TUGAS AKHIR - KS141501

Identifikasi Entitas Data pada Industri Diving untuk Menstrukturkan Informasi menggunakan Analisis Frekuensi

Structuring Information of Diving Industry by Identifying Data Entity using Frequency Analysis

ROBBIGH FAUBENDRI
NRP 5213 100 164

Dosen Pembimbing I
Rully Agus Hendrawan, S.Kom., M.Eng.

Dosen Pembimbing II
Amna Shifia Nisafani, S.Kom, M.Sc

DEPARTEMEN SISTEM IN*FORM*ASI
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2018

TUGAS AKHIR – KS141501

Identifikasi Entitas Data pada Industri Diving untuk Menstrukturkan Informasi menggunakan Analisis Frekuensi

ROBBIGH FAUBENDRI
NRP 5213 100 164

Dosen Pembimbing I
Rully Agus Hendrawan, S.Kom., M.Eng.

Dosen Pembimbing II
Amna Shifia Nisafani, S.Kom, M.Sc

DEPARTEMEN SISTEM IN*FORMA*SI
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember
Surabaya 2018

FINAL PROJECT – KS141501

Structuring Information of Diving Industry by Identifying Data Entity using Frequency Analysis

ROBBIGH FAUBENDRI
NRP 5213 100 164

Supervisor I
Rully Agus Hendrawan, S.Kom., M.Eng.

Supervisor II
Amna Shifia Nisafani, S.Kom, M.Sc

INFORMATION SYSTEMS DEPARTMENT
Faculty of Information and Communication Technology
Sepuluh Nopember Institute of Technology
Surabaya 2018

LEMBAR PENGESAHAN

IDENTIFIKASI ENTITAS DATA PADA INDUSTRI DIVING UNTUK MENSTRUKTURKAN INFORMASI MENGGUNAKAN ANALISIS FREKUENSI

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember

Oleh :

ROBBIGH FAUBENDRI
NRP 5213 100 164

Surabaya, 9 Januari 2018

PLH KETUA
DPEARTEMEN SISTEM INFORMASI

Edwin Riksakomara, S.Kom, MT.
NIP 196907252003131001

LEMBAR PERSETUJUAN

IDENTIFIKASI ENTITAS DATA PADA INDUSTRI DIVING UNTUK MENSTRUKTURKAN INFORMASI MENGGUNAKAN ANALISIS FREKUENSI

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Informasi dan Komunikasi
Institut Teknologi Sepuluh Nopember

Oleh :

ROBBIGH FAUBENDRI
NRP 5213 100 164

Disetujui Tim Penguji : Tanggal Ujian : 10 Januari 2017
Periode Wisuda : Maret 2017

Rully A. Hendrawan, S.Kom., M.Eng	**(Pembimbing I)**
Amna Shifia Nisafani, S.Kom, M.Sc	**(Pembimbing II)**
Arif Wibisono, S.Kom., M.Sc	**(Penguji I)**
Mahendrawati Er., S.T., M.Sc., Ph.D	**(Penguji II)**

Identifikasi Entitas Data pada Industri Diving untuk Menstrukturkan Informasi menggunakan Analisis Frekuensi

Nama Mahasiswa :Robbigh Faubendri
NRP :5213 100 164
Jurusan :Sistem Informasi FTIF-ITS
Pembimbing 1 :Rully Agus Hendrawan, S.Kom., M.Eng.
Pembimbing 2 :Amna Shifia Nisafani, S.Kom, M.Sc

ABSTRAK

Diving adalah sebuah aktivitas wisata yang dapat dinikmati oleh para turis dengan cara menyelam dan menikmati alam bawah laut. Dalam melaksanakan wisata diving, tentunya perlu untuk mengetahui informasi - informasi terkait dengan akitivitas tersebut. Informasi yang diperlukan seperti wilayah yang berpotensi untuk di datangi, jenis terumbu karang dan ikan yang tersedia, penginapan yang terdekat, instruktur diving yang tersedia, informasi sertifikasi yang dapat dilakukan, dan banyak lagi informasi yang harus didapatkan jika ingin melakukan diving. Semua informasi tersebut membutuhkan suatu situs yang berfungsi sebagai wadah yang mudah diakses oleh para diver. Kemudahan pengguna pengguna memahami informasi yang terdapat dalam situs tersebut dibutuhkan kondisi informasi yang terstruktur. Dalam menyediakan informasi terstruktur, perlunya dilakukan penenetuan entitas data yang tepat

Oleh karena itu, dalam penelitian kali ini akan dilakukan proses Idenftifikasi entitas data yang akan dilakukan berdasarkan beberapa website kajian yang berkaitan dengan industri diving. Sebelum dilakukannya perancangan, akan dilakukan proses ektraksi data atau informasi pada tiap halaman – halaman webiste yang dijadikan sebagai kajian. Data yang didapatkan akan di pecah menjadi kata kunci. Kata kunci tersebut akan di hitung berdasarkan frekuensi dan akan dilakukan analisis frekuensi. Idenftifikasi Entitas Data akan mengacu kepada

standarisasi metadata. Penggunaan standarisasi metadata akan disesuaikan dengan kebutuhan sistem dan industri diving. Setelah menghasilkan sebuah entitas data, maka dilakukan validasi model. Validasi model akan dilakukan dengan pemeriksaan mengguanakan Quality of Assurance.

Hasil analisa dan pereancangan dari model data berdasarkan website sebagai sumber data, didapatkan entitas sebanyak 14 entitas. Website yang digunakan adalah sebanyak 27 website Entitas dan atribut sudah disesuaikan dengan standar schema.org.

Kata Kunci : Entitas Data, Analisis Frekuensi, Standarisasi, Validasi, Quality of Assurance

STRUCTURING INFORMATION OF DIVING INDUSTRY BY IDENTIFYING DATA ENTITY USING FREQUENCY ANALYSIS

Nama Mahasiswa : Robbigh Faubendri
NRP : 5213 100 164
Jurusan : Sistem Informasi FTIF-ITS
Pembimbing I : Rully Agus Hendrawan, S.Kom., M.Eng.
Pembimbing II : Amna Shifia Nisafani, S.Kom, M.Sc

ABSTRACT

Diving is an activity that can be enjoyed by the tourists by way of diving and enjoy the underwater world. In carrying out diving tours, it is necessary to know the information related to the activity. Required information such as potential areas to go to, types of coral reefs and fish available, nearby lodging, available diving instructors, certification information that can be done, and more information to be obtained if you want to dive. All such information requires a site that serves as a container that is easily accessible to the diver. The ease with which users understand the information contained in the site requires a structured information condition. In providing structured information, the need for penenetuan schema and data model right.

Therefore, in this research this will be done process of data model design which will be done based on several research website related to diving industry. Before the design, the process of extraction of data or information will be done on every web page that will be used as data source. That data will be separate and generate keywords. That keyword will be calculated in frequency. Data enitity design will refer to metadata standardization. The use of metadata standardization will be adjusted to the needs of the diving system and industry. After generating a data entity, model validation is done. Validation of the model will be done by checking the Quality of Assurance.

The results of analysis and design of the data model based on the website as a data source, obtained as many as 15 entities. Website used is as much as 27 websites Entities and attributes have been adapted to schema.org standards.

Keyword : Model Data, Standarization, Validation, Quality of Assurance

KATA PENGANTAR

Alhamdulillah atas karunia, rahmat, barakah, dan jalan yang telah diberikan Allah SWT selama ini sehingga penulis mendapatkan kelancaran dalam menyelesaikan tugas akhir dengan judul:

Identifikasi Entitas Data pada Industri Diving untuk Menstrukturkan Informasi menggunakan Analisis Frekuensi

Terima kasih atas pihak-pihak yang telah mendukung, memberikan saran, motivasi, semangat, dan bantuan baik materi maupun spiritual demi tercapainya tujuan pembuatan tugas akhir ini. Secara khusus penulis akan menyampaikan ucapan terima kasih yang sedalam-dalamnya kepada:

1. Allah SWT yang senatiasa melimpahkan berkah dan rahmat-Nya selamat penulis menempuh masa perkuliahan
2. Bapak Dr. Ir. Aris Tjahyanto, M.Kom selaku Ketua Jurusan Sistem Informasi ITS Surabaya
3. Bapak Rully Agus Hendrawan, S.Kom., M.Eng. dan Ibu Amna Shifia Nisafani, S.Kom, M.Sc selaku dosen pembimbing yang meluangkan waktu, memberikan ilmu, petunjuk, dan motivasi untuk kelancaran tugas akhir ini.
4. Bapak Arif Wibisono, S.Kom., M.Sc. selaku dosen wali penulis yang selalu memberikan motivasi, dan masukkan sehingga penulis terus mengusahakan yang terbaik selama perkuliahan dan pengerjaan Tugas Akhir.
5. Daddy, Umi, Abang, dan adek yang telah memberikan motivasi, semangat, keyakinan, dan doa sehingga penulis dapat menyelesaikan pendidikan S1 ini dengan baik
6. Nurfadillah Rohmatin Anya, yang telah membantu dan mendukung penulis di masa – masa perkuliahan hingga sekarang dengan sabar.
7. Keluarga Umi dan Abi di Surabaya, yang penulis anggap sebagai keluarga sendiri. Sudah membantu penulis semasa perkuliahan
8. Seluruh dosen Jurusan Sistem Informasi ITS yang telah memberikan ilmu yang sangat berharga bagi penulis.

9. Harun, Fahrizal, dan Eki, yang telah membantu penulis dalam melaksanakan tugas akhir ini.
10. Ari, Kevin, Ashma, Ervi, Safrin, Rani, dan Dhara yang merupakan teman seperjuangan terdekat penulis selama menjadi mahasiswa yang selalu menemani dan memberikan semangat dari awal perkuliahan hingga penyelesaian tugas akhir ini.
11. Alvin dan Adi, teman seperjuangan semester akhir, sama – sama berjuang hingga titik darah penghabisan.
12. Tim Pasta Kangen yang telah mendukung saya dan semoga kedepannya kita makin maju.
13. Berbagai pihak yang membantu dalam penyusunan Tugas Akhir ini dan belum dapat disebutkan satu per satu dengan dukungan, semangat, dan kebersamaan.

Penyusunan laporan ini masih jauh dari sempurna, untuk itu saya menerima adanya kritik dan saran yang membangun untuk perbaikan di masa mendatang. Semoga buku tugas akhir ini dapat memberikan manfaat pembaca.

Surabaya, 8 Januari 2018
Penulis,

(Robbigh Faubendri)

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

Pada bab pendahuluan akan diuraikan proses identifikasi masalah penelitian yang meliputi latar belakang masalah, perumusan masalah, batasan masalah, tujuan tugas akhir, manfaat kegiatan tugas akhir dan relevansi terhadap pengerjaan tugas akhir. Berdasarkan uraian pada bab ini, harapannya gambaran umum permasalahan dan pemecahan masalah pada tugas akhir dapat dipahami.

1.1 Latar belakang

Kawasan segitiga terumbu karang merupakan suatu permukaan bumi yang mempengaruhi kehidupan laut sehingga memiliki keberagaman jenis biota laut yang ada di sekitarnya. Hal tersebut dapat dimanfaatkan menjadi wisata alam bawah laut yang dapat dinikmati oleh para turis dalam negeri maupun luar negeri. Indonesia, Timor Leste, Filipina, Sabah – Malaysia, Papua Niugini, dan Kepulauan Soloman adalah wilayah yang termasuk dalam segitiga terumbu karang [1]. Salah satu aktivitas wisata yang dapat dilakukan adalah *diving*. Wisata *diving* merupakan aktivitas menikmati alam bawah laut dengan cara menyelam dan menggunakan peralatan yang lengkap seperti baju selam, sepatu katak, masker kacamata, jaket penyangga oksigen, tabung oksigen, regulator, dan pemberat badan [2].

Dalam melaksanakan wisata *diving*, tentunya perlu untuk mengetahui informasi – informasi terkait dengan aktivitas tersebut. Informasi yang diperlukan seperti wilayah yang berpotensi untuk di datangi, jenis terumbu karang dan ikan yang tersedia, penginapan yang terdekat, instruktur *diving* yang tersedia, informasi sertifikasi yang dapat dilakukan, dan banyak lagi informasi yang harus didapatkan jika ingin melakukan *diving*. Semua informasi tersebut membutuhkan suatu situs yang berfungsi sebagai wadah yang mudah diakses oleh para *diver*. Kemudahan pengguna penggna memahami informasi yang terdapat dalam situs tersebut dibutuhkan kondisi informasi

yang terstruktur. Di internet, sudah terdapat beberapa *website* yang menyediakan informasi mengenai *diving*, tetapi informasi yang ada belum tersetruktur. Pada *website Bali Reef Divers* [3], menjelaskan mengenai tempat wisata yang ada di Bali. Saat melihat halaman *website* tersebut, informasi – informasi penting yang ada ditampilkan secara kalimat yang panjang dan tidak menampilkan informasi penting secara langsung. Sehingga membutuhkan waktu untuk membaca dan memahami terlebih dahulu.

Oleh karena itu, dalam penelitian kali ini akan dilakukan proses identifikasi entitas data yang akan dilakukan berdasarkan beberapa *website* kajian yang berkaitan dengan industri *diving*. Maka dari itu, diperlukan metode identifikasi entitas data dengan *website* sebagai sumber data. Pada tiap halaman *website* akan didapatkan sebuah kata – kata yang penting dalam penyampaian suatu informasi terkait dengan industri *diving.* Untuk mendapatkan kata – kata penting yang ada pada *website*, dilakukan dengan metode *web scrapping.* Supaya model data ini dapat dimanfaatkan seluas – luasnya maka kami membangunnya mengacu pada standar. Setiap entitas diturunkan dari skema standar. Setelah dilakukan Idenftifikasi Entitas Data, maka dilakukan validasi model. Validasi model akan dilakukan dengan pemeriksaan mengguanakan *Quality of Assurance*.

Penelitian ini diharapkan dapat menjadi gambaran bagi para pengembang dalam membuat suatu sistem informasi untuk industri *diving.* Model data yang dihasilkan dari beberapa *website* kajian, dapat menjadikan sistem informasi mengenai industri *diving* lebih terstruktur dan informatif. Dan juga dari metode identifikasi entitas data dengan *website* sebagai sumber data dapat digunakan oleh pengembang sistem dari segala bidang.

1.2 Rumusan masalah

Berdasarkan uraian latar belakang, maka rumusan permasalahan yang menjadi fokus dan akan diselesaikan dalam tugas akhir ini antara lain:

1. Bagaimana metode yang dapat dilakukan dalam identifikasi entitas data dengan menggunakan *website* sebagai sumber data?
2. Apa saja *website* mewakili industri *Diving* yang dijadikan sebagai sumber data?
3. Apa saja kata kunci yang paling sering muncul pada tiap *website* yang dijadikan sebagai sumber data?
4. Bagaimana entitas berserta atribut dari industri *Diving* yang dirancang melalui halaman – halaman *website* yang menjadi kajian?
5. Bagaimana tingkat validasi dari entitas data yang sudah dirancang?

1.3 Batasan masalah

Batasan masalah yang dapat diperhatikan dalam penyelesaian Tugas Akhir ini adalah sebagai berikut :

1. Sumber data yang digunakan dalam penelitian ini berasal dari *website* yang bekaitan dengan industri *diving*
2. Untuk pencarian *website* yang dijadikan sebagai sumber data, lokasi yang digunakan adalah Asia (Philipina, Malaysia, Indonesia, Australia).
3. Keyword pencarian yang dilakukan melalui Google adalah *Diving* dan *Snorkling.*
4. Untuk pemilihan *website* yang dijadikan sebagai sumber data, dipilih berdasarkan 10 teratas hasil pencarian Google.
5. Tidak menggunakan kebutuhan dari pengguna atau praktisi.

1.4 Tujuan

Tujuan dari penelitian Tugas Akhir ini adalah

1. Membuat model data sesuai dengan informasi pada *website* – *website* di industri *diving.*

2. Melakukan validasi terhadap model data yang sudah penulis buat pada penelitian ini.
3. Membuat metode pembuatan model data data *website* sebagai sumber data.

1.5 Manfaat

Manfaat yang diharapkan dapat diperoleh dari tugas akhir ini adalah:

Bagi Developer Sistem Informasi pada Industri *Diving*

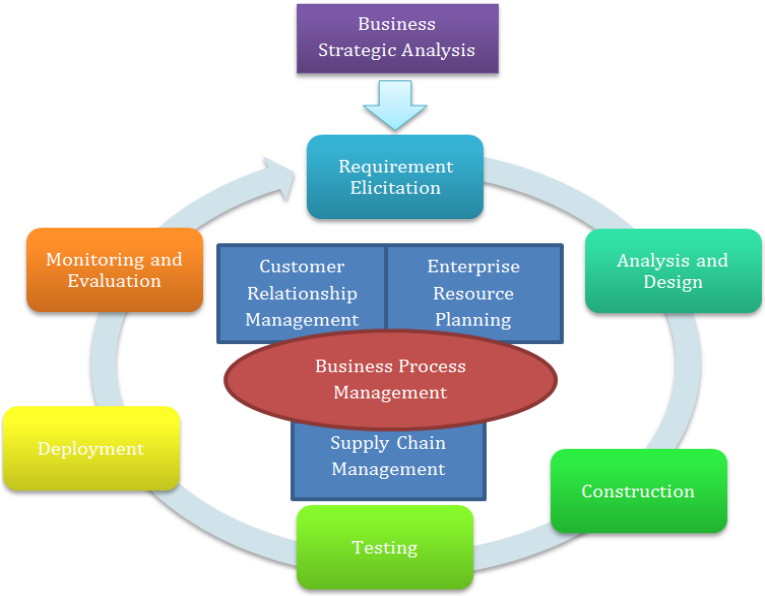
Mendapatkan gambaran mengenai struktur sistem informasi yang harus dikembangkan dan manfaatnya bagi *client* (pelaku industry *diving* yang membutuhkan sistem informasi.

