan semua dokumen tersinkronisasi.
 - 3.2 *Model Data to Support Keyword Search*
Penyimpanan *keyword* dalam bentuk *array* pada dokumen yang sama sebagai *text field*. Dikombinasikan dengan *multi-key index* untuk mendukung *keyword search*.
- 5 *DBMS selection*
Memilih *database* yang sesuai dengan kebutuhan dari sistem.
- 6 *Implementation*
Proses implementasi *database* ke organisasi yang membutuhkan *database*.

Lingkungan Pengembangan

Perangkat keras berupa komputer personal dengan spesifikasi sebagai berikut:

- Processor Intel Core i3.
- RAM 2 GB.
- 640 GB HDD.
- Mouse dan keyboard.

Perangkat lunak:

- MongoDB 2.6.5.
- OS Windows 8 Pro 64-bit.
- RStudio 0.98.1091.

HASIL DAN PEMBAHASAN

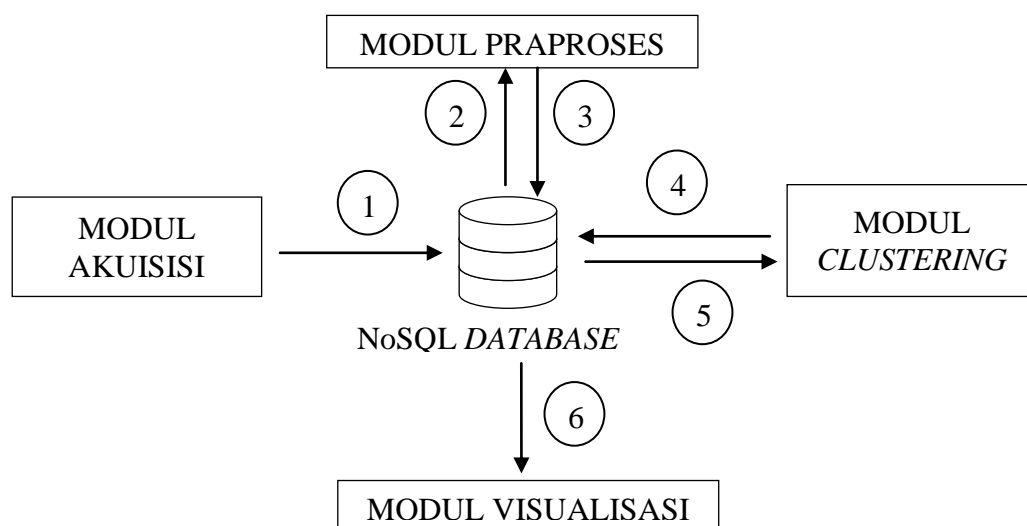
Database Planning

Pada tahapan ini dimulai dengan pengumpulan tujuan umum pembangunan *database*, yaitu untuk mempermudah pertukaran data antar modul pada pembangunan Sistem Informasi Geografik Pertanian Indonesia. Setelah tujuan umum terdefiniskan kemudian dilanjutkan dengan pengumpulan kebutuhan sistem dari segi pertukaran data di *database*. Kebutuhan ini didapatkan dari proses yang terdapat dalam sistem yaitu akuisisi, praproses data, *clustering* dan visualisasi. Berikut merupakan kebutuhan keempat proses terhadap *database*:

- 1 Menyimpan hasil akuisisi (Tahap akuisisi)
- 2 Mengambil hasil akuisisi (Tahap praproses)
- 3 Menyimpan hasil praproses (Tahap praproses)
- 4 Mengambil hasil praproses (Tahap *clustering*)
- 5 Menyimpan hasil *clustering* (Tahap *clustering*)
- 6 Mengambil hasil *clustering* (Tahap visualisasi)

System Definition

Tahapan *system definition* mengidentifikasi bagaimana kebutuhan *user* dipenuhi oleh *database system* dengan didefinisikan sebagai *job role* atau proses bisnis. Batasan pada sistem dapat dilihat pada Gambar 5.