Bagi Developer Sistem Informasi secara umum

Mendapatkan gambaran mengenai tahap – tahap dalam idenfikasi entitas data dengan *website* sebagai sumber data.

1.6 Relevansi

Laboratorium Sistem *Enterprise* (SE) Jurusan Sistem Informasi ITS memliki empat topik utama yaitu *customer relationship management (*CRM), enterprise resource planning (ERP), *supply chain management* (SCM) dan *business process management* (BPM) seperti yang terdapat pada gambar 1.1. Dalam tugas akhir yang dikerjakan oleh penulis mengambil *customer relationship management* (CRM) sebagai topik utama. Dan selain itu, topik penelitian ini mendukung penelitian dosen pembimbing pertama.



Gambar 1.1 Kerangka Kerja Riset di Laboratorium Sistem Enterprise

(Halaman ini sengaja dikosongkan)

BAB II
TINJAUAN PUSTAKA

Bab ini akan menjelaskan mengenai penelitian sebelumnya dan dasar teori yang dijadikan acuan atau landasan dalam pengerjaan tugas akhir ini. Landasan teori akan memberikan gambaran secara umum dari landasan penjabaran tugas akhir.

2.1 Penelitian terkait

Selama pengerjaan tugas akhir ini, terdapat beberapa penelitian yang telah dilakukan sebelumnya yang dapat dijadikan sebagai bahan kajian maupun referensi untuk studi literatur. Penelitian tersebut lalu dikaji untuk dilihat dari gambaran umum, tujuan hasil, dan keterkaitannya dengan penelitian tugas akhir ini. Hasil dari kajian tersebut dapat kita lihat pada tabel berikut ini

Tabel 2.1 Penelitian Sebelumnya 1

Judul	starER: A Conceptual Model for Data Warehouse Design [4]
Nama, Tahun	David W. Embley, Douglas M. Campbell, Yuan Jiang, Stephen W. Liddle, Yiu-Kai Ng
Gambaran umum penelitian	Pada penelitian ini, membuat sebuah model *data warehouse* yang berdasarkan kebutuhan pelanggan dan membuat model yang bebas dari keselahan, mudah dipahami, dan mudah untuk memperluas model data yang dibuat. Peneliatan ini akan membuat sebuah *Entity Relationship Model* yang dilakukan berdasarkan star skema maka akan dihasilkan starER.

Keterkaitan penelitian	Keterkaitan penelitian ini adalah membuat sebuah *conceptual model* dengan menggunakan *data warehouse* yang ada. Pada pembuatan conceptual model akan ditentukan entitas, *relationship*, dan atribut yang didapatkan dari *data warehouse* yang dijadikan sebagai kajian.

Tabel 2.2 Penelitian Sebelumnya 2

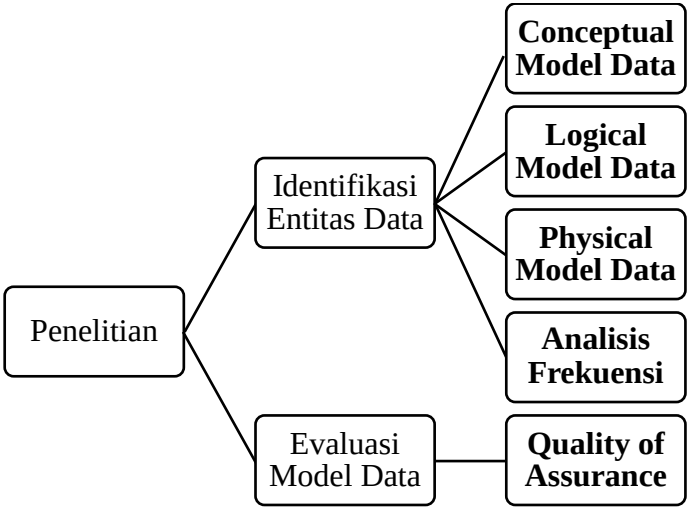
Judul	Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice [5]
Nama, Tahun	Daniel L Moody
Gambaran umum penelitian	Pada penelitian ini menejelaskan mengenai evaluasi terhadap kualitas model data . Penilaian dilakukan dengan melihat dari beberapa matriks / kategori yang sudah ditentukan.
Keterkaitan penelitian	Penelitian ini menggunakan beberapa faktor kualitas dalam melakukan penilaian model data, yaitu ketepatan (*correctness*), kelengkapan (*completeness*), integritas (*integrity*), fleksibleitas (*flexibelity*), kemudahan untuk dipahami (*understandability*), kesederhanaan (*simplicity*), integrasi (*integration*), implementasi *(implementability*)

Tabel 2.3 Penelitian Sebelumnya 3

Judul	Analisis Metadata Halaman Productdisplay Pada Pasar Online Di Indonesia [6]
Nama, Tahun	Junior Wijaya
Gambaran umum penelitian	Pada penelitian ini berfokus kepada pembahasan terkait metadata dari beberapa pasar online. Dari setiap pasar online kemungkinan memiliki metadata yang digunakan oleh pasar – pasar online di Indonesia besera atribut – atribut yang terkait dengan masing – masing metadata.
Keterkaitan penelitian	Kertakaitan penelitian ini adalah menggunakan standarisasi metadata dalam menggunakan penamaan pada suatu metadata. Standarisasi yang digunakan adalah *schema.org.* Penggunaan dari *schema.org* dapat membantuk *search engine* dalam menafsirkan suatu informasi dari suatu *web*, dan mempermudah pengguna dalam mencari informasi yang digunakan.

2.2 Dasar teori

Bagian ini membahas teori dan konsep yang berkaitan dengan penelitian tugas akhir. Relevansi dan dasar teori yang dibahas pada penelitian ini dapat dilihat pada Gambar 1:



Gambar 2.1 Bagan Dasar Teori

Landasan teori disusun untuk menjawab pertanyaan berikut:

1. Hal apa saja yang harus dilakukan dalam Idenftifikasi Entitas Data?
2. Dalam pembuatan metadata, standarisasi apa saja yang akan digunakan?
3. Bagaimana cara menghasilkan form dari model data yang sudah dibuat?
4. Bagaimana cara untuk melakukan validasi pada data model?

2.2.1 *Conceptual Data Model*

Conceptual Data Model merupakan bagian dari aktivitas yang akan dilakukan pada *Conceptual Database Design*, yang merupakan proses untuk melakukan perancangan model pada data yang ada pada perusahaan / atau pun pada suatu sistem yang akan dibangun. Pada *Conceptual Data Model* mengandung beberapa data, yaitu : entitas, *relationship*, atribut, *keys*, dan batasan (*constraint*) [7]. Pada penelitian kali ini, akan dilakukan pembuatan *Conceptual Data Model* dengan

menghasilkan sebuah dokumentasi yang berbentuk *Entity Relationship Diagram* dan *Data Dictionary* yang dihasilkan dari data perusahaan yang sudah didapatkan. Tahapan yang dapat dilakukan pada *Conceptual Data Model* adalah [7]:

1. Identifikasi entitas
2. Identifikasi *relationship*
3. Identifikasi atribut yang terkait dengan entitas
4. Menetapkan domain atribut
5. Menetapkan *primary key*, *alternate key*, dan *candidate key* pada tiap entitas

2.2.2 *Logical Data Model*

Logical Data Model merupakan bagian dari aktivitas yang akan dilakukan pada *Logical Database Design*, yang merupakan mengubah hasil dari *conceptual* menjadi sebuah model data *logical* dengan menghasilkan sebuah hubungan antara entitas, *relationship*¸dan atribut [7]. Dan dalam pembuatan *Logical Data Model* akan dilakukan validasi terhadap model untuk melihat kebenaran dari struktur yang sudah dibuat dan telah mendukung transaksi yang seharusnya dapat dilakukan oleh sistem. Tahapan yang dapat dilakukan pada *Logical Data Model* adalah [7]:

1. Mendapatkan relasi antar Atribut
2. Memvalidasi relasi menggunakan normalisasi
3. Memvalidasi relasi terhadap transaksi pengguna
4. Melakukan cek integritas pada batasan

2.2.3 *Physical Data Model*

Physical Data model merupakan pembuatan model yang menjelaskan mengenai bentuk fisik dari tiap atribut dimana hasil dari *Logical Data Model* akan diubah secara spesifik agar dapat di implementasikan pada suatu DBMS [7]. Pada *Physical Data Model*, akan diidentifikasi batasan (*constraint*) dari tiap atribut sesuai dengan DBMS dan atribut yang dapat memiliki nilai *NULL* atauapun *NOT NULL*.

2.2.4 *schema.org*

Schema.org merupakan sebuah kolaborasi, aktivitas komunitas dengan tujuan untuk membuat, memelihara, dan mempromosikan skema dari sturktrur data pada internet, *web*page, dan pesan email [8]. *Schema.org* menyediakan standarisasi pada pembuatan metadata pada sebuah sistem ataupun halaman *website*. Penggunaan standarisasi tersebut berguna untuk meningkatkan kemudahaan dalam melakukan pencarian terhadap konten yang ada pada *website* pada *search engine* yang dilakukan oleh pengguna [9]. Dalam penelitian kali ini, akan menggunakan standarisasi yang sudah disediakan oleh *schema.org* dalam pembuatan model data dan metadata.

2.2.5 Analisis Frekuensi

Analisis Frekuensi merupakan suatu analisis yang melihat seberapa banyak kata kunci muncul pada beberapa halaman website yang berkaitan dengan industri diving [10]. Analisis frekuensi merupakan bagian dari statistik deskriptif. Statistik deskriptif digunakan untuk menampilkan kesimpulan dari suatu data yang ada. Setelah mengetahui seberapa banyak kata kunci yang terlibat pada beberapa halaman website, akan diketahui beberapa temuan kesimpulan yang dapat membantu dalam melakukan analisis [11]. Kesimpulan tersebut berupa total dari kesulurahan data, data dengan frekuensi tertinggi, dan persentase perbandingan antara frekuensi suatu kata kunci dengan total keseluruhan kata kunci.

2.2.6 *Slovin*

Slovin merupakan sebuah formula matematika untuk menuntukan sebuah ukuran sampel dari suatu data yang akan diuji [12]. Rumus yang digunakan adalah sebagai berikut.

$$n = \frac{N}{1 + Ne^2}$$

n = Ukurang sampel

N = Jumlah data

e = Toleransi eror

Untuk nilai dari toleransi eror yang digunakan adalah sebesar 0.05 [13]. Formula slovin ini digunakan saat penggunaan data yang akan dianalisa menjadi model data.

2.2.7 *Quality of Assurance*

Quality of Assurance merupakan suatu aktivitas untuk mengulas kembali hasil dari perancangan yang sudah di bangun dan akan diukur kualitas dari model data yang dihasilkan [14]. Pada penelitian kali ini, penilian kualitas dengan menggunakan *Quality of Assurance*, berfokus pada penilaian seberapa baik model data yang sudah dibuat pada penilitian. Dan kategori yang digunakan sebagai penilaian, antara lain [15]:

Tabel 2.4 *Quality of Assurance* *[15]*

No.	Metriks	Tingkat Kepentingan	Y/N	Keterangan
1	Dapatkah Cakupan dari entitas yang akan dibangun terdefinisikan.	Esensial		Data model yang dirancang harus mempunyai keselarasan yang jelas dengan bisnis dan pengguna
2	Dapatkah kebutuhan pengguna terdefinisikan	Esensial		Kebutuhan pengguna harus bisa didefinisikan.
3	Apakah data model memenuhi kebutuhan?	Esensial		Data model harus dapat memenuhi kebutuhan
4	Dapat kah business rules yang lengkap terdefinisikan?	Esensial		Sangat penting business rules yang didapat adalah sah.
5	Apakah terdapat daftar istilah yang lengkap?	Esensial		Sangat penting istilah yang didapat adalah sah

6	Pengecekan normalisasi 1. Dapatkah nilai tiap data item pada tabel hanya berasal dari primary key? (Y) 2. Dapatkah data item apapun berasal dari item lain? (N) 3. Apakah nama kolom terdapat perulangan didalam tabel yang sama? (N) 4. Apakah item yang sama muncul didalam lebih dari satu tabel? (N)	*Desirable*		Model data masih tergolong baik meskipun semua pengecekan gagal.
7	Apakah model data relevan dengan Design Patterns?	*Desirable*		Model data masih tergolong baik jika design patterns tidak bisa di identifikasi.
8	Apakah dilakuakn penyesuaian dengan menggunakan standarisasi?	*Desirable*		Tidak terlalu esensial
9	Apakah model mudah untuk dibaca dan dimengerti?	*Desirable*		
10	Dapatkah grup data yang berulang teridentifikasi?	Tidak bersifat kritis		Tidak bersifat kritis. Biasanya indikasi dari tahap normalisasi.dapat diterima untuk sebuah desain *database*.

				Contoh seperti, Address Lines seriing berulang.
11	-Dapatkah data asal dapat teridentifikasi	Tidak bersifat kritis		Tidak bersifat kritis
12	Apakah model data menggunakan penamaan yang standar?	Tidak bersifat kritis		

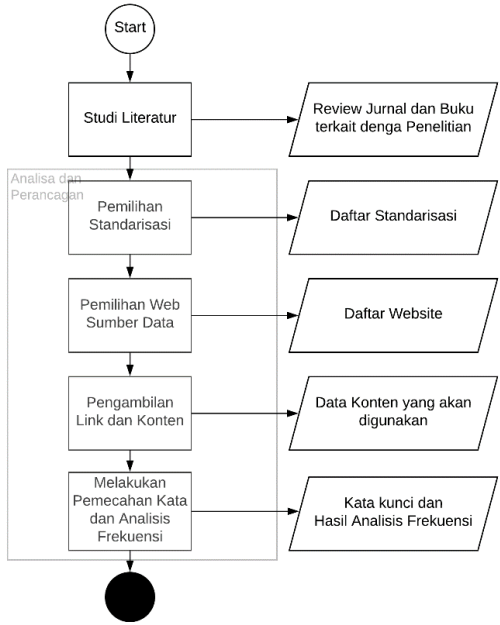
(Halaman sengaja dikosongkan)

BAB III
METODOLOGI PENELITIAN

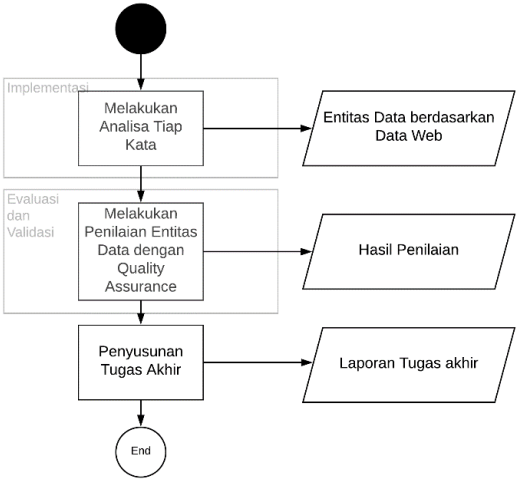
Pada bab metode penelitian akan dijelaskan mengenai tahapan – tahapan apa saja yang dilakukan dalam pengerjaan tugas akhir ini beserta deskripsi dan penjelasan tiap tahapan tersebut. Lalu disertakan jadwal pengerjaan tiap tahapanan.

3.1 Diagram Metodologi

Pada gambar dibawah ini, merupakan alur metodologi yang akan digunakan sebagai panduan untuk menyelesaikan penelitian tugas akhir ini.



Gambar 3.1 Diagram Metodologi Penelitian 1



Gambar 3.2 Diagram Metodologi Penelitian 2

3.2 Uraian metodologi

Pada bagian ini dijelaskan tahapan-tahapan pembuatan tugas akhir secara rinci sesuai dengan metodologi yang telah digambarkan sebelumnya.

3.2.1 Studi Literatur

Pada tahap awal ini, akan dilakukan pengkajian studi literatur yang bertujuan untuk mencari informasi mengenai solusi yang bisa dilakukan terkait dengan permasalahan yang ada dan juga dapat memahami dasar teori yang terkait dengan penelitian ini. Pengumpulan data dan informasi sendiri dilakukan dengan cara membaca beberapa referensi dari buku dan penelitan – penelitian terkait yang telah dilakukan sebelumnya.

3.2.2 Analisis Dan Perancangan

Setelah dilakukannya studi literatur, akan dilakukan pengumpulan data yang dilakukan dengan cara melakukan pencarian *website* kajian. pencarian *website* kajian dilakukan dengan cara melakukan pencarian menggunakan *search engine*

Google dengan menggunakan kata kunci : *diving* dan Snorkling, yang ada di negara : Philipina, Malaysia, Indonesia, Australia. Dari hasil pencarian tersebut akan diambil beberapa *website* yang masuk dalam 10 pencarian teratas di Google dengan menggunakan beberapa keyword yang telah ditentukan sebelumnya.