Gambar 5 Batasan sistem

Pada Gambar 5 ada empat proses yang meliputi penggunaan *database*. Proses akuisisi akan menyimpan DATA ke dalam *database*. Proses praproses akan mengambil DATA, kemudian setelah dilakukan praproses hasilnya akan disimpan ke dalam TDM URL, TDM TWEET dan TDF. Proses *clustering* akan mengambil data TDM URL dan TDM TWEET kemudian hasil *clustering* akan disimpan

kembali kedalam *database*. Proses terakhir yaitu visualisasi akan mengambil DATA dan TDF untuk dilakukan visualisasi.

Requirement Collection and Analysis

Tahapan akuisisi

Tahapan akuisisi data dilakukan oleh Pratama (2015) yang menghasilkan 70 buah *tweet*. Dari 70 *tweet* tersebut memiliki 19 atribut. Atribut dari data *tweet* yang terambil dapat dilihat pada Tabel 5. Atribut nomor 1 sampai nomor 16 didapatkan secara otomatis dari Twitter menggunakan *library* *twitteR* dan tiga atribut yaitu URL, URL2, dan *content* ditambahkan secara manual.

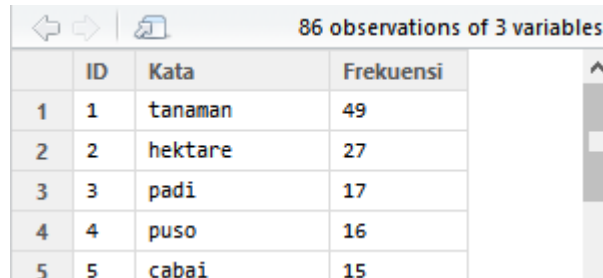
Tabel 5 Atribut pada data Twitter

No	Atribut	Keterangan
1	<i>Text</i>	Status yang pengguna buat
2	<i>Favorite</i>	<i>True</i> jika status merupakan <i>favorite</i>
3	<i>FavoritedCount</i>	Jumlah status yang difavoritkan
4	<i>ReplyToSN</i>	<i>Screen name</i> pengguna yang membalas status
5	<i>Created</i>	waktu status dibuat
6	<i>Truncated</i>	<i>True</i> jika status terpotong
7	<i>ReplyToSID</i>	ID status lain yang dibalas
8	<i>Id</i>	ID status
9	<i>ReplyToUID</i>	ID pengguna statusnya dibalas
10	<i>StatusSource</i>	Perantara sumber pengguna untuk <i>tweets</i>
11	<i>ScreenName</i>	<i>Screen name</i> pengguna yang memasang status
12	<i>RetweetCount</i>	Jumlah status yang di- <i>retweet</i>
13	<i>IsRetweet</i>	<i>True</i> jika merupakan status yang me- <i>retweet</i>
14	<i>Retweeted</i>	<i>True</i> jika merupakan status yang <i>retweet</i>
15	<i>Longitude</i>	Koordinat status di garis bujur
16	<i>Latitude</i>	Koordinat status di garis lintang
17	<i>URL</i>	URL yang terdapat pada <i>tweets</i>
18	<i>URL2</i>	URL pertama jika URL lebih dari satu
19	<i>Content</i>	Pemisahan konten dengan meta data

Tahapan praproses

Praproses dilakukan oleh Husnayain (2015) membutuhkan atribut *text* dan *content* dari hasil yang didapatkan oleh Pratama (2015). Atribut *text* di penelitian ini disebut data *tweet* karena praproses hanya dari teks twitter, sedangkan atribut *content* disebut data konten url karena praproses menggunakan data yang berasal dari *content* yang ada di URL pada suatu *tweet*. Dari 70 data yang dihasilkan Pratama (2015) akan diambil 51 buah *tweet* secara subyektif yang berhubungan dengan pertanian. Dari 51 *tweet* tersebut nantinya akan dipraproses sehingga

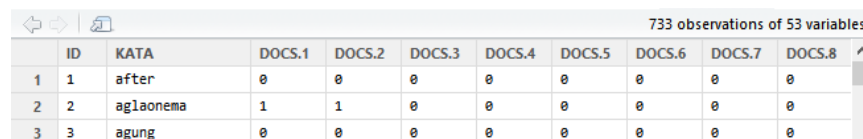
menghasilkan dua buah bentuk data, yaitu frekuensi perolehan *term* dari *tweet*/konten URL (*Term Document Frequency*/TDF) dan *term document matrix* (TDM). Contoh bentuk TDF yang didapat untuk data konten URL dapat dilihat pada Gambar 6.



	ID	Kata	Frekuensi
1	1	tanaman	49
2	2	hektare	27
3	3	padi	17
4	4	puso	16
5	5	cabai	15

Gambar 6 TDF data konten URL

Pada Gambar 6 terlihat bahwa *term* “tanaman” memiliki jumlah kemunculan kata sebanyak 49 buah, kata “hektare” memiliki jumlah kemunculan kata sebanyak 27 buah, kata “padi” sebanyak 17 buah, kata “puso” sebanyak 16 buah dan kata “cabai” sebanyak 15 buah. Selain perolehan *term* dari *tweet*/konten URL, tahapan praproses juga menghasilkan TDM. TDM merupakan matriks jumlah kemunculan suatu kata pada dokumen. Baris matriks menunjukkan kata yang ada pada data *tweet* dan data konten URL, sedangkan kolom matriks menunjukkan dokumen yang ada pada data tersebut. Contoh data konten URL dapat dilihat pada Gambar 7.



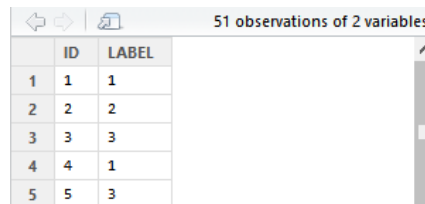
	ID	KATA	DOCS.1	DOCS.2	DOCS.3	DOCS.4	DOCS.5	DOCS.6	DOCS.7	DOCS.8
1	1	after	0	0	0	0	0	0	0	0
2	2	aglaonema	1	1	0	0	0	0	0	0
3	3	agung	0	0	0	0	0	0	0	0

Gambar 7 TDM konten URL

Pada Gambar 7 menunjukan TDM URL sendiri terdiri dari 733 kata dan terdapat 53 *variable*. Dicontohkan untuk kata “after” pada DOCS 1 adalah 0, yang artinya adalah pada *tweet* ke 1 tidak memiliki kata “after” di dalamnya, kata “after” dapat ditemukan pada *tweet* ke 10 dan ke 51 dengan masing-masing berjumlah satu buah.

Tahapan *clustering*

Clustering dilakukan oleh Denatari (2015) membutuhkan TDF yang dihasilkan oleh Husnayain (2015). Tahapan *clustering* menghasilkan label dari masing-masing *tweet*. Bentuk data yang dihasilkan dapat dilihat pada Gambar 8.



	ID	LABEL
1	1	1
2	2	2
3	3	3
4	4	1
5	5	3

Gambar 8 Label data

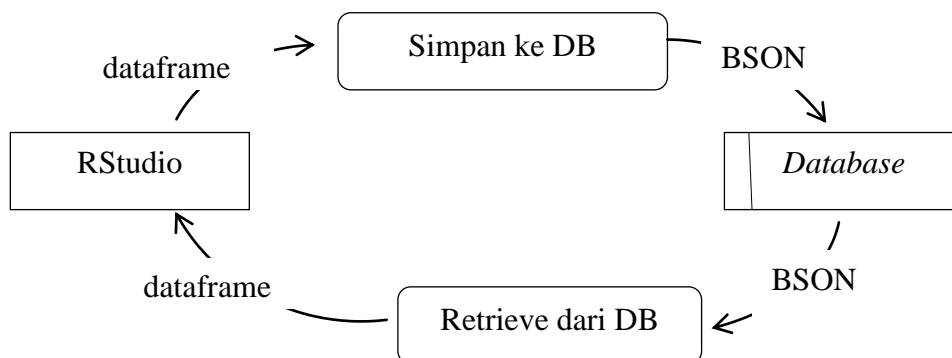
Pada Gambar 8, jumlah data ada sebanyak 51 buah karena telah dikurangi dengan *tweet* yang dinilai secara subyektif tidak relevan dengan pertanian. Atribut LABEL merupakan nomor dari *cluster* sedangkan ID merupakan ID *tweet*, sehingga pada baris ke 1 *tweet* ke-1 masuk ke dalam *cluster* nomor 1.