Setelah didapatkan nya *website* dan halaman halaman yang akan dikaji, akan dilakukan tahapan ekstraksi data terhadap informasi informasi yang ada pada halaman *website* tersebut. Tiap halaman – halaman tersebut akan diambil konten yang adayang dapat digunakan dalam identifikasi entitas data. Setelah itu akan dilakukan pemecahan konten menjadi beberapa kata kunci. Kata kunci tersebut akan dibagi berdasarkan suku kata.

3.2.3 Idenftifikasi Entitas Data

Pada tahap Idenftifikasi Entitas Data, akan dilakukan Idenftifikasi Entitas Data berdasarkan hasil dari ekstraksi informasi dari *website* yang dijadikan sebagai kajian. Untuk membantu dalam Idenftifikasi Entitas Data, dibutuhkan kebutuhan pelanggan terhadap sistem. Kebutuhan pelanggan didapatkan dari praktisi yang sudah ahli dalam bidang *diving*. Dalam Idenftifikasi Entitas Data, terdapat beberapa poin penting yang akan diambil, yaitu Entitas, Atribut, dan Data. Tujuan dari tahap ini adalah untuk melihat struktur data dari tiap halaman *website* yang menjadi kajian dan informasi yang ada pada *website* tersebut..

3.2.4 Evaluasi dan Validasi Model Data

Untuk melihat seberapa baik model data yang telah dibuat, dilakukan penilaian model data berdasarkan *Quality of Assurance*. Hasil dari penghitungan validasi yang telah dilakukan, dapat dimasukkan kedalam penilaian *Quality of Assurance*. Terdapat beberapa poin penting yang dapat dijadikan penilaian model data yang harus dijawab dengan Ya(*yes)*/Tidak(*no*). Dan dalam melakukan penilaian pada poin yang mempunyai kepentingan esensial, nilai yang harus

didapatkan adalah Ya (*yes)*, jika tidak model data yang dirancang dinyatakan tidak valid.

3.2.5 Penyusunan Buku Tugas Akhir

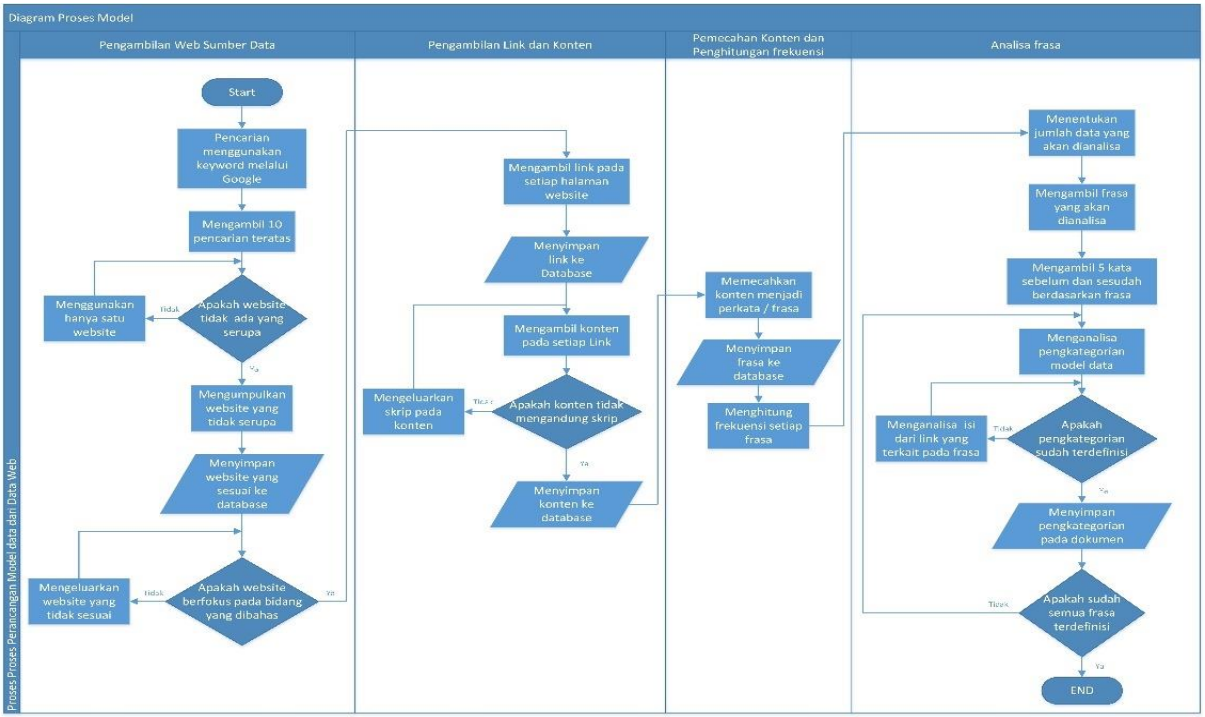
Setelah semua tahap telah dilakukan mengenai Idenftifikasi Entitas Data pada industri *diving* dan telah didapatkan hasil validasi dari model data yang dibuat, akan dilakakuan proses penyusunan laporan tugas akhir. Laporan tugas akhir akan berisikan penjelasan mengenai penelitian, langkah – langkah pengerjaan penelitian, hasil dan kesimpulan dari penelitian secara mendetil.

BAB IV
ANALISA DAN PERANCANGAN

Pada bab ini dijelaskan perancangan awal yang diperlukan sebelum melakukan penelitian tugas akhir. Bab ini mencakup metode-metode dalam persiapan penelitian tugas akhir disertai penjelasannya

4.1 Model Proses Idenftifikasi Entitas Data dari Data Webisite

Idenftifikasi Entitas Data dilakukan untuk membantu dalam pembuatan suatu sistem. Model data dirancang melalui kebutuhan – kebutuhan yang diperlukan oleh orang – orang yang terlibat langsung dalam sistem tersebut. Situs atau *website* dapat dijadikan sebagai sumber data. Pada penelitian kali ini, dilakukan proses – proses dalam Idenftifikasi Entitas Data, dari data *web* hingga menjadi serangkaian entitas. Dalam penelitian ini, dilakukan metode sebagai berikut.



Gambar4.4.1 Model Proses Perancangan Entitas

4.2 Pemilihan Standarisasi

Standarisasi metadata digunakan untuk menstandarkan penggunaan metadata pada suatu sistem yang bergunakan untuk mengoptimalkan dari penggunaan metadata pada sistem. Saat ini, sudah banyak standarisasi yang tersedia dan dapat digunakan sesuai dengan cakupan sistem yang akan dibangun. Penggunaan standarisasi metadata pada penilitan kali ini digunakan untuk pembuatan tiap entitas dan juga hubungan tiap entitas yang ada. Maka dari itu, setelah dilakukannya analisis terhadap standarisasi metadata yang ada, dalam penelitian kali ini akan digunakan, yaitu *schema.org*

Pada standar *schema.org* terdapat elemen elemen yang dapat digunakan dalam penggunaan metadata. Pengguna dapat menggunakan elemen elemen tersebut sesuai dengan fungsi atau cakupan dari sistem ataupun *website* yang dibangun, *schema.org* telah menyediakan pengkategorian sesuai dengan keperluan dan cakupan dari jenis data. Dalam penelitian kali ini, *schema.org* berfungsi untuk memberikan penamaan entitas yang sesuai dan juga atribut atribut apa saja yang harus ada pada entitas tersebut.

4.3 Pengambilan Sumber Data

Untuk melakukan Idenftifikasi Entitas Data pada industri *diving*, dibutuhkan sumber kajian sebagai data informasi yang dapat berguna dalam Idenftifikasi Entitas Data. Sumber kajian dari Idenftifikasi Entitas Data ini menggunakan *website* yang berkaitan dengan industri *diving.*

4.3.1 Pemilihan *Web*

Dalam pemilihan *website* yang akan digunakan, penulis akan memilih berdasarkan hasil dari 10 teratas pencarian *Google* dengan menggunakan kata kunci *diving* dan *snorkling.* Dan *website* yang didapatkan adalah sebagai berikut.

Tabel 4.1 Daftar Website Sumber Data 1

No.	Nama Website
1	www.padi.com
2	www.divezone.net
3	www.diveadvisor.com
4	www.scubatravel.co.uk
5	www.islandsofthephilippines.com
6	www.scubadiving.com
7	www.madmonkeyhostels.com
8	www.divein.com
9	www.diveworldwide.com
10	www.island-divers-caluya.com
11	www.dive-the-world.com
12	www.dive-indonesia.co.id
13	www.allindonesiatravel.com
14	www.divetheworldindonesia.com/
15	www.liveaboard.com
16	www.lumbalumbadiving.com/
17	www.blueseadivers.com
18	www.divinglembongan.com/
19	www.wonderlfulmalaysia.com
20	www.tripadvisor.com
21	www.diveadventures.com.au
22	www.asiadivingvacation.com
23	www.divetheworldmalaysia.com

24	www.luxuryescapes.com
25	www.dive-malaysia.com
26	www.diving.asn.au
27	www.divingcairns.com.au
28	www.diveinaustralia.com.au
29	www.prodivecairns.com.au
30	www.corporate.olympics.com.au
31	www.yongaladive.com.au
32	www.outoftownblog.com
33	www.lonelyplanet.com
34	www.tropicalsnorkeling.com
35	www.lazada.com.ph
36	www.edition.cnn.com
37	www.8list.ph
38	www.action-philippines.com
39	www.aurora.ph
40	www.coraltraingleadventures.com
41	www.snorkeling-report.com
42	www.theamazingindonesia.com
43	www.gonomad.com
44	www.indonesia.travel
45	www.travelfish.org
46	www.symbiosis-travel.com
47	www.tropicalsnorkeling.com

48	www.islands.com
49	www.wildernesstravel.com
50	www.traveltips.usatoday.com
51	www.touropia.com
52	www.1stopmalaysia.net
53	www.thetaaras.com
54	www.asiancorrespondent.com
55	www.thrillophilia.com
56	www.huffingtonpost.com
57	www.lazada.com.my
58	www.australiangeographic.com.au
59	www.westernaustralia.com
60	www.cntraveler.com
61	www.visitnsw.com
62	www.tusadive.com
63	www.sydney.com
64	www.snorkeltrips.com.au
65	www.divethereef.com
66	www.hamiltonisland.com.au

Dari 66 *website* yang didapatkan berdasarkan 10 pencarian teratas dengan menggunakan mesin pencari *Google*, akan dipilih beberapa *website* yang memiliki topik utama mengenai industri *diving*. Maka dari itu, didapatkan 25 *website* yang akan menjadi bahan penelitian.

Tabel 4.2 Daftar Website Sumber Data 2

No.	Nama Website
1	www.padi.com
2	www.divezone.net
3	www.diveadvisor.com
4	www.scubatravel.co.uk
5	www.scubadiving.com
6	www.divein.com
7	www.diveworldwide.com
8	www.dive-the-world.com
9	www.dive-indonesia.co.id
10	www.divetheworldindonesia.com
11	www.lumbalumbadiving.com
12	www.blueseadivers.com
13	www.divinglembongan.com
14	www.diveadventures.com.au
15	www.asiadivingvacation.com
16	www.divetheworldmalaysia.com
17	www.dive-malaysia.com
18	www.divingcairns.com.au
19	www.prodivecairns.com.au
20	www.yongaladive.com.au
21	www.tropicalsnorkeling.com
22	www.coraltraingleadventures.com
23	www.tropicalsnorkeling.com

24	www.tusadive.com
25	www.snorkeltrips.com.au
26	www.island-divers-caluya.com
27	www.divethereef.com

4.3.2 Pengambilan *Link* dan Konten pada *Web*

Dan setelah mendapatkan *website* yang menjadi sumber data dari Idenftifikasi Entitas Data, akan dilakukan pengambilan – yang ada pada masing masing *website* yang sudah dipilih. Pada *website* tersebut, tidak hanya yang akan diambil, tetapi konten dari tiap halaman / juga akan diambil. Pengambilan dan konten pada *web* tersebut akan di simpan pada *database* yang diberi nama '*dive*industry'.

Didalam *database* tersebut terdapat 2 tabel yang menjelaskan masing masing *web* dan juga atribut / kolom, yaitu :

Tabel 4.3 Struktur Tabel *Website* dan *link_web*

Nama Tabel	Nama Atribut	Tipe Data	Keterangan
Website	Id_*web* (PK)	Int (12)	Kode unik tiap *website*
	nama_*website*	Varchar (255)	Nama *website* yang menjadi sumber data
	url	Varchar (1000)	URL / *Link* dari tiap *website*.
Link_web	Id_*link* (PK)	Int(12)	Kode unik tiap dari *website*
	Id_*web*	Int(12)	Kode unik tiap *website*

	Link	Varchar (1000)	Berisi *link* yang ada pada tiap *website*
	Flag	Tinyint (1)	Sebagai penanda jika *link* tersebut telah ditelurusi. Dan nilai yang dipunyai adalah 0 jika belum ditelusuri, dan 1 jika sudah ditelusuri.
	Content	text	Berikan konten dari tiap halaman atau dari *website.*

Metode yang digunakan dalam melakukan pengambilan adakah metode *crawling.* Penggunaan metode *crawling* dilakukan dengan pengkodingan PHP.

```
function nyimpen( $url, $id, $n=0 ) {
   global $db, $c; //manggil db
   $domain1 = explode( '/', $url )[2]; //pemecah
   $c->bc = $c->get( $url );

      if ($url) $db->u( 'link_web', 'SET
      Content=? WHERE link=?', array(
      strip_tags( $c->bc ),$url  ) );
      //pergantian flag menjadi 1

      // Get text
      for ( $x=1; $x<=200;$x++){
         $link = $c->xp( 'href="', '"', $x );
         $opt = explode( '/', $link ); //pemecah

         if ( $opt[0] != 'http:' && $opt[0] !=
         'https:' ) {
```

Dari hasil pengambilan pada *website*, *link* tersebut akan disimpan pada *database.* Penyimanan *link* akan disimpan pada tabel *_web* dan kolom .

```
$f = $db->f( 'link_web', 'link', 'WHERE link=?',
$link );

if ( ! $f ) $db->i('link_web',
'id_web,link,flag', array( $id, $link, 0 ) );
//menyimpan link
```

Konten yang diambil berupa *text* atau tulisan murni. Pengambilan konten dari *web* tersebut juga akan disimpan pada *database* dan tabel yang sama.

```
$html=preg_replace('/<script\b[^>]*>(.*?)<\/scri
pt>/is', "", $c->bc); //mereplace content script

$html=preg_replace('/<style\b[^>]*>(.*?)<\/style
>/is', "", $html); //mereplace content css

if ($url) $db->u( 'link_web', 'SET
flag=1,Content=? WHERE link=?', array(
strip_tags( $html ),$url ) ); //pergantian flag
menjadi 1 dan pengambilan konten

$f = $db->f( 'link_web', '*', 'WHERE flag=0' );
```

4.3.3 Melakukan Analisa *Text Frequency*

Setelah didapatkannya seluruh konten yang ada pada tiap *website*, akan dilakukan penghitungan frekuensi kata atau frase dari seluruh konten yang sudah didapatkan. Kata – kata yang sudah di pecah akan disimpan pada *database* "*Dive*industry". Pada *database* tersebut akan ditambakan tabel untuk tiap kata konten dan pada tabel "*Link_web*" akan ditambahakan "flag_konten" sebagai penanda.

Tabel 4.4 Tabel *link_web* dan *link_keyword*

Nama Tabel	Nama Atribut	Tipe Data	Keterangan
Link_web	Flag_konten1	Tinyint (1)	Sebagai penanda jika *link* tersebut telah

			ditelurusi untuk penguraian 1 kata. Dan nilai yang dipunyai adalah 0 jika belum ditelusuri, dan 1 jika sudah ditelusuri.
	Flag_konten2	Tinyint (1)	Sebagai penanda jika *link* tersebut telah ditelurusi untuk penguraian 2 kata. Dan nilai yang dipunyai adalah 0 jika belum ditelusuri, dan 1 jika sudah ditelusuri.
	Flag_konten3	Tinyint (1)	Sebagai penanda jika *link* tersebut telah ditelurusi untuk penguraian 3 kata. Dan nilai yang dipunyai adalah 0 jika belum ditelusuri, dan 1 jika sudah ditelusuri.
	Flag_konten4	Tinyint (1)	Sebagai penanda jika *link* tersebut telah ditelurusi untuk penguraian 4 kata. Dan nilai yang dipunyai adalah 0 jika belum ditelusuri, dan 1 jika sudah ditelusuri.

	Flag_konten5	Tinyint (1)	Sebagai penanda jika *link* tersebut telah ditelurusi untuk penguraian 5 kata. Dan nilai yang dipunyai adalah 0 jika belum ditelusuri, dan 1 jika sudah ditelusuri.
*Link*_ keyword	Id_key (FK)	Int (12)	Kode unik untuk tiap kata
	Keyword	Varchar (1000)	Berisi kata yang didapatkan dari konten
	frekuensi	Int (12)	Berisi jumlah dari tiap kata yang muncul pada tiap konten.
*Link*_ keyword2	Id_key2 (FK)	Int (12)	Kode unik untuk tiap kata
	Keyword	Varchar (1000)	Berisi kata yang didapatkan dari konten
	Frekuensi	Int (12)	Berisi jumlah dari tiap kata yang muncul pada tiap konten.
*Link*_ keyword3	Id_key3 (FK)	Int (12)	Kode unik untuk tiap kata
	Keyword	Varchar (1000)	Berisi kata yang didapatkan dari konten

	Frekuensi	Int (12)	Berisi jumlah dari tiap kata yang muncul pada tiap konten.
*Link_* keyword4	Id_key4 (FK)	Int (12)	Kode unik untuk tiap kata
	Keyword	Varchar (1000)	Berisi kata yang didapatkan dari konten
	Frekuensi	Int (12)	Berisi jumlah dari tiap kata yang muncul pada tiap konten.
*Link_* keyword5	Id_key5 (FK)	Int (12)	Kode unik untuk tiap kata
	Keyword	Varchar (1000)	Berisi kata yang didapatkan dari konten
	Frekuensi	Int (12)	Berisi jumlah dari tiap kata yang muncul pada tiap konten.

Konten yang sudah didapat, akan dipecah menjadi per 1 kata, 2 kata, 3 kata, 4 kata, dan 5 kata. Tiap kata yang sama akan ditambahkan pada jumlah frekuensi dari kata tersebut.

```
foreach ( $f as $r ) { //looping, satu satu
   $r['Content'] = str_replace( "\n", ' ',
   $r['Content'] ); //menghilangkan enter

   $r['Content'] = preg_replace('!\s+!', ' ',
   $r['Content'] ); //menghilangkan dobel
   spasimenggunakan requler expression

   $pecah = explode( ' ', $r['Content'] ); //get
   yang mau dipecah

   foreach ( $pecah as $p) {
      $kata = str_replace( $diapus, '', $p );
      //menghilangkan karakter

      $kata = trim( $kata ); //menghilangkan
      enter dan spasi

      if ( ! $kata ) continue; //jika kosng,
      lanjut

      foreach ($stopword as $sw) { //sw satu
      persatu

      if ( $kata == $sw ) continue 2;
         }
```

Pemecahanan tiap kata pada tiap konten, akan disimpan pada *database* yang sama.

```
$cek = $db->f( 'link_keyword', '*', 'WHERE
keyword=?', $kata ); //mengecek

if ( ! $cek ) $db->i('link_keyword',
'keyword,frekuensi', array( $kata, 0 ) );
//menambahkan keyword

else $db->u( 'link_keyword', 'SET
frekuensi=frekuensi+1 WHERE id_key=? ', [$cek-
>id_key] ); //jika sudah ada keyword, tambahkan
frekuensi

$db->u( 'link_web', 'SET flag_konten=1 WHERE
id_link=?', $r['id_link'] );
```

4.3.4 Melakukan Analisa Tiap *Text*

Setelah mendapatkan frasa – frasa tiap konten dari *website* yang ada dan menghitung frekuensi dari keseluruhan, akan diambil frekuensi tertinggi dari tiap frasa. Tiap frasa yang dipilih akan dilakukan analisa untuk mengetahui kategori yang sesuai. Kategori data yang perlu di analisis antara lain, entitas, atribut, dan data. Entitas merupakan sebuah objek yang ada di dunia nyata dan menjadi hal utama dalam mengambarkan sebuah data. Sedangankan atribut adalah represntasi beberapa properti yang menjelaskan suatu entitas. Dan data adalah sebuah informasi yang dapat dimasukkan pada entitas dan atribut yang sesuai. Karakteristik dari entitas dan atribut sebagai berikut [16].

Entitas :

1. *Presence* :mengetahui komponen dari dunia *diving*
2. *Significance* :mengetahui kebutuhan dari dunia *diving*
3. *Visibility* :bisa bersifat yang terhitung maupun yang tak terhitung

Atribut :

1. Menjelaskan entitas yang ada
2. Harus berkaitan dengan entitas yang ada.

Untuk mengetahui frasa yang sudah didapatkan termasuk kedalam kategori yang sesuai, dengan cara mengambil beberapa kata sebelum dan sesudah frasa tersebut pada sumber data. Dengan didapatkannya bantuan beberapa frasa pendukung tersebut, dapat membantu dalam melakukan analisa pengkategorian. Frasa yang akan digunakan dan di analisa adalah frasa yang memiliki frekuensi paling banyak.

```
$f = $db->r( 'link_web', '*'  );

$fk = $db->r( 'link_keyword3', '*', 'where
flag_analisis=0 order by rand()');
   foreach ($fk as $k){

foreach ($f as $x){

   $x['Content'] = str_replace( "\n", ' ',
$x['Content'] ); //menghilangkan enter
   $x['Content'] = preg_replace('!\s+!', ' ',
$x['Content'] ); //menghilangkan dobel
spasimenggunakan requler expression

   $kataLengkap = $x['Content'];
   $kataKunci = $k['keyword'];

   $kataSebelum = null;
```

```
   $kataSesudah = null;

   $kataExplode = explode($kataKunci,
$kataLengkap);

   if(count($kataExplode) >= 2)
   {
         if(!empty($kataExplode[0]))
         {
             $kataSebelum = $kataExplode[0];
         }
         if(!empty($kataExplode[1]))
         {
             $kataSesudah = $kataExplode[1];
         }
   }
```

Pada hasil dari pengambilan kata awal dan terakhir, akan di simpan pada tabel yang berbeda. Tabel tersebut diberi nama ‘analisa’. Tiap analisa per *keyword* akan disediakan satu tabel

‘analisa’. Didalam tabel tersebut terdapat kolom id_analisis, id_key, id_*link*, analisis, dan frekuensi.

Tabel 4.5 Tabel Analisis

Nama Tabel	Nama Atribut	Tipe Data	Keterangan
Analisis	Id_analisis	Int(12)	Kode unik untuk tiap data analisis
	Id_key	Int(12)	Kode unik yang didapatkan dari tabel keyword
	Id_*link*	int(12)	Kode unik yang didapatkan dari tabel *link_web*
	Analisa	Varchar (1000)	Hasil pengambilan kata awal dan kata akhir berdasarkan keyword
	frekuensi	Int(12)	Jumlah data analisa yang sama
Analisis2	Id_analisis2	Int(12)	Kode unik untuk tiap data analisis
	Id_key2	Int(12)	Kode unik yang didapatkan dari tabel keyword
	Id_*link*	int(12)	Kode unik yang didapatkan dari tabel *link_web*
	Analisa2	Varchar (1000)	Hasil pengambilan kata awal dan kata

			akhir berdasarkan keyword
	Frekuensi	Int(12)	Jumlah data analisa yang sama
Analisis3	Id_analisis3	Int(12)	Kode unik untuk tiap data analisis
	Id_key3	Int(12)	Kode unik yang didapatkan dari tabel keyword
	Id_*link*	int(12)	Kode unik yang didapatkan dari tabel *link_web*
	Analisa3	Varchar (1000)	Hasil pengambilan kata awal dan kata akhir berdasarkan keyword
	Frekuensi	Int(12)	Jumlah data analisa yang sama
Analisis4	Id_analisis4	Int(12)	Kode unik untuk tiap data analisis
	Id_key4	Int(12)	Kode unik yang didapatkan dari tabel keyword
	Id_*link*	int(12)	Kode unik yang didapatkan dari tabel *link_web*
	Analisa4	Varchar (1000)	Hasil pengambilan kata awal dan kata akhir berdasarkan keyword

	frekuensi	Int(12)	Jumlah data analisa yang sama
Analisis5	Id_analisis5	Int(12)	Kode unik untuk tiap data analisis
	Id_key5	Int(12)	Kode unik yang didapatkan dari tabel keyword
	Id_*link*	int(12)	Kode unik yang didapatkan dari tabel *link_web*
	Analisa5	Varchar (1000)	Hasil pengambilan kata awal dan kata akhir berdasarkan keyword
	Frekuensi	Int(12)	Jumlah data analisa yang sama

4.3.4.1 Melakukan Analisa melalui Text

Dalam pengkatagorian tiap kata kunci yang akan dianalisa dengan menggunakan beberapa kata sebelum dan kata sesudah nya. Maka dari itu akan terbentuk sebuah potongan dari sebuah kalimat ataupun pada bagian konten dari halaman website tersebut. Untuk secara umum, struktur dari suatu kata adalah :

$$Subjek + Predikat/Kata\ Kerja + Objek + Keterangan$$

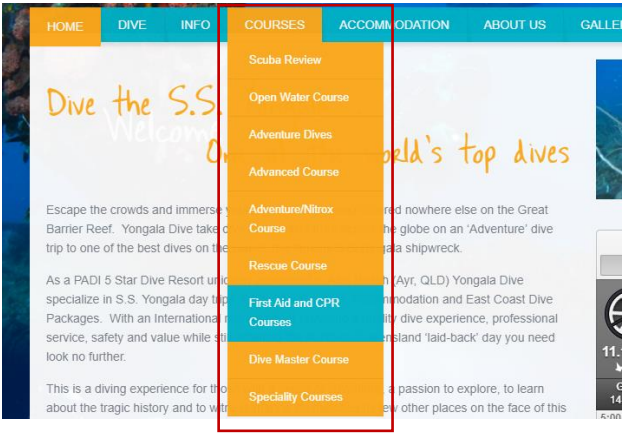
Untuk keterkaitan antara struktur kalimat dan kata kunci dalam menentukan kategori tergantung dengan konteks dari kata kunci tersebut. Didalam sebuah teks atau narasi, entitas secara umum diwakili oleh suatu kata benda / *nouns.* Dan atribut dapat diklarifikasikan sebagai kata sifat yang menjelaskan entitas tersebut [17].

4.3.4.2 Melakukan Analisa melalui Halaman Website

Setelah dilakukannya analisa melaui teks dan kata kunci tersebut tidak dapat dikatagorikan, maka dapat dilakukan dengan cara melakukan pengecakan terhadap *link* halaman website yang mempunyai kata kunci tersebut. *Link* tersebut bisa didapatkan dari dari program analisa melalui teks yang sebelumnya sudah dilakukan. Pada program tersebut akan menyimpan *id_link* dari halaman *website* yang dituju. Sehingga dapat dilakukan pengecekan pada berdasarkan halaman *website* yang mengandung kata kunci tersebut.

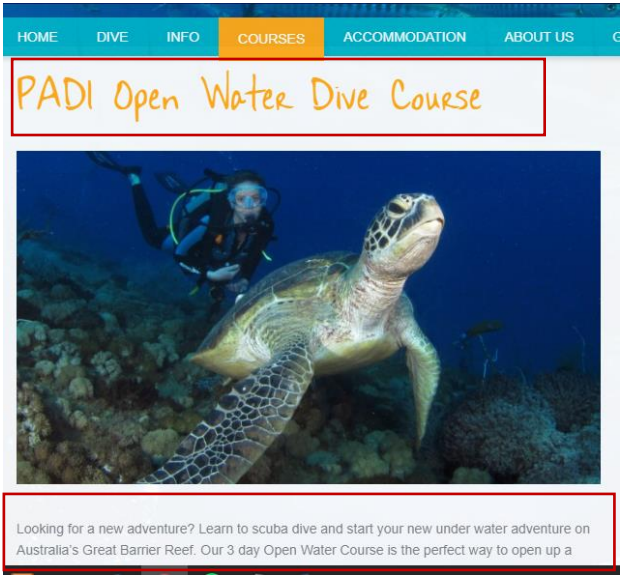
Dalam melakukan analisa kata kunci pada halaman *website* dapat dilakukan dengan cara melihat letak dari kata kunci tersebut berada. Melalui letak dari kata kunci tersebut, dapat diketahui kategori yang tepat untuk kata kunci yang di analisa. Letak dari kata kunci mempunyai pendefinisian dan pengkatagorian yang berbeda.

Yang pertama adalah letak kata kunci yang berada pada *menu bar*. *Menu bar* biasa nya terletak pada bagian atas dari halaman *webisite*. Untuk letak kata kunci yang berada pada *menu bar* dari halaman *website* tersebut, dapat dikatagorikan sebagai entitas yang ada pada industri *diving*. Hal tersebut dapat dikatakan karena menu tersebut menghimpun beberapa data - data yang berkaitan. Dan untuk bagian *drop list* atau beberapa daftar yang muncul pada menu, dapat dikatagorikan sebagai data dari atribut. Hal tersebut dapat dikatakan karena *drop list* tersebut terdapat sebuah data yang berkaitan dengan menu yang ada diatasnya. Dan dapat disebut juga atribut nama dari entitas menu tersebut.



Gambar 4.2 Analisa melalui Halaman Webisite 1

Selanjutnya adalah pada bagian isi dari dari suatu halaman *website*. Isi biasanya terletak pada bagian tengah dari halaman *website*. Jika suatu kata kunci berada diantar kalimat isi dari suatu halaman website, kata kunci tersebut dapat dikatagorikan sebagai data dari suatu atribut dan isi dari halaman tersebut merupakan sebuah data. Untuk menentukan atribut yang mengumpulkan data tersebut, dapat dilihat dari judul isi dari halaman *website*. Jika judul tersebut mengandung salah satu data dari entitas, maka data tersebut merupakan atribut yang dinaungi oleh entitas yang berkaitan. Tapi jika judul dari isi halaman *website* tidak ada sama sekali kaitannya dengan salah satu entitas, maka isi tersebut termasuk dalam entitas artikel.

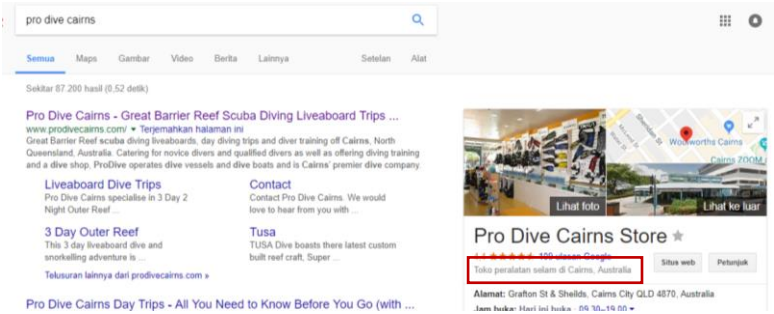


Gambar 4.3 Analisa melalui Halaman Website 2

4.3.4.3 Melakukan Analisa melalui Pencarian Google

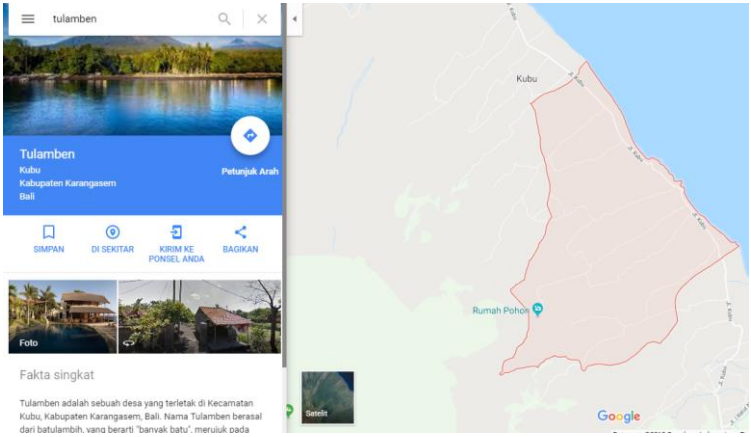
Jika pengkatagorian tidak dapat dilakukan dengan menganalisa halaman *website* yang terkait, maka dapat dilakukan analisa dengan cara melakukan pencarian kata kunci mengunakan mesin pencari *Google*. Terdapat beberapa fitur yang ada pada mesin pencari *Google* dapat dimanfaatkan sebagai pertimbangan dalam melakukan analisa pengkategorian model data terhadap kata kunci yang dipilih.

Kata kunci yang dipilih dapat dicari dengan menggunakan *Google Search*. Dari pencarian *Google* tersebut akan didapatkan sebuah informasi yang sudah di kumpulkan oleh *Google*. Dari informasi tersebut dapat membantu dalam menentukan pengkatagorian sebuah kata kunci yang digunakan. Informasi tersebut dapat memunculkan bahwa kata kunci tersebut merupakan suatu data dari atribut dan entitas yang berkaitan.



Gambar 4.4 Analisa melalui Pencarian *Google* 1

Dan fungsi lainnya yang dapat digunakan untuk menganalisa suatu kata kunci adalah menggunakan *Google Maps.* Untuk penggunaan *Google Maps* ini berkaitan dengan nama suatu tempat. Jika kata kunci yang dianalisa terdapat pada suatu kawasan, maka kata kunci tersebut merupakan sebuah data dari atribut nama suatu tempat.



Gambar 4.5 Analisa melalui Pencarian Google 2

4.3.4.4 Entitas dan Atribut yang didapatkan dari Data

Dalam pengkatagorian suatu kata kunci, terdapat beberapa entitas dan atribut yang belum dapat didefinsikan. Jika suatu

kata kunci teridentifikasi sebagai suatu data, maka data tersebut adalah bagian dari suatu atribut. Jika atribut dari data belum teridentifikasi sebelumnya, maka atribut tersebut akan muncul dan dimasukkan pada atribut pada suatu entitas. Dan juga jika entitas dari atribut belum teridentifikasi, maka entitas tersebut akan muncul dan dimasukkan pada daftar entitas yang ada pada industri *diving.*

4.3.4.5 Entitas dan Atribut yang didapatkan dari *Schema.org*

Dalam melakukan Idenftifikasi Entitas Data berdasarkan data *web* ini, akan mengacu pada suatu standarisasi metadata, yaitu *schema.org*. Penggunaan *schema.org* untuk memaksimalkan penggunaan properti yang tepat dalam Idenftifikasi Entitas Data agar dapat dengan mudah ditemukan oleh pengguna saat melakukan pencarian pada mesin pencari. Tipe metadata yang disediakan oleh *schema.org* akan digunakan jika ada kesesuaian dalam entitas pada model data. Jika ada entitas yang dirancang sesuai dengan tipe metadata *schema.org,* maka properti yang ada pada tipe tersebut akan digunakan sebagai atribut suatu entitas.