Tahapan visualisasi

Visualisasi dilakukan oleh Husein (2015) membutuhkan atribut TEXT, CREATED yang berasal dari penelitian Pratama (2015), data TDF dari penelitian Husnayain (2015) dan LABEL dari penelitian Denatari (2015).

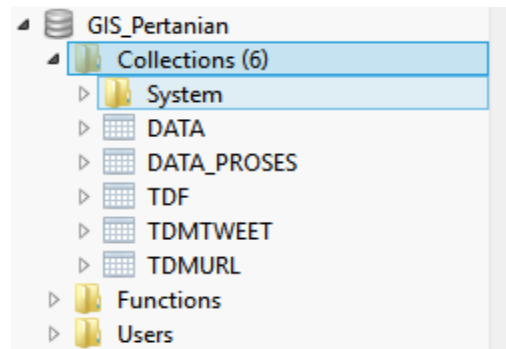
Database Design

Fokus utama dari penelitian ini adalah untuk menghubungkan *database* pada RStudio dan MongoDB, adapun bentuk aliran data dapat dilihat pada Gambar 9.

Gambar 9 Alur transaksi data dari RStudio ke *database* NoSQL

Pada Gambar 9, data pada RStudio berupa *dataframe* akan diubah dahulu menjadi BSON agar dapat dimasukkan ke dalam *database*, sedangkan untuk melakukan pengambilan data ke RStudio, data diubah dahulu dari BSON ke dalam *dataframe*, proses memasukan dan mengambil data menggunakan *package* “*rmongodb*”.

Data yang dihasilkan oleh Pratama (2015), Husnayain (2015), dan Denatari (2015) akan disimpan ke dalam lima buah *collection* pada *database* GIS_Pertanian. Struktur *database* yang akan dihasilkan terdapat pada Gambar 10.

Gambar 10 Struktur *database*

Pada Gambar 10 dapat dilihat ada empat buah *collection* yaitu DATA, DATA_PROSES, TDF, TDMTWEET dan TDMURL. Dikarenakan bentuk data yang tidak memiliki repetisi dan datar sehingga tidak menggunakan *embedded*, *reference* maupun *tree* dalam perancangannya. Bentuk data yang disimpan di MongoDB tetap sama dengan data yang dihasilkan di RStudio agar mempermudah pengolahan. Bentuk data TDMURL dan TDMTWEET memiliki bentuk yang sama, terdiri dari ID, KATA, DOCS ke-n (Pada TDMTWEET disebut ID ke-n). Berikut adalah bentuk data TDM:

```
{
  "ID" : "727",
  "KATA": "yogyakarta",
  "DOCS 1" : "0",
  "DOCS 2" : "0",
  "DOCS 3" : "0",
  "DOCS 4" : "0",
  "DOCS 5" : "0",
}
```

Collection TDF terdiri dari ID, Kata, dan Frekuensi. Data pada *collection* TDF merupakan hasil dari proses praproses. Berikut adalah bentuk data pada *collection* TDF:

```
{
  "ID" : "25",
  "Kata": "tanah",
  "Frekuensi" : "5"
}
```