BAB V
IMPLEMENTASI ANALISA

Pada bab ini akan dilakukan tahap implementasi dari analisa yang telah dirancang pada bab sebelumnya. Analisa akan dilakuakan pada tiap kata kunci yang telah diproses pada pemecahan kata kunci.

5.1 Kata Kunci yang Digunakan

Dalam perancangan entitas dan atribut, kata kunci yang didapatkan dari konten web akan digunakan sebagai bahan analisa. Untuk jumlah kata kunci yang digunakan sebagai bahan analisa perancagan model data, ditentukan dengan cara metode *slovin*. Metode *slovin* dilakukan dengan mengambil sampel dari jumlah data yang ada. Data yang digunakan adalah kata kunci yang sebelumnya sudah didapatkan. Penggunaan toleransi eror dalam penghitungan *slovin* adalah 0.5

Tabel 5.1 Pemilihan Data Sampel

	Jumlah Data	Jumlah Sample
link_keyword	67901	397
link_keyword2	212808	399
link_keyword3	232470	399
link_keyword4	219600	399
link_keyword5	137778	398

Data sampel yang digunakan sebagai bahan analisa akan menghasilkan pengkatagorian entitas, atribut, dan data. Dan untuk kata kunci yang di pilih adalah 25 kata kunci dengan frekuensi tertinggi dan selebihnya menggunakan kata kunci secara acak.

5.2 Analisa Idenftifikasi Entitas Data

Pada sub bab ini akan dilakukan implementasi analisa indentifikasi entitas data yang sebelumnya sudah di rancang.

5.2.1 Analisa Menggunakan Teks

Metode analisa pertama yang dapa dilakukan adalah analisa menggunakan teks yang berkaitan. Berikut merupakan hasil analisa tiap kata kunci dengan menggunakan teks sebagai bahan analisa.

5.2.1.1 Entitas Diver

Atribut role merupakan sebuah peran yang dilakukan oleh *person* , didapatkan dari kata kunci '*Diver'* yang diidentifaksi sebagai data dari *role*.

Tabel 5.2 Analisa Atribut *role*

Id_key	Id_link	Keyword	frekuensi	Analisis
149	69527	Divers	4884	dive site is reserved for experienced (divers) only. It has even happened
149	7682			due to the small numbers of (divers) visiting the country.Cristian Magnani My
149	4929			is an ideal holiday destination for (divers) with a non-diving partner. Fiji

Dari tabel 5.2, terlihat bahwa *diver* melakukan aktivitas yang ada di industri *diving.* Dan diver dibeberapa kalimat tersebut bertindak sebagai kata benda (*noun*). Maka dari itu, *diver* dapat dijadikan sebagai entitas karena berperan penuh didalam industri *diving* dan menunjukkan kata benda.

5.2.1.2 Atribut *name_dive_operator*

Atribut *name_operator* merupakan nama dari orang yang terlibat dengan industri *diving*, didapatkan dari kata kunci '*dive instructor*' yang mempunyai frekuensi sebanyak 415 kali muncul pada berbagai halaman website.

Tabel 5.3 Analisa Atribut *name_person*

Id_key	Id_link	Keyword	Frekuensi	Analisis
11793	79894	Dive insturctor	415	Fishman says. "As a dive instructor, I work every day to bring

Dari tabel 5.3, Fishan berkata bahwa dia adalah seorang *dive instructor*. *Dive instructor* merupakan peran yang dilakukan oleh seseorang untuk melatih para penyelam [18]. Dan *dive instructor* mempunyai peran di industri *diving*. Dan data 'Fishman' mewakili atribut *name_person* sebagai data.

5.2.1.3 Atribut *type_ destination*

Atribut *type_destination* merupakan tipe dari *destination* yang ada pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Great Barrier Reef.*

Tabel 5.4 Analisa Atribut *type_destination*

Id_key	Id_link	keyword	Frekuensi	Analisis
14202	65055	Great Barrier Reef	2279	Reports. Scuba Diving Conditions on the (Great Barrier Reef) , Australia. Great Advice, toll-free! USA: 1-800-207-2453 Australia:

Dari tabel 5.4, dijelaskan bahwa terdapat tempat dive yaitu *great barrier reef*, sebuah karang atau *reef*. Untuk mengetahui tipe dari nama tempat tersebut dapat diketahui jika nama tempat

tersebut diikuti oleh kata tambahan dibelakangnya. Seperti *Great Barrier Reef*, nama tempat Reef dengan tipe island. *Reef* merupakan kata keterangan tempat. Maka dari itu, *reef* merupakan tipe dari tempat yang bisa diselami oleh penyelam. *reef* dijadikan sebagai data yang mewakili atribut *type_ destination*.

5.2.1.4 Atribut name_promo

Atribut *name_promo* merupakan nama dari *courses* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'summer buddy promo'.* .

Tabel 5.5 Analisa Atribut name_promo

Id_key	Id_link	keyword	Frekuensi	Analisis
55699	5895	Summer buddy promo	47	Season Special" Sams Tours Palau "2017 (Summer Buddy Promo") View All Specials Follow us on Facebook

Dari tabel 5.5, terdapat sebuah promosi di Pulau dengan nama *summer buddy promo*. Kata *promo* menjelaskan 2 kata sebelumnya, dan di identifikasi sebagai nama dari promo tersebut. Maka dari itu, dibentuk lah atribut *name_promo. Summer buddy promo* dijadikan sebagai data dari atribut *name-_promo.*

5.2.1.5 Atribut detail_promo

Atribut *detail_promo* merupakan penjelasan dari *promo* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'4 Nights + '* .

Tabel 5.6 Analisa Atribut detil_promo

Id_key	Id_link	keyword	frekuensi	Analisis
52153	5992	"4 Nights +	33	Buddy Promo" Layang Layang Dive Resort ("4

				Nights +) 1 Night Free" View All Specials Follow

Dari tabel 5.6, terdapat sebuah promosi di Layang layang Dive Resort dengan nama *buddy deal*. Promosi yang didapatkan yaitu *'4Nights + 1Night Free'*. *'4Nights + 1Night Free'* diartikan sebagai penjelasan dari promo *buddy deal*. Maka dari itu, dibentuk lah atribut *detail_promo*. *'4Nights + 1Night Free'* dijadikan sebagai data dari atribut *detail_promo*.

5.2.1.6 Atribut type_animal

Atribut *type_animal* merupakan tipe dari *animal* yang dapat dinikmati para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'fish'*.

Tabel 5.7 Analisa Atribut type_animal

Id_key	Id_link	keyword	frekuensi	Analisis
3999	274	fish	4100	essentially underwater cities, crammed with different (fish) , coral and invertebrate species. They

Dari tabel 5.7, dijelaskan bahwa dibawah air terdapat ikan (*fish*). Ikan merupakan salah satu jenis atau tipe hewan (*animal)*. Maka dari itu, *fish* merupakan tipe dari dari *animal*. *fish* dijadikan sebagai data yang mewakili atribut *type_animal*.

5.2.1.7 Atribut price_resort

Atribut *price_resort* merupakan biaya dari *resort* yang dapat dinikmati para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'(400 Resort Credit'*.

Tabel 5.8 Analisa Atribut price

Id_key	Id_link	Keyword	frekuensi	Analisis

63906	4989	400 Resort Credit	36	Nights Pay 9 Nights Plus SBD $ (400 Resort Credit) per Bungalow View Details Fatboys Resort, Gizo

Dari tabel 5.8, terdapat sebuah harga dengan kurs dolar. Dan kata selanjutnya mengenai *resort*. Dan dari hal itu berarti menjelaskan harga dari suatu *resort*. Maka dari itu, terbentuklah atribut *price*.

5.2.2 Analisa Menggunakan Halaman Website

Berikut merupakan hasil analisa tiap kata kunci melalui halaman website.

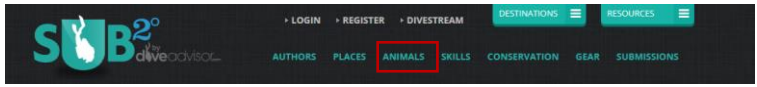
5.2.2.1 Entitas Animal

Entitas animal mendeksripsikan hewan yang hidup di industri *diving* dan para *diver* dapat melihatnya. Entitas *animal* didapatkan dengan analisa terhadap kata kunci '*animal'* dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.9 Analisa Entitas animal

Id_key	Id_link	keywords	frekuensi	Analisis
8262	48219	Animals	507	App SUB2o by DiveAdvisor authors places (animals) skills conservation gear submissions Home

Dari tabel 5.9, terdapat *animal* pada industri *diving*. Banyak hewan yang hidup di linkungan industry *diving*. Untuk membuktikannya, dapat dilihat langsung melalui halaman website yang terkait.



Gambar 5.1 Analisa Entitas Animal

Dari gambar 5.1, terlihat bahwa *animals* terletak pada bagian menu dari suatu *website*. Menu dari suatu *website* menghimpun beberapa data yang penting. Dibutuhkan suatu entitas untuk menghimpun data hewan yang dapat dilihat oleh penyelam. Maka dari itu, *animal* dijadikan sebagai entitas karena merupakan kompenen penting di Industri *diving*.

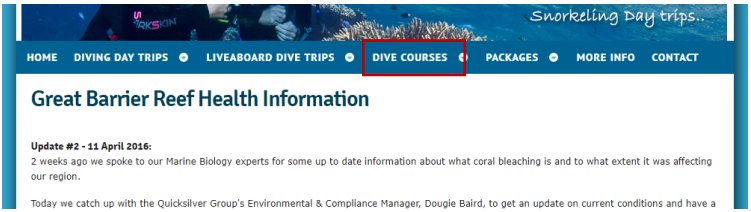
5.2.2.2 Entitas courses

Entitas *courses* mendeksripsikan pelatihan atau ilmu yang bisa didapatkan oleh para *diver* yang diberikan oleh *dive center*. Entitas *courses* didapatkan dengan analisa terhadap kata kunci '*courses'* dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.10 Analisa Entitas course

Id_key	Id_link	keyword	frekuensi	Analisis
56	66725	courses	1369	Ribbon Reefs Coral Sea Liveaboard Dive (Courses) Referral Training Learn to Dive

Dari tabel 5.10, terdapat *courses* pada industri *diving*. *Courses* sendiri mempunyai arti pelatihan. Untuk membuktikannya, dapat dilihat langsung melalui halaman website yang terkait.



Gambar 5.2 Analisis Entitas *courses*

Dari gambar 5.2, terlihat bahwa *courses* terletak pada bagian menu dari suatu *website*. Menu dari suatu *website* menghimpun beberapa data yang penting. Maka dari itu, *course* dijadikan

sebagai entitas karena merupakan kompenen penting di Industri *diving*.

5.2.2.3 Entitas shop

Entitas *store* mendeksripsikan toko peratalan pendukung melakukan aktivitas di industri *diving*. Entitas *shop* didapatkan dengan analisa terhadap kata kunci '*shop'* dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.11 Analisa Entitas *Shopping_center*

Id_key	Id_link	keyword	Frekuensi	Analisis
53	48171	shop	1253	Register or LOGIN List Your Dive (Shop) or Liveaboard Sorry, but we

Dari tabel 5.11, terdapat *shop* pada industri *diving*. *Shop* sendiri mempunyai dalam bahasa Indonesia mempunyai arti tempat berbelanja. Tempat berbelanja penting untuk para penyelam dalam memehuni kebutuhan saat menyelam. Untuk membuktikannya dapat dilihat dari halaman website yang berkaitan.

Gambar 5.3 Analisa Entitas *Shop*

Dari gambar 5.3, terlihat bahwa *shops* terletak pada bagian menu dari suatu *website*. Menu dari suatu *website* menghimpun beberapa data yang penting. Dibutuhkan suatu entitas untuk menghimpun data *shop* yang dapat digunakan oleh penyelam. Maka dari itu, *shop* dijadikan sebagai entitas karena merupakan kompenen penting di Industri *diving*.

5.2.2.4 Entitas Resort

Entitas resort mendeksripsikan tempat penginapan yang bisa digunakan oleh para *diver* atau pengunjung. Entitas *resort* didapatkan dengan analisa terhadap kata kunci '*resort'* dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.12 Analisa Enttias *Resort*

Id_key	Id_link	keyword	frekuensi	Analisis
55	28628	Resort	3661	Diving! Menu Home Dive Destinations Dive (Resort) s Special Deals About Us Home

Dari tabel 5.12, dijelaskan bahwa terdapat *resort*. *Resort* sendiri mempunyai dalam bahasa Indonesia mempunyai arti tempat beristirahat. Tempat beristirahat penting untuk para penyelam atau pengunjung. Untuk membuktikannya, dapat dilihat dari halaman website yang berkaitan.

Gambar 5.4 Analisa Entitas *Resort*

Dari gambar 5.4, terlihat bahwa *resort* terletak pada bagian menu dari suatu *website*. Menu dari suatu *website* menghimpun beberapa data yang penting. Dibutuhkan suatu entitas untuk menghimpun data *resort* yang dapat digunakan oleh penyelam. Menu dari suatu *website* menghimpun beberapa data yang penting. Maka dari itu, *resort* dijadikan sebagai entitas karena merupakan kompenen penting di Industri *diving*.

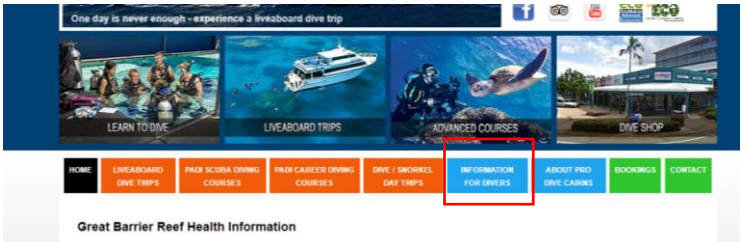
5.2.2.5 Entitas Information

Entitas *information* mendeskripsikan informasi informasi penting yang diberikan kepada penyelam. Entitas *information* didapatkan dengan analisa terhadap kata kunci '*information'* dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.13 Analisa Entitas *Information*

Id_key	Id_link	keyword	frekuensi	Analisis
216	319	information	2455	Dive Cairns News Employment Contact Us (INFORMATION) FOR DIVERS Beginners Guide to

Dari tabel 5.14, dijelaskan bahwa terdapat informasi (*information*) penting mengenai penyelam. *Information* berarti untuk membantu dalam kegiatan menyelam. Untuk mengetahui lebih lanjut, dapat dilihat dari halaman website yang berkaitan.



Gambar 5.5 Analisa Entitas *Information*

Terlihat dari gambar 5.5, bahwa *information* terdapat pada *Menu Bar*. Menu dari suatu *website* menghimpun beberapa data yang penting. Maka dari itu, *information* dijadikan sebagai entitas karena merupakan kompenen penting di Industri *diving* untuk para penyelam.

5.2.2.6 Entitas day_trips

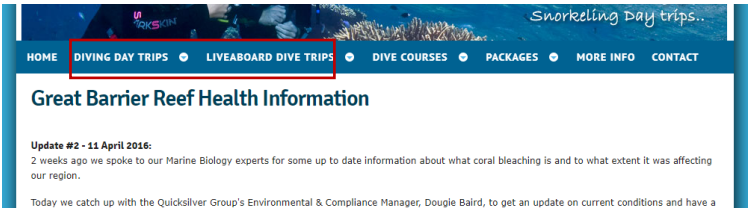
Entitas *day_trips* mendeskripsikan paket yang dapat diambil oleh para penyelam. Entitas *dive_package* didapatkan dengan analisa terhadap kata kunci '*day trip'* dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.14 Analisa Entitas *day_trips*

Id_key	Id_link	keyword	frekuensi	Analisis

20235	66725	dive trips	1747	+61 7 4041 7536 Home Diving (Day Trip) s

Pada tabel 5.15, dijelaskan bahwa terdapat *day trip* pada industri. Untuk mmembuktikan, dapat melihat halaman website yang berkaitan.



Gambar 5.6 Analisa Entitas *Day trip*

Terlihat dari gambar 5.6, bahwa *day trip* terdapat pada *Menu Bar*. Menu dari suatu *website* menghimpun beberapa data yang penting. Maka dari itu, *day_trip* dijadikan sebagai entitas karena merupakan kompenen penting di Industri *diving* untuk para penyelam.

5.2.2.7 Entitas destinations

Entitas *destination* mendeksripsikan tempat tempat yang bisa dikunjungi oleh para *diver* atau pengunjung. Entitas dive_spot didapatkan dengan analisa terhadap kata kunci '*destination*' dengan mengambil 5 kata sebelum dan sesudah nya.

Tabel 5.15 Analisa Entitas dive_spot

Id_key	Id_link	Keyword	Frekuensi	Analisis
50015	4908	Destinations	50015	DIVE CENTRE LOGIN TRAVEL INSURANCE (Destinations) Antarctica Australia Queensland South Australia

Dari kata kunci tersebut pada tabel 5.15, dijelaskan *destination* terdapat pada beberapa negara. *Destination* sendiri mempunyai arti tujuan suatu tempat. Tujuan tempat penting dalam industri *diving*. Untuk membuktikannya, dapat dilihat dari halaman website yang berkaitan.

Gambar 5.7 Analisa Entitas *destination*

Terlihat dari gambar 5.7, bahwa *day trip* terdapat pada *Menu Bar*. Menu dari suatu *website* menghimpun beberapa data yang penting. Maka dari itu, *destinations* dijadikan sebagai entitas karena merupakan kompenen penting di Industri *diving*.

5.2.2.8 Atribut Telephone

Atribut *telephone* merupakan nomor telpon yang ada pada *dive center*. Atribut ini didapatkan dari analisa terhadap kata kunci '2888'.

Tabel 5.16 Analisa Atribut *Telephone* 1

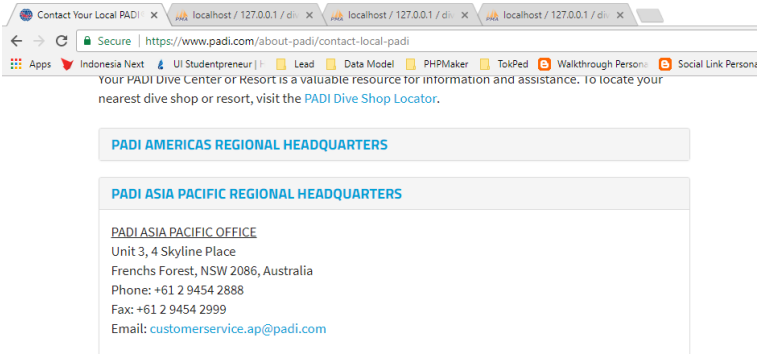
Id_key	Id_link	keyword	frekuensi	Analisis
2588	74	2888	4	2086, Australia Phone: +61 2 9454 (2888) Fax: +61 2 9454 2999

Dari kalimat pada tabel 5.17, kata kunci '2888' merupakan suatu nomor telepon. Untuk memastikan atribut *telephone* termasuk dalam entitas *dive_center*, akan dilihat pada link yang tertera.

Tabel 5.17 Analisa Atribut *Telephone*2

Id_web	Id_link	Link
1	74	https://www.padi.com/about-padi/contact-local-padi

Dari link pada tabel 5.17 terlihat bahwa data ‘2888’ merupakan nomor telepon dari PADI, maka dari itu ‘2888’ merupakan data atribut *telephone* dan entitas *dive_center*, karena PADI merupakan bagian dari *dive center*.



Gambar 5.8 Analisa Atribut *Telephone*

5.2.2.9 Atribut *address_organization*

Atribut *address merupakan* alamat dari *organization*, didapatkan dari kata kunci ‘NSW’ yang diidentifikasi sebagai data dari *address.*

Tabel 5.18 Analisa Atribut *address 1*

Id_key	Id_link	keyword	Frekuensi	Analisis
14143	74	NSW	995	Place Frenchs Forest, (NSW) 2086, Australia Phone:

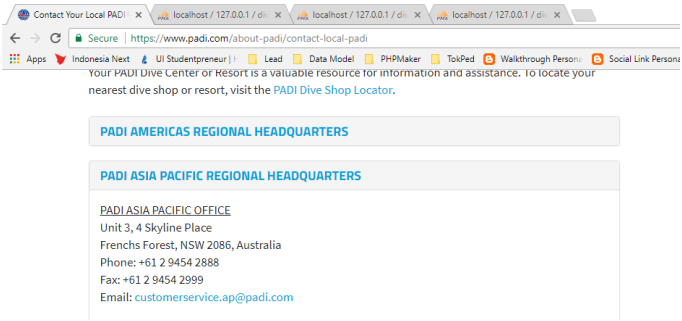
Dari tabel 5.18terlihat bahwa ‘NSW’ merupakan suatu alamat dan data dari atribut *address*. Untuk memastikan atribut *address*

termasuk dalam entitas *organization*, akan dilihat pada link yang tertera.

Tabel 5.19 Analisa Atribut *address 2*

Id_web	Id_link	Link
1	47	https://www.padi.com/about-padi/contact-local-padi

Dari link pada tabel 5.19 terlihat bahwa data 'NSW' merupakan bagian dari alamat PADI, maka dari itu 'NSW' merupakan data atribut *address* dan entitas *organization*, karena PADI merupakan bagian dari *dive center*.



Gambar 5.9Analisa Atribut *address*

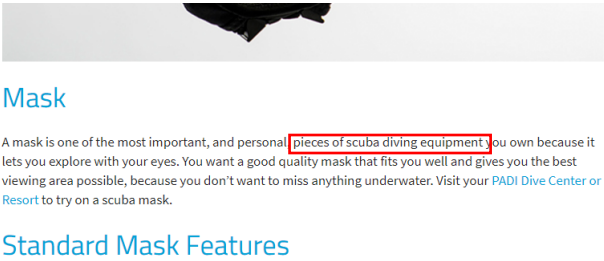
5.2.2.10 Atribut *detail_equipment*

Atribut *detail_equipment* merupakan nama dari *equipment* yang ada pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'diving equipment'*.

Tabel 5.20 Analisa Atribut *detail_equipment*

Id_key	Id_link	keyword	frekuensi	Analisis
4173	80698	Diving equipment	30	important, and personal, pieces of scuba (diving equipment) you own because it lets

Dari kalimat pada tabel 5.20, terlihat bahwa seperti sebuah konten artikel. Untuk membuktikannya, dapat dilihat dari *link* halaman *website* yang mengandung kalimat tersebut.



Gambar 5.10 Analisa Atribut *detail_equipment*

Pada gambar 5.10, merupakan halaman *website* yang akan dianalisis. Terlihat bahwa '**important, and personal, pieces of scuba (diving equipment) you own because it lets'** ditempatkan pada isi halaman dan menjelaskan item '*mask*' yang merupakan sebuah *'equipment'*. Hal itu menandakan bahwa dijadikan sebuah penjelasa dari *equipment*. Maka dari itu, terbentuklah atribut *detail_equipment*. '**important, and personal, pieces of scuba (diving equipment) you own because it lets** ' dijadikan sebagai data yang mewakili atribut *detail_equipment.*

5.2.2.11 Atribut *title_article*

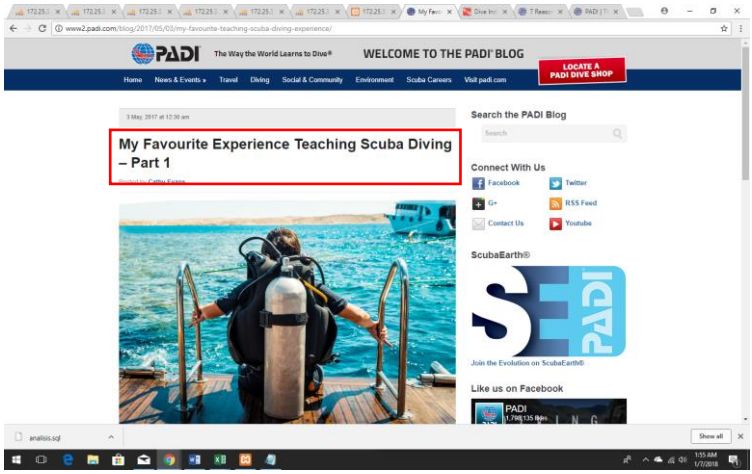
Atribut *title_article* merupakan judul dari *article* yang dapat dibaca para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Favourite Experience Teaching Scuba'.*

Tabel 5.21 Analisa Atribut *title_article*

Id_key	Id_link	Keyword	Frekuensi	Analisis
1880	127	Favourite Experience	16	My (Favourite Experience Teaching Scuba) Diving - Part 1 -

		Teaching Scuba		

Pada kalimat pada tabel 5.22, terlihat bahwa seperti sebuah judul artikel. Untuk membuktikannya, dapat dilihat dari *link* halaman *website* yang mengandung kalimat tersebut.



Gambar 5.11 Analisa Atribut *title_article*

Pada gambar 5.11, merupakan halaman *website* yang akan dianalisis. Terlihat bahwa ***'My (Favourite Experience Teaching Scuba) Diving - Part 1 –'*** ditempatkan pada bagian atas halaman. Hal itu menandakan bahwa dijadikan sebuah judul pada sebuah artikel. Maka dari itu, terbentuklah atribut *title_article*. ***'My (Favourite Experience Teaching Scuba) Diving - Part 1 –'*** dijadikan sebagai data yang mewakili atribut *title_article.*

5.2.2.12 Atribut *article_body*

Atribut *article_body* merupakan isi konten dari *article* yang dapat dibaca para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Pro can bring paychecks'*.

Tabel 5.22 Analisa Atribut *article_body*

Id_key	Id_link	Keyword	Frekuensi	Analisis
1570	258	Pro can bring paychecks	10	Brooke Morton Sure, becoming a PADI (Pro can bring paychecks) in paradise, but it also

Dari kalimat pada tabel 5.22, terlihat bahwa seperti sebuah konten artikel. Untuk membuktikannya, dapat dilihat dari *link* halaman *website* yang mengandung kalimat tersebut.

Gambar 5.12 Analisa Atribut *article_body*

Pada gambar 5.12, merupakan halaman *website* yang akan dianalisis. Terlihat bahwa '***Brooke Morton Sure, becoming a PADI (Pro can bring paychecks) in paradise, but it also'*** ditempatkan pada isi halaman. Hal itu menandakan bahwa dijadikan sebuah isi konten pada sebuah artikel. Maka dari itu, terbentuklah atribut *article_body*. '***Brooke Morton Sure, becoming a PADI (Pro can bring paychecks) in paradise, but it also*** ' dijadikan sebagai data yang mewakili atribut *title_article.*

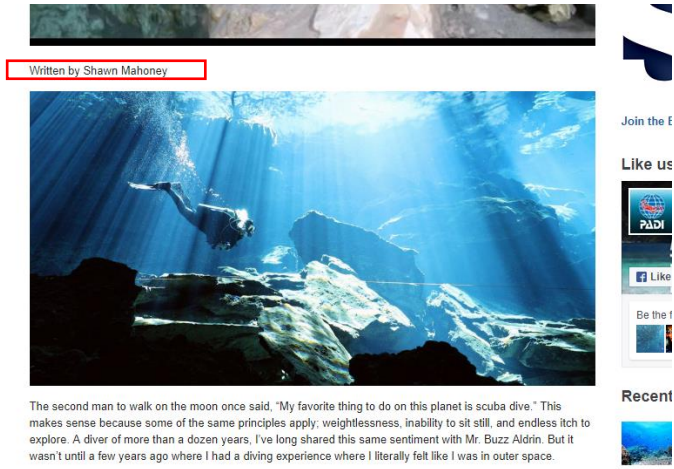
5.2.2.13 Atribut *author_article*

Atribut *author* merupakan penulis dari *article* yang dapat dibaca para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Mexico Written by Shawn Mahoney*.

Tabel 5.23 Analisa Atribut *author*

Id_key	Id_link	Keyword	Frekuensi	Analisis
1570	258	Mexico Written by Shawn Mahoney The	2	padi.com Diving the Cenotes (Caves) in (Mexico Written by Shawn Mahoney) The second man to walk

Dari kalimat pada tabel 5.23, terlihat bahwa seperti sebuah konten artikel. Untuk membuktikannya, dapat dilihat dari *link* halaman *website* yang mengandung kalimat tersebut.



Gambar 5.13 Analisa Atribut *author*

Pada gambar 5.13, merupakan halaman *website* yang akan dianalisis. Terlihat bahwa ***'Written by Shawn Mahoney'*** menjelaskan bahwa artikel ditulis oleh Shawn Mahoney. Maka

dari itu, terbentuklah atribut *author*. ‘**Shawn Mahoney** ‘ dijadikan sebagai data yang mewakili atribut *author*.

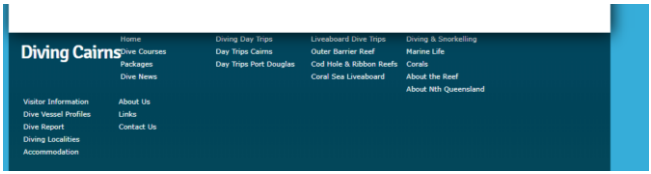
5.2.2.14 Atribut *name_dive_trip*

Atribut *name_dive_trip* merupakan nama dari *dive_trip* yang dapat diambil oleh para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *‘Day Trips Port Douglas.*

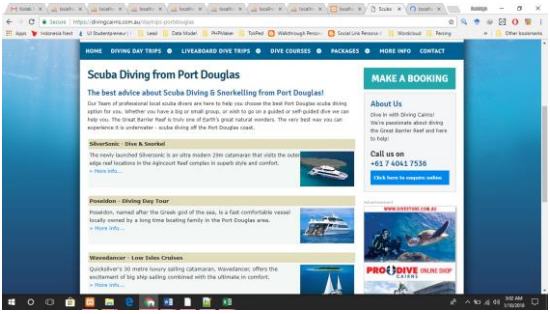
Tabel 5.24 Analisa Atribut name_dive_trips

Id_key	Id_link	keyword	frekuensi	Analisis
20734	67695	Day Trips Port Douglas	848	Diving Day Trips Day Trips Cairns (Day Trips Port Douglas) Liveaboard Dive Trips Outer Barrier

Pada tabel 5.25, seperti sebuah pilihan paket. Untuk membuktikannya dapat dilihat dari link yang berkaitan.



Gambar 5.14 Analisa atribut *name_day_trips*



Gambar 5.15 Analisa Atribut *name_day_trips* 2

Dari gambar 5.15, terlihat bahwa '*Day Trips Port Douglas'* merupakan bagian kumpulan informasi. Jika piluhan tersebut di klik, maka akan mengarah kepada pilihan paket menyelam. Maka dari itu, '*Day Trips Port Douglas'* dijadikan sebagai atribut *name_dive_package.*

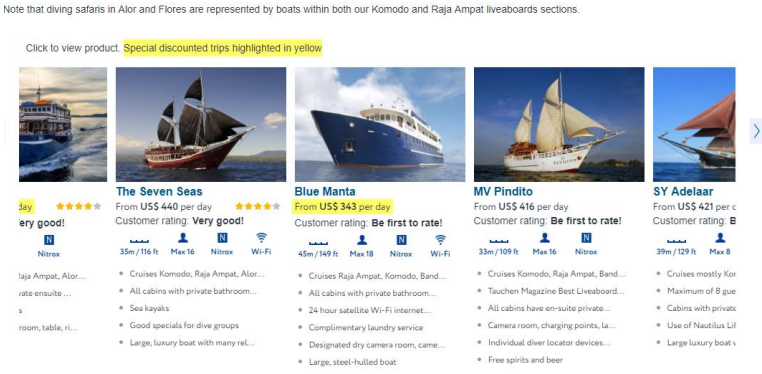
5.2.2.15 Atribut *name_transportation*

Atribut *name_tranportation* merupakan nama dari *transportation* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'MV Leo Liveaboard'*.

Tabel 5.25 Analisa Atribut *name_transportation*

Id_key	Id_link	Keyword	Frekuensi	Analisis
186003	67551	Blue Manta	81	MV Ambai KLM Sea Safari 8 (Blue Manta) SY Adelaar High Budget Dewi

Dari kalimat pada tabel 5.25, dijelaskan bahwa terdapat beberapa nama transportasi. *Blue Manta* merupakan sebuah transportasi (*transportasi*). Untuk membuktikannya dapat dilihat dari website yang berkaitan.



Gambar 5.16 Analisa Atribut *name_transportation*

Terlihat pada gambar 5.16, bahwa *blue manta* merupakan sebuah tranportasi dengan jenis kapal. Maka dari itu, terbentuklah atribut *name_transportation*.

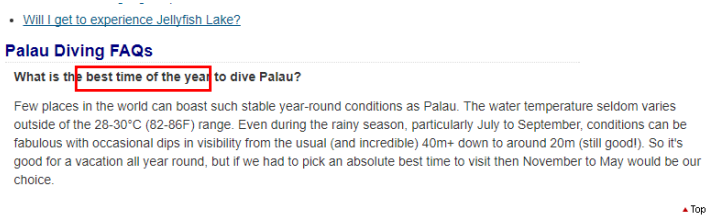
5.2.2.16 Atribut *detil_information*

Atribut *detil_information* merupakan detail dari *information* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'best time'* .

Tabel 5.26 Analisa Atribut detil_information

Id_key	Id_link	keyword	frekuensi	Analisis
20734	67695	best time	96	enlarge Diving FAQs What is the (best time) of the year to dive

Pada tabel 5.26, terdapat sebuah *FaQ* dan pertanyaan mengenai menyelam. Untuk membuktikannya dapat dilihat dari link yang berkaitan.



Gambar 5.17 Analisa atribut *detil_information*

Dari gambar 5.17, terlihat bahwa '*what is the best time of the year'* merupakan bagian dari *FaQ*. Dan hal tersebut merupakan isi dari informasi penting. Maka dari itu, terbentuklah atribut detail_information.