Bentuk dokumen pada *collection* DATA didapat dari hasil akuisisi dan memiliki 18-19 atribut seperti pada Tabel 2 (tergantung ada atau tidaknya URL2), sedangkan *collection* DATA_PROSES ada *field* LABEL yang merupakan hasil *clustering* dan memiliki jumlah dokumen lebih sedikit karena telah dipilih secara manual oleh Husnayain (2015) dan Denatari (2015).

```

{
  "ID": "5.97E+17",
  "TEXT": "Diskusi Umum Re Cara Menanam Cabai Terpedas dari
  Biji UmyDylan menulis Butuh berapa hari ya untuk muncul
  http t co DQPdXiCD9D",
  "FAVORITED": "FALSE",
  "FAVORITECOUNT": "0",
  "REPLYTOSN": "NA",
  "CREATED": "05-10-2015 7 20 00 AM",
  "TRUNCATED": "FALSE",
  "REPLYTOSID": "NA",
  "REPLYTOUID": "NA",
  "STATUSSOURCE": " a href http twitterfeed com rel nofollow
  >twitterfeed< a>",
  "SCREENNAME": "tanaman id",
  "RETWEETCOUNT": "0",
  "ISRETWEET": "FALSE",
  "RETWEETED": "FALSE",
  "LONGITUDE": "NA",
  "LATITUDE": "NA",
  "URL": "http t co DQPdXiCD9D",
  "CONTENT": "TANAMAN\nCARA MENANAM CABAI TERPEDAS DARI
  BIJI\nPeraturan Forum Selamat Datang di Forum Diskusi
  Umum\nForum ini adalah ajang diskusi sharing pengalaman
  dan tanya jawab"
  "LABEL": "1"
}

```

DBMS Selection

Tahapan selanjutnya dalam penelitian ini adalah *DBMS selection*, ada dua jenis *database* yang berkembang saat ini yaitu *relational database* dan *NoSQL database*. *NoSQL database* memiliki banyak kekurangan dari segi kehandalan (Tabel 1) dibandingkan dengan *relational database*, tetapi lebih cocok digunakan untuk data media sosial karena strukturnya yang dapat mengikuti perubahan bentuk data, oleh karena itu penelitian ini akan menggunakan *database NoSQL* dalam pembangunan *database*.

NoSQL database memiliki empat jenis *database* yaitu *database NoSQL* berbasis *key-value*, *database NoSQL* berbasis *document*, *database NoSQL* berbasis *column* dan *graph-oriented*. Berdasarkan Tabel 3, *database NoSQL* berbasis *key-value* tidak digunakan karena dari segi penyimpanan hanya dapat menyimpan satu jenis *object* saja, tidak sesuai dengan karakteristik data media sosial yang beragam, *database NoSQL* berbasis *document* dapat digunakan penyimpanan *multiple object* hal tersebut sesuai dengan karakteristik data media sosial, *database NoSQL* berbasis *column* tidak digunakan karena biasa digunakan untuk data yang dapat dipartisi, tidak sesuai dengan karakteristik data media sosial sedangkan *graph-oriented* biasa digunakan untuk untuk memperlihatkan keterhubungan antar *object* pada *database*, karena penelitian ini tidak menggambarkan hubungan antar *object* maka *graph-oriented* tidak digunakan. Penelitian ini akan menggunakan *database NoSQL* berbasis *document* karena karakteristiknya yang sesuai dengan data media sosial.

NoSQL berbasis dokumen paling populer yaitu CouchDB dan MongoDB, fungsi dan karakteristik keduanya tidak banyak memiliki perbedaan (Tabel 4).

Karena pengolahan data dari media sosial pada sistem GIS Pertanian Indonesia menggunakan RStudio, sehingga perlu dilihat dukungan dari *database* untuk penggunaan RStudio sebagai aplikasi pengolahan data. CouchDB mendukung untuk penggunaan RStudio $\geq 3.0.0$ dengan *package* 'couchDB', sedangkan MongoDB dapat digunakan untuk RStudio $\geq 2.1.0$ dengan *package* 'rmongodb'. Dari segi ukuran maksimum dari *document* maupun *collection* MongoDB tidak memiliki batasan, sedangkan untuk CouchDB apabila ukuran melebihi *default* yang ditentukan, sebelumnya harus dikonfigurasi dahulu di awal. Dari pertimbangan hal-hal tersebut maka penelitian ini menggunakan MongoDB karena lebih sesuai dengan karakteristik data yang mengalami perpindahan cepat antar modul dan ukurannya yang besar.

Implementation

Fungsi *conection*

Untuk dapat dilakukan pengolahan data maka dilakukan proses menyambungkan antara RStudio dan MongoDB. Di tahapan ini diperlukan *package* 'rmongodb' sebagai penghubung antara RStudio dan MongoDB. Dibangunlah fungsi *conection* untuk memenuhi kebutuhan tersebut.

```
> conection <- function()
+ {
+   library(rmongodb)
+   mongo = mongo.create(host = "localhost")
+   mongo.is.connected(mongo)
+   mongo.get.databases(mongo)
+ }
> conection()
[1] "Example"      "GIS_Pertanian" "MediaSosial"   "test"         "dbs"
> |
```

Gambar 11 Fungsi *conection* antara RStudio dan MongoDB

Pada Gambar 11 terlihat bahwa koneksi berhasil dilakukan dan pada MongoDB telah ada 5 buah *database*, yaitu Example, GIS_Pertanian, MediaSosial, test dan dbs. *Database* yang digunakan pada penelitian ini adalah GIS_Pertanian.

Fungsi *insert* data dari RStudio ke MongoDB

Setelah terhubung antara RStudio dan MongoDB, proses penyimpanan dapat dilakukan. Gambar 12 merupakan contoh penggunaan fungsi *insert* untuk memasukan data.

```

> insert <- function (var, dataset, dbs.dbc)
+ {
+   colnames(dataset)[var] <- "ID"
+   datalist <- list()
+   datalist <- apply( dataset, 1, function(x) c( datalist, x ) )
+   res <- lapply( datalist, function(x) mongo.bson.from.list(x) )
+   if (mongo.is.connected(mongo)) {
+     mongo.insert.batch(mongo, "dbs.dbc", res )}
+ }
> insert(53, tdm_url, GIS_Pertanian.TDMURL)
[1] TRUE
> insert(53, tdm_tweet, GIS_Pertanian.TDTWEET)
[1] TRUE
> insert(3, frekuensi_kata, GIS_Pertanian.TDF)
[1] TRUE
> |

```

Gambar 12 Fungsi *insert* ke *database*

Pada Gambar 12 proses *insert* telah berhasil dilakukan untuk *dataset* *tdm_url* dengan *variable* sebanyak 53 buah ke lokasi tujuan *database* GIS_Pertanian dan *collection* TDMURL, *dataset* *tdm_tweet* dengan *variable* sebanyak 53 buah ke lokasi tujuan *database* GIS_Pertanian dan *collection* TDMTWEET, serta *frekuensi_kata* dengan *variable* sebanyak 3 buah ke lokasi tujuan *database* GIS_Pertanian dan *collection* TDF.

Fungsi *retrieve* data ke RStudio dari MongoDB

Pada tahapan ini data yang sudah disimpan di MongoDB, dimasukkan kembali ke MongoDB untuk dilakukan proses *clustering*. Proses *clustering* menggunakan *collection* TDMTWEET dan TDMURL, berikut adalah contoh penggunaan fungsi untuk mengambil TDMURL

```

> retrieve_col <- function (dbs, dbs.dbc, dbc)
+ {
+   library(plyr)
+   mongo.get.database.collections(mongo, db = "dbs")
+   DBNS = "dbs.dbc"
+   dbc = data.frame(stringsAsFactors = FALSE)
+   cursor = mongo.find(mongo, DBNS)
+   while (mongo.cursor.next(cursor))
+   {
+     tmp = mongo.bson.to.list(mongo.cursor.value(cursor))
+     tmp.df = as.data.frame(t(unlist(tmp)), stringsAsFactors = F)
+     dbc = rbind.fill(dbc, tmp.df)
+   }
+ }
> retrieve_col(GIS_Pertanian, GIS_Pertanian.TDMURL, TDMURL)
> |