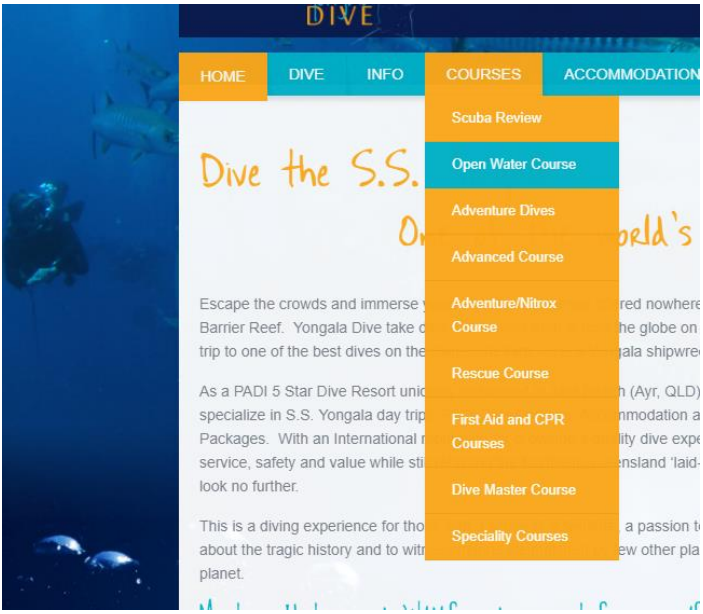
5.2.2.17 Atribut name_course

Atribut *name_courses* merupakan nama dari *courses* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Open Water'*.

Tabel 5.27 Analisa Atribut name_course

Id_key	Id_link	Keyword	Frekuensi	Analisis
619	39421	Open Water	1840	LifeInfoFAQ’sBlogDiver SafetyEco TourismReef and Coral factsCoursesScuba Review (Open Water) Course Adventure DivesAdvanced CourseAdventure/Nitrox CourseRescue CourseFirst Aid and CPR

Dari kalimat pada tabel 5.27, terlihat bahwa terdapat beberapa *course* yang ada, salah satu nya adalh *Open water Course*. *Open water diver course* merupakan sebuah pelatihan (*course*). Untuk membuktikannya dapat dilihat dari halaman website yang terkait.



Gambar 5.18 Analisa Atribut *name_course*

Terlihat dari gambar 5.18, terdapat beberapa *drop list* dari *course. Drop list* tersebut merupakan data dari entitas *course.* Maka dari itu, terbentuklah atribut *name_course*. *Open water diver course* dijadikan sebagai data yang mewakili atribut *name_course*.

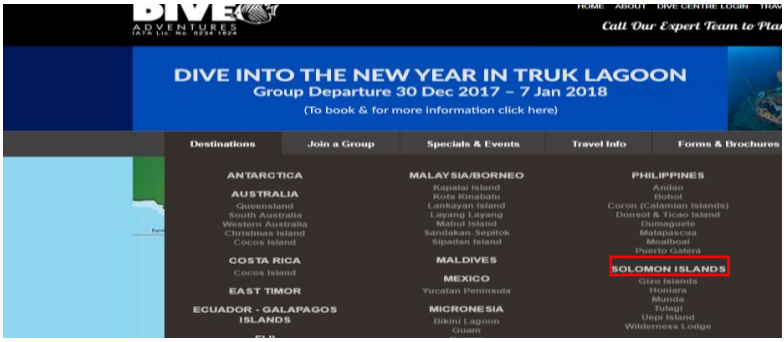
5.2.2.18 Atribut area

Atribut *area* merupakan daerah / area dari situs menyelam. Atribut tersebut didapatkan dari kata kunci sebelumnya '*Solomon Island'*.

Tabel 5.28 Analisa Atribut area

Id_key	Id_link	Keyword	Frekuensi	Analisis
26350	4908	Solomon Islands	1448	Island Dumaguete Malapascua Moalboal Puerto Galera (Solomon Islands) Gizo Islands Honiara Munda Tulagi

Dari kalimat pada tabel 5.28, dijelaskan situs selam Raja Ampat berada di suatu daearah yaitu *Solomon Island*. Untuk membuktikannya, dapat dilihat dari halaman website yang terkait.



Gambar 5.19 Analisa Atribut area

Terlihat dari gambar diatas, *Solomon Island* terletak pada *drop list* dari menu *destinations*. . *Drop list* tersebut merupakan data dari entitas *destination.* Maka dari itu, *Solomon island* merupakan daerah / area dari tempat yang bisa diselami oleh penyelam. West Papua dijadikan sebagai data yang mewakili atribut *area*.

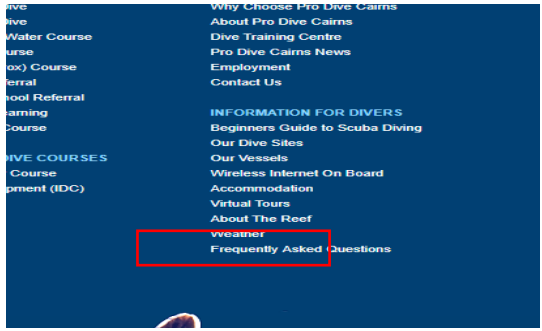
5.2.2.19 Atribut *type_information*

Atribut *type_information* merupakan nama dari *courses* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Frequently Asked Questions'* .

Tabel 5.29 Analisa Atribut *type_information*

Id_key	Id_link	keyword	frekuensi	Analisis
92	319	Frequently Asked Questions	109	Virtual Tours About The Reef Weather (Frequently Asked Questions) Great Barrier Reef Scuba Diving Liveaboard Trips

Frequently Asked Question merupakan suatu informasi mengenai pertanyaan – pertanyaan yang kerap kali ditanyakan oleh penyelam, dan disertakan dengan jawaban nya. *FaQ* dapat disebut dengan salah satu jenis informasi yang dapat membantu penyelam. Untuk mmbuktikannya dapat dilihat pada halaman website yang terkait.



Gambar 5.20 Analisa Atribut *type_information*

Terlihat dari gambar diatas, terdapat beberapa *drop list* dari *Information. Drop list* tersebut merupakan data dari entitas *information.*Maka dari itu, terbentuklah atribut *type_information. FaQ* merupakan salah satu data yang mewakilkan atribut *type_information.*

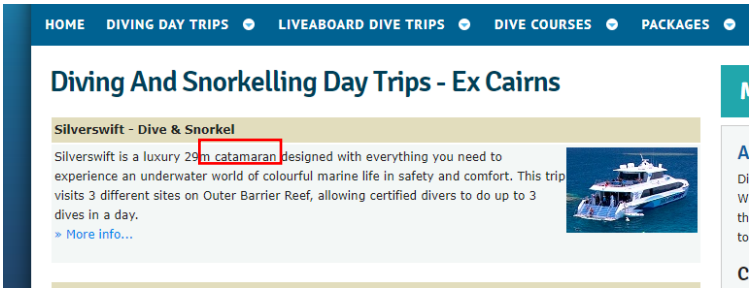
5.2.2.20 Atribut *detail_day_trip*

Atribut *detail_day_trip* merupakan detil dari *dive_package* yang dapat digunakan para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'catamaran'* .

Tabel 5.30 Analisa Atribut *detil_day_trip*

Id_key	Id_link	keyword	Frekuensi	Analisis
11718	66734	catamaran	99	Snorkel Silverswift is a luxury 29m (catamaran) designed with everything you need

Dari kalimat pada tabel 5.30, seperti menjelaskan suatu hal. Untuk pendeskripsian lebih jelas, dapat dilihat melalui halaman webiste yang berkaitan.



Gambar 5.21 Analisa Atribut *detail_day_trip*

Terlihat dari gambar 5.21, bahwa *catamaran* merupakan bagian dari penjelasan suatu *day trips*. Maka dari itu, *catamaran* beserta isinya merupakan sebuah atribut *detail_day_trip.*

5.2.3 Analisa Menggunakan *Google*

Berikut merupakan hasil analisa pada tiap kata kunci melalui mesin pencari *Google*.

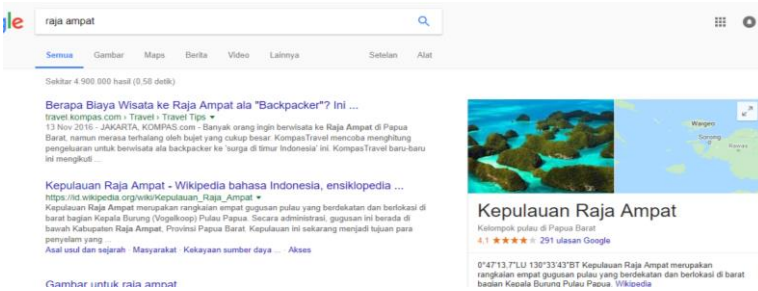
5.2.3.1 Atribut *name_ dive_spot*

Atribut *name_ dive_spot* merupakan nama dari *place* yang ada pada indsutri diving. Atribut tersebut didapatkan dari kata kunci '*Raja Ampat*'.

Tabel 5.31 Analisa Atribut name_dive_spot

Id_key	Id_link	keyword	frekuensi	Analisis
16375	4846	Raja Ampat	1544	The World Indonesia Diving Sites Komodo (Raja Ampat) Sulawesi > Bunaken Island Lembeh Strait

Dari kalimat pada tabel 5.31, dijelaskan bahwa Raja Ampat merupakan nama tempat menyelam yang ada di Indonesia. Untuk membuktikannya, dapat dilakukan dengan cara melakukan pengecekan lebih lanjut melalui pencarian *Google.*



Gambar 5.22 Analisa Atribut *name_dive_spot*

Dari gambar 5.22, terlihat jelas bahwa Raja Ampat merupakan sebuah pulau, terdapat artikel mengenai wisata disana. Maka

dari itu, Raja ampat merupakan data yang mewakili atribut *name_dive_spot*.

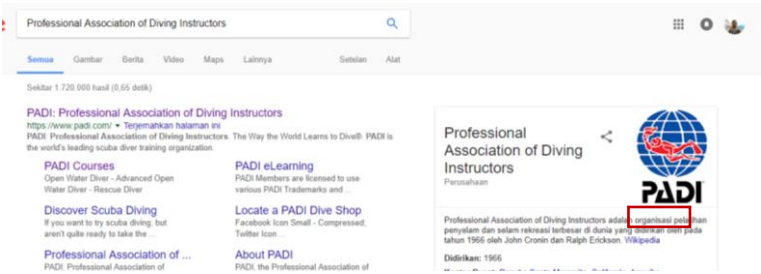
5.2.3.2 Atribut *name_organization*

Atribut name_organization merupakan nama dari *organization* yang ada, didapatkan dari kata kunci ‘PADI’ yang diidentifaksi sebagai data dari name_ *organization*.

Tabel 5.32 Analisa Atribut *name_organization*

Id_key	Id_link	keyword	frekuensi	Analisis
92	72	PADI	6842	About (PADI) Dive Centers and Resorts

Dari kalimat pada tabel 5.32, PADI merupakan salah satu dari *organization* pada industri *diving*. Untuk membuktikannya, dapat dilakukan dengan cara melakukan pengecekan lebih lanjut melalui pencarian *Google*.



Gambar 5.23 Analisa Atribut *name_organization*

Dari gambar 5.23, terlihat keterangan bahwa PADI adalah suatu organisasi diving. Maka dari itu PADI merupakan data dari *name_ organization* yang juga menjelaskan entitas *organization.*

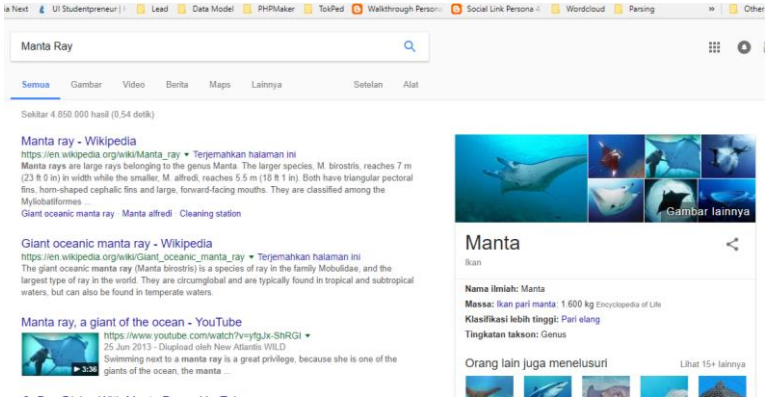
5.2.3.3 Atribut *name_animal*

Atribut *name_animal* merupakan nama dari *animal* yang dapat dinikmati para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'ray'*.

Tabel 5.33 Analisa Atribut *name_animal*

Id_key	Id_link	Keyword	Frekuensi	Analisis
8267	7660	Ray	1434	First we snorkelled with the Manta (Ray) s which was an amazing yet

Dari kalimat pada tabel 5.34, dijelaskan bahwa seseorang menyelam dengan ikan pari manta (*Manta Ray)*. *Manta Ray* merupakan salah satu nama hewan (*animal*) yang bertipe ikan (*fish)*. Untuk membuktikannya, dapat dilakukan dengan cara melakukan pengecekan lebih lanjut melalui pencarian *Google.*



Gambar 5.24 Analisa Atribu *name_animal*

Maka dari itu, *manta ray* merupakan nama dari dari *animal*. *Manta ray* dijadikan sebagai data yang mewakili atribut name_*animal*.

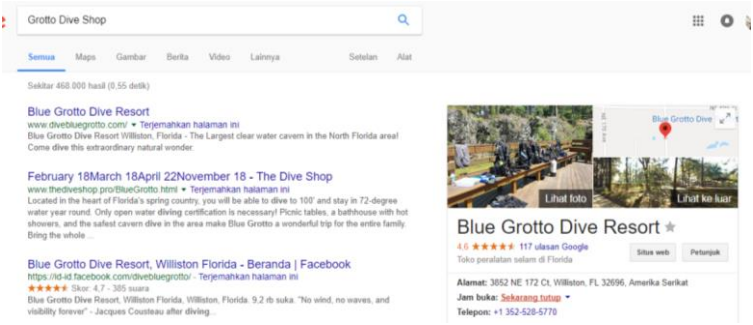
5.2.3.4 Atribut name_store

Atribut *name_store* merupakan nama dari *store* yang dapat dinikmati para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Grotto Dive Shop'*.

Tabel 5.34 Analisa Atribut *name_store*

Id_key	Id_link	Keyword	Frekuensi	Analisis
144104	48304	Grotto Dive Shop	7	Sunday, Sep. 2, 2012 Location: Blue (Grotto Dive Shop) : Adventure Outfitters Rating: view log Bill Miller

Dari kalimat pada tabel 5.35, dijelaskan bahwa terdapat sebuah shop bernama *'Grotto Dive Shop' Grotto Dive Shop* didefinisikan sebagai nama sebua *store*. Untuk membuktikannya, dapat dilakukan dengan cara melakukan pengecekan lebih lanjut melalui pencarian *Google*.



Gambar 5.25 Analisis Atribut *name_store*

Dari gambar 5.25, terlihat keterangan bahwa *Grotto Dive Shop* merupakan tempat Maka dari itu, terbentuklah atribut *name_store*. *Grotto Dive Shop* dijadikan sebagai data yang mewakili atribut *name_store*.

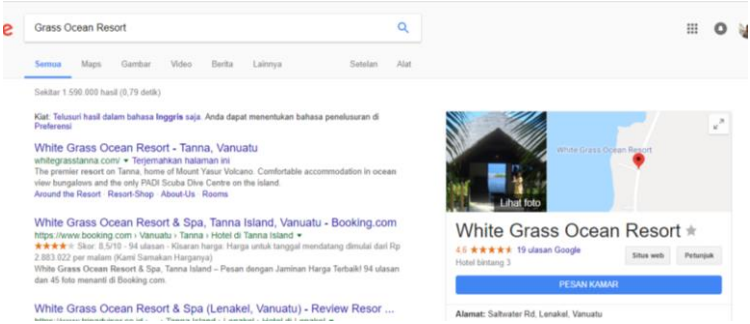
5.2.3.5 Atribut *name_resort*

Atribut *name_resort* merupakan nama dari *resort* yang dapat dinikmati para penyelam pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'Grass Ocean Resort'*.

Tabel 5.35 Analisa Atribut *name_resort*

Id_key	Id_link	Keyword	Frekuensi	Analisis
67847	5410	Grass Ocean Resort	83	In Tanna we stayed at White (Grass Ocean Resort) which now has an onsite dive operator

Dari kalimat pada tabel 5.35, dijelaskan bahwa seorang penyelam tinggal di suatu tempat bernama *White Grass Ocean Resort. White Grass Ocean Resort* didefinisikan sebagai nama sebua *resort*. Untuk membuktikannya dapat dilihat dari hasil pencarian *Google.*



Gambar 5.26 Analisa Atribut *name_resort*

Terlihat dari gambar 5.26, bahwa *White Grass Ocean Resort* merupakan sebuah penginapan. Maka dari itu, terbentuklah atribut *name_resort*. *White Grass Open Ocean* dijadikan sebagai data yang mewakili atribut *name_resort*.