```

Gambar 13 Fungsi *retrive* *collection* TDMURL

Pada Gambar 13 berhasil diambil data TDMURL dari lokasi *database* GIS_Pertanian dan *collection* TDMURL. Data yang berupa BSON pada *database* diubah dulu kedalam list untuk kemudian diubah lagi ke dalam bentuk *dataframe* untuk mempermudah pengolahan data oleh modul lain. Hasil *dataset* yang berhasil diambil dapat dilihat pada Gambar 14.

_id	ID	KATA	DOCS 1	DOCS 2	DOCS 3
74394216	727	yogyakarta	0	0	0
0	728	yup	1	1	0
0	729	walang	0	0	0
0	730	warna	2	2	0
0	731	wartawan	0	0	0

Gambar 14 Hasil *retrieve* TDMURL

TDMURL pada Gambar 14 mengalami perubahan bentuk data dari data semula yang didapat dari tahapan praproses (Gambar 7), terdapat atribut tambahan yaitu *_id*. Atribut *_id* terbentuk secara otomatis di MongoDB ketika suatu data disimpan dan otomatis menjadi *primary key* pada *collection*.

SIMPULAN DAN SARAN

Simpulan

Penelitian ini berhasil membangun *database* sederhana untuk SIG ber*database* pertanian dari media sosial Twitter. *Database* yang dihasilkan memiliki lima *collection* yaitu TDF, TDMTWEET, TDMURL, DATA dan DATA_PROSES. *Collection* DATA semua *field* berasal dari proses akuisisi sedangkan *collection* TDF, TDMTWEET dan TDMURL berasal dari proses praproses, sedangkan *collection* DATA_PROSES merupakan hasil dari praproses dan *clustering* yang diseleksi manual secara subjektif mengenai keterkaitannya dengan bidang pertanian. Penelitian ini juga sudah berhasil membangun tiga fungsi untuk *retrieve* dan *insert* data serta dapat *connection* antara RStudio dan MongoDB untuk memudahkan proses transportasi data. Dengan adanya fungsi-fungsi yang telah mempermudah proses pertukaran dan integrasi data pada SIG Pertanian.

Saran

Penelitian ini masih memiliki kekurangan yaitu belum dilakukan pengoptimalan *query* dan *indexing* serta belum dapat menangani data *stream*. Penelitian yang akan datang diharapkan dapat menerapkan pengoptimalan *query* dan *indexing* serta mengembangkan lebih lanjut untuk penanganan data *stream* sehingga dapat mendukung pertukaran data antara RStudio dan MongoDB lebih baik lagi.

DAFTAR PUSTAKA

- Adler RB, Rodman G. 2006. *Understanding Human Communication*. New York (US): Oxford University Press.
- Boyd D, Ellison N. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*. [Internet][diunduh 2015 Jan 1]; 13(1) Tersedia pada <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00393.x/full/>
- Cattell R. 2010. Scalable SQL and NoSQL data stores. *ACM SIGMOD* [Internet][diunduh 2015 Jan 8]; 39(4): 12–27. Tersedia pada <http://www.sigmod.org/publications/sigmod-record/1012/pdfs/04.surveys.cattell.pdf>.
- Chodorow K, Dirolf M. 2010. *MongoDB: The Definitive Guide*. Sebastopol(US): O'Reilly Media Inc.
- Connolly TM, Begg CE. 2015. *Database Systems: A Practical Approach to Design, Implementation, and Management*. Ed ke-6. Edinburgh Gate(UK): Pearson Education Limited.
- Denatari RB. 2015. Clustering data teks Twitter untuk kasus pertanian di Indonesia [Skripsi]. Bogor (ID): Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Husein MN. 2015. Visualisasi hasil clustering data berbasis media sosial untuk kasus pertanian di Indonesia [Skripsi]. Bogor (ID): Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Husnayain I. 2015. Praproses data teks Twitter pada Sistem Informasi Geografis untuk kasus pertanian di Indonesia [Skripsi]. Bogor (ID): Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- [Kemenkominfo]. 2013. Kominfo: Pengguna internet di Indonesia 63 juta orang [Internet]. [diakses 2015 Jan 08]. Tersedia pada: http://kominfo.go.id/index.php/content/detail/3415/Kominfo+%3A+Pengguna+Internet+di+Indonesia+63+Juta+Orang/0/berita_satker.
- [Kemenkominfo]. 2014. Kominfo: Pengguna internet Indonesia nomor enam dunia [Internet]. [diakses 2015 Jan 08]. Tersedia pada: http://kominfo.go.id/index.php/content/detail/4286/Pengguna+Internet+Indonesia+Nomor+Enam+Dunia/0/sorotan_media#.VK-dlyusW3g.
- Mohamed AM, Altrafi OG, Ismail MO. 2014. Relational vs. NoSQL databases: A survey. *International Journal of Computer and Information Technology*. [Internet][diunduh 2015 Jan 1]; 3(3). Tersedia pada <http://www.ijcit.com/archives/volume3/issue3/Paper030320.pdf>.
- [MongoDB]. 2015. MongoDB Documentation Release 3.0.7. [internet] [diunduh 2015 Nov 15] Tersedia pada <https://docs.mongodb.org/v3.0/MongoDB-manual-v3.0.pdf>.
- Nayak A, Poriya A, Poojary D. 2013. Type of NoSQL databases and its comparison with relational databases. *International Journal of Applied Information Systems*. [Internet][diunduh 2014 Des 1]; 5(4):16-19. Tersedia pada <https://www.ijais.org/archives/volume5/number4/434-0888>.
- Prajapati V. 2013. *Big Data Analytics with R and Hadoop*. Birmingham(UK): Packt Publishing Ltd.

sebelum kurung berikan spasi

skripsi itu "s"-nya kecil

- Pratama FS. 2015. Akuisisi data pada sistem informasi geografis berbasis Twitter untuk kasus pertanian di indonesia [skripsi]. Bogor (ID): Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- Ramalho LG. 2011. From ISIS to CouchDB: Databases and data models for bibliographic records. *The Code 4 Lib Journal*. [Internet][diunduh 2015 Jan 1]; 13(7). Tersedia pada <http://journal.code4lib.org/articles/4893#merriman2010>.
- Rhoades EB, Hall K. 2007. The agricultural blogosphere: A snapshot of new agricultural communicators online. *Journal of Applied Communications*, 91(3 dan 4): 37-56.

RIWAYAT HIDUP

Penulis dilahirkan di Yogyakarta pada tanggal 7 September 1993 dan merupakan anak pertama dari dua bersaudara, dari Agus Fahri Husein, S.S dan Eli Rosidah, S.H. Pada tahun 2011, penulis menamatkan pendidikan di SMA Negeri 2 Krakatau Steel Cilegon. Penulis lulus seleksi masuk Institut Pertanian Bogor (IPB) pada tahun yang sama melalui jalur SNMPTN Undangan dan diterima sebagai mahasiswa di Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam.

Selama aktif menjadi mahasiswa, penulis merupakan anggota Divisi Internal Himpunan Mahasiswa Ilmu Komputer 2013/2014. Penulis aktif mengikuti berbagai kegiatan kemahasiswaan sebagai panitia, salah satunya menjadi Kepala Divisi *Fundraising* pada IT Today 2013. Penulis juga melaksanakan kegiatan Praktik Kerja Lapangan di Pusat Data, Informasi dan Standardisasi Badan Pengkajian dan Penerapan Teknologi (PDIS BPPT), Serpong pada Juni - Agustus 2014.