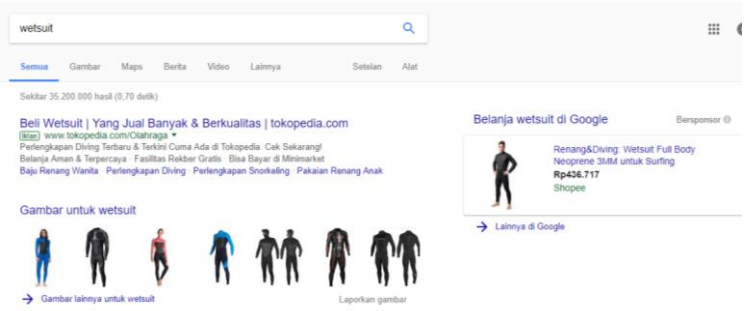
5.2.3.6 Atribut *name_equipment*

Atribut *name_equipment* merupakan nama dari *equipment* yang ada pada indsutri diving. Atribut tersebut didapatkan dari kata kunci *'mask'*.

Tabel 5.36 Analisa *Atribut name_equipment*

Id_key	Id_link	keyword	frekuensi	Analisis
3960	7682	wetsuit	103	you will need a 7 mm (wetsuit) , a 5mm one in summer.

Dari kalimat pada tabel 5.36, dijelaskan bahwa *wetsuit* merupakan sesuatu yang dibutuhkan oleh seseorang. Untuk membuktikannya, dapat dilihai dari pencarian *Google*.



Gambar 5.27 Analisa Atribut name_equipmnet

Terlihat dari gambar 5.27, hasil dari pencarian *wetsuit* mengarah pada baju yang digunakan untuk menyelam. Maka para penyelam dapat menggunakan *wetsuit* untuk melakukan aktivitas menyelam. *Name_equpiment* dijadikan sebagai atribut dan *wetsuit* sebagai salah satu data.

5.2.3.7 Atribut *role_dive_operator*

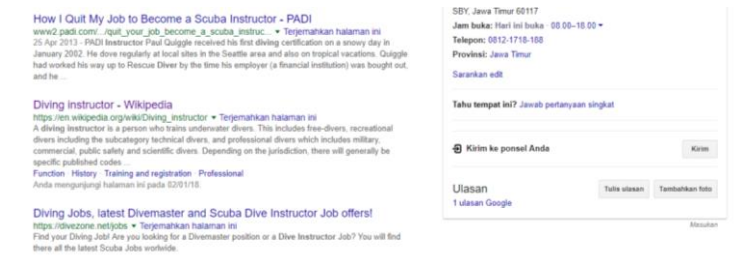
Atribut *role_dive_operator* merupakan nama dari orang yang terlibat dengan industri *diving*, didapatkan dari kata kunci '*dive*

instructor' yang mempunyai frekuensi sebanyak 415 kali muncul pada berbagai halaman website.

Tabel 5.37 Analisa Atribut name_person

Id_key	Id_link	Keyword	Frekuensi	Analisis
11793	79894	Dive insturctor	415	Fishman says. "As a dive instructor, I work every day to bring

Dari kalimat pada tabel 5.37, dijelaskan bahwa sesorang bernama *fishman* bekerja sebagai *dive instructor*. Untuk mengetahui maksud dari *dive instructor* dapat dilakukan dengan pencarian *Google*.



Gambar 5.28 Analisa *role_dive_operator*

Dari hasil pencarian *Google*, terdapat penjelasan pada *Wikipedia* yang menjelaskan mengenai *dive instructor*. *Dive instructor* adalah seseorang yang memberikan pelatihan kepada para penyelam. *Dive instructor* mempunyai peran dalam memberikan layanan kepada para penyelam dan pengerak dari industri *diving*. *Role_dive_operator* dijadikan sebagai atribut dan *dive instructor* merupakan data dari atribut tersebut.

5.2.4 Entitas yang diturunkan dari Data dan Atribut

Terdapat beberapa entitas yang belum teridentifikasi dengan menggunakan kata kunci yang ada. Berikut beberapa tambahan entitas yang berdasarkan data dan atribut.

5.2.4.1 Entitas *organization*

Entitas *organization* mendeskripsikan suatu tempat segala informasi mengenai industri *diving*. Entitas ini muncul karena adanya atribut *name_organization, telephone_organization, address_organization*

5.2.4.2 Entitas *dive_operator*

Entitas *dive_operator* medeskripsikan orang orang yang mempunyai peran dan membantu para diver di industri diving Entitas ini muncul karena adanya atribut *name_dive_operator dan role_operator*

5.2.4.3 Entitas *equipment*

Entitas *equipment* mendeksripsikan peralatan yang dapat digunakan oleh para *diver* atau pengunjung saat melakukan aktivitas di industri *diving*. Entitas ini muncul karena adanya atribut *name_equipment dan detail_equipment.*

5.2.4.4 Entitas *Article*

Entitas artcile mendeksripsikan informasi informasi yang berkaitan dengan industri *diving* yang dapat dibaca oleh para *diver*. Entitas ini muncul karena adanya *title_article, author, dan article_body.*

5.2.4.5 Entitas *Transportation*

Entitas *transportation* merupakan suatu kumpulan kendaraan yang dapat digunakan oleh parah penyela atau pengunjung. Entitas ini muncul karena adanya atribut *name_transportation.*

5.2.4.6 Entitas *Promotion*

Entitas promotion mendeskripsikan penawaran menarik kepada para penyelam. Entitas ini muncul karena adanya atribut *name_promo, dan detail_promo.*

(halaman sengaja dikosongkan)

BAB VI
HASIL DAN PEMBAHASAN

Pada bab ini akan membahas mengai hasil dari analisa pengkatagorian model data pada industri *diving.* Pengkatagorian model data tersebut antara lain, entitas, atribut, dan data. Dan juga akan dilakukan penilaian *quality of assurance* terhadap hasil dari rancangan model data tersebut.

6.1 Hasil Analisis Frekuensi

Pada sub bab ini akan dijelaskan hasil mengenai analisis frekuensi yang sudah dilakukan. Analisis frekuensi akan dilakukan pada tiap suku kata yang berasal dari konten halaman *website* yang sudah di pecah.

6.1.1 Analisis Frekuensi pada 1 Suku Kata

Analisis frekuensi akan dilakukan dengan cara penghitungan kemunculan suatu halaman *website* yang digunakan. Setelah didapatkan frekuensi dari tiap kata, akan dilakukan penjumlahan keseluruhan frekuensi. Dan total dari keseluruhan frekuensi 1 suku kata adalah sebanyak 1454918 kata.

Selanjutnya seluruh kata kunci akan diurutkan berdasarkan frekuensi dari yang tertinggi ke yang terendah. Dan akan dikalkulasikan persentase frekuensi relatif, yang merupakan persentase dari perbandingan antara jumlah frekuensi kata kunci dengan total keseluruhan kata kunci

Tabel 6.1 Analisis Frekuensi 1 Suku Kata

No	Id_key	Keyword	Frekuensi	Presentase
1	49	Dive	28090	1.931%
2	4	Diving	26523	1.823%
3	75	Scuba	12412	0.853%
4	4108	Reef	9955	0.684%
5	80	Travel	7828	0.538%

6	7	PADI	6842	0.470%
7	37286	Cloudflare	6721	0.462%
8	2476	snorkeling	5659	0.389%
9	908	ISLAND	5290	0.364%
10	960	Sea	5279	0.363%
11	1234	Best	5242	0.360%
12	149	Divers	4884	0.336%
13	12250	brochures	4572	0.314%
14	153	Day	4540	0.312%
15	3999	Fish	4100	0.282%
16	977	GREAT	3980	0.274%
17	1095	Events	3923	0.270%
18	14	World	3854	0.265%
19	37285	US	3723	0.256%
20	4107	Coral	3680	0.253%
21	1570	Dives	3676	0.253%
22	55	Resort	3661	0.252%
23	41460	&amp	3528	0.242%
24	7671	Reefs	3460	0.238%
25	1375	Marine	3374	0.232%

Dari tabel 6.1, menunjukkan 25 kata kunci teratas berdasarkan frekuensi. 25 kata kunci ini akan dilakuka identifikasi mengenai entitas data. Dan dari analisa frekuensi pada 1 suku kata, didapatkan kata *'Dive'* yang mempunyai frekuensi paling tinggi dengan presentase kemunculan 1.931% dari keseluruhan kata kunci 1 suku kata

6.1.2 Analisis Frekuensi pada 2 Suku Kata

Analisis frekuensi akan dilakukan dengan cara penghitungan kemunculan suatu halaman *website* yang digunakan. Setelah didapatkan frekuensi dari tiap kata, akan dilakukan penjumlahan keseluruhan frekuensi. Dan total dari keseluruhan frekuensi 2 suku kata adalah sebanyaka 1.829.150 kata.

Selanjutnya seluruh kata kunci akan diurutkan berdasarkan frekuensi dari yang tertinggi ke yang terendah. Dan akan dikalkulasikan persentase frekuensi relatif, yang merupakan persentase dari perbandingan antara jumlah frekuensi kata kunci dengan total keseluruhan kata kunci.

Tabel 6.2 Analisis Frekuensi 2 Suku Kata

No	Id_key	Keyword	Frekuensi	Presentase
1	16224	Barrier Reef	4763	0.2604%
2	156901	Diving Cairns	3535	0.1933%
3	16223	Great Barrier	3266	0.1786%
4	48095	Dive Adventures	2941	0.1608%
5	17136	liveaboard dive	2576	0.1408%
6	10612	marine life	2514	0.1374%
7	72	Scuba Diving	2493	0.1363%
8	156831	4041 7536	2239	0.1224%
9	31854	About Us	2206	0.1206%
10	17162	Coral Sea	1879	0.1027%
11	26235	Papua New	1866	0.1020%
12	716	Open Water	1840	0.1006%
13	156912	Cairns	1809	0.0989%
14	18032	dive trips	1747	0.0955%

15	45014	liveaboard diving	1732	0.0947%
16	17156	Cod Hole	1727	0.0944%
17	16375	Raja Ampat	1544	0.0844%
18	156855	Vessel Profiles	1539	0.0841%
19	17305	Outer Barrier	1524	0.0833%
20	84816	diving day	1521	0.0832%
21	26236	New Guinea	1501	0.0821%
22	156841	&amp Ribbon	1476	0.0807%
23	18872	Sea Liveaboard	1473	0.0805%
24	156840	Hole &amp	1473	0.0805%
25	18812	Call Us	1467	0.0802%

Dari tabel diatas, menunjukkan 25 kata kunci teratas berdasarkan frekuensi. 25 kata kunci ini akan dilakukan identifikasi mengenai entitas data. Dan akan dikalkulasikan persentase frekuensi relatif, yang merupakan persentase dari perbandingan antara jumlah frekuensi kata kunci dengan total keseluruhan kata kunci. Dan dari analisa frekuensi pada 2 suku kata, didapatkan kata *'Barrier Reef'* yang mempunyai frekuensi paling tinggi dengan presentase kemunculan 0.2604% dari keseluruhan kata kunci 2 suku kata

6.1.3 Analisis Frekuensi pada 3 Suku Kata

Analisis frekuensi akan dilakukan dengan cara penghitungan kemunculan suatu halaman *website* yang digunakan. Setelah didapatkan frekuensi dari tiap kata, akan dilakukan penjumlahan keseluruhan frekuensi. Dan total dari keseluruhan frekuensi 2 suku kata adalah sebanyaka 973.713 kata.

Selanjutnya seluruh kata kunci akan diurutkan berdasarkan frekuensi dari yang tertinggi ke yang terendah. Dan akan dikalkulasikan persentase frekuensi relatif, yang merupakan

persentase dari perbandingan antara jumlah frekuensi kata kunci dengan total keseluruhan kata kunci.

Tabel 6.3 Analisis Frekuensi 3 Suku Kata

No	Id_key	Keyword	Frekuensi	Presentase
1	14202	Great Barrier Reef	2280	0.2342%
2	23069	Papua New Guinea	2014	0.2068%
3	61980	Galapagos Islands Fiji	1337	0.1373%
4	119069	'My Dive Shop'	1035	0.1063%
5	15521	LIVEABOARD DIVE TRIPS	948	0.0974%
6	61971	Christmas Island Cocos	935	0.0960%
7	61979	East Timor Ecuador	926	0.0951%
8	174557	7 4041 7536	924	0.0949%
9	62092	Philippines Solomon Islands	908	0.0933%
10	174602	Dive Vessel Profiles	897	0.0921%
11	15251	Outer Barrier Reef	896	0.0920%
12	162128	Diving day trips	882	0.0906%
13	174575	&amp Ribbon Reefs	865	0.0888%
14	174573	Cod Hole &amp	863	0.0886%
15	174574	Hole &amp Ribbon	863	0.0886%
16	16674	Coral Sea Liveaboard	862	0.0885%
17	174564	Cairns Day Trips	854	0.0877%
18	73198	day trips Cairns	853	0.0876%
19	174561	Day Trips Day	852	0.0875%
20	174562	Trips Day Trips	852	0.0875%

21	174563	Trips Cairns Day	850	0.0873%
22	174611	Queensland Visitor Information	850	0.0873%
23	123778	marine life corals	849	0.0872%
24	174565	Day Trips Port	849	0.0872%
25	174566	Trips Port Douglas	849	0.0872%

Dari tabel diatas, menunjukkan 25 kata kunci teratas berdasarkan frekuensi. 25 kata kunci ini akan dilakukan identifikasi mengenai entitas data. Dan dari analisa frekuensi pada 3 suku kata, didapatkan kata *' Great Barrier Reef'* yang mempunyai frekuensi paling tinggi dengan presentase kemunculan 0.2342% dari keseluruhan kata kunci 3 suku kata

6.1.4 Analisis Frekuensi pada 4 Suku Kata

Analisis frekuensi akan dilakukan dengan cara penghitungan kemunculan suatu halaman *website* yang digunakan. Setelah didapatkan frekuensi dari tiap kata, akan dilakukan penjumlahan keseluruhan frekuensi. Dan total dari keseluruhan frekuensi 2 suku kata adalah sebanyaka 1.007.06 kata.

Selanjutnya seluruh kata kunci akan diurutkan berdasarkan frekuensi dari yang tertinggi ke yang terendah. Dan akan dikalkulasikan persentase frekuensi relatif, yang merupakan persentase dari perbandingan antara jumlah frekuensi kata kunci dengan total keseluruhan kata kunci.

Tabel 6.4 Analisis Frekuensi 4 Suku Kata

No	Id_key	Keyword	Frekuensi	Presentase
1	52811	Niue Papua New Guinea	1160	0.1152%
2	52868	Papua New Guinea Philippines	1158	0.1150%

3	52872	Solomon Islands Sri Lanka	1150	0.1142%
4	52869	New Guinea Philippines Solomon	1149	0.1141%
5	52870	Guinea Philippines Solomon Islands	1149	0.1141%
6	52871	Philippines Solomon Islands Sri	1149	0.1141%
7	52864	Galapagos Islands Fiji Indonesia	1147	0.1139%
8	13093	Cod Hole &amp Ribbon	984	0.0977%
9	13437	Hole &amp Ribbon Reefs	923	0.0917%
10	156173	Day Trips Day Trips	851	0.0845%
11	156172	Diving Day Trips Day	849	0.0843%
12	156174	Trips Day Trips Cairns	849	0.0843%
13	156175	Day Trips Cairns Day	849	0.0843%
14	156176	Trips Cairns Day Trips	848	0.0842%
15	156177	Cairns Day Trips Port	848	0.0842%
16	156178	Day Trips Port Douglas	848	0.0842%
17	156179	Trips Port Douglas Liveaboard	848	0.0842%
18	156180	Port Douglas Liveaboard Dive	848	0.0842%
19	156181	Douglas Liveaboard Dive Trips	848	0.0842%
20	156167	+61 7 4041 7536	847	0.0841%
21	156182	Liveaboard Dive Trips Outer	847	0.0841%
22	156183	Dive Trips Outer Barrier	847	0.0841%

23	156184	Trips Outer Barrier Reef	847	0.0841%
24	156185	Outer Barrier Reef Cod	847	0.0841%
25	156186	Barrier Reef Cod Hole	847	0.0841%

Dari tabel diatas, menunjukkan 25 kata kunci teratas berdasarkan frekuensi. 25 kata kunci ini akan dilakukan identifikasi mengenai entitas data. Dan dari analisa frekuensi pada 4 suku kata, didapatkan kata *' Niue Papua New Guinea'* yang mempunyai frekuensi paling tinggi dengan presentase kemunculan 0.1152% dari keseluruhan kata kunci 4 suku kata

6.1.5 Analisis Frekuensi pada 5 Suku Kata

Analisis frekuensi akan dilakukan dengan cara penghitungan kemunculan suatu halaman *website* yang digunakan. Setelah didapatkan frekuensi dari tiap kata, akan dilakukan penjumlahan keseluruhan frekuensi. Dan total dari keseluruhan frekuensi 2 suku kata adalah sebanyaka 773.867 kata.

Selanjutnya seluruh kata kunci akan diurutkan berdasarkan frekuensi dari yang tertinggi ke yang terendah. Dan akan dikalkulasikan persentase frekuensi relatif, yang merupakan persentase dari perbandingan antara jumlah frekuensi kata kunci dengan total keseluruhan kata kunci.

Tabel 6.5 Analisis Frekuensi 5 Suku Kata

No	Id_key	Keyword	Frekuensi	Presentase
1	38906	New Guinea Philippines Solomon Islands	1526	0.1972%
2	38905	Papua New Guinea Philippines Solomon	1525	0.1971%
3	38907	Guinea Philippines Solomon Islands Sri	1525	0.1971%
4	38908	Philippines Solomon Islands Sri Lanka	1525	0.1971%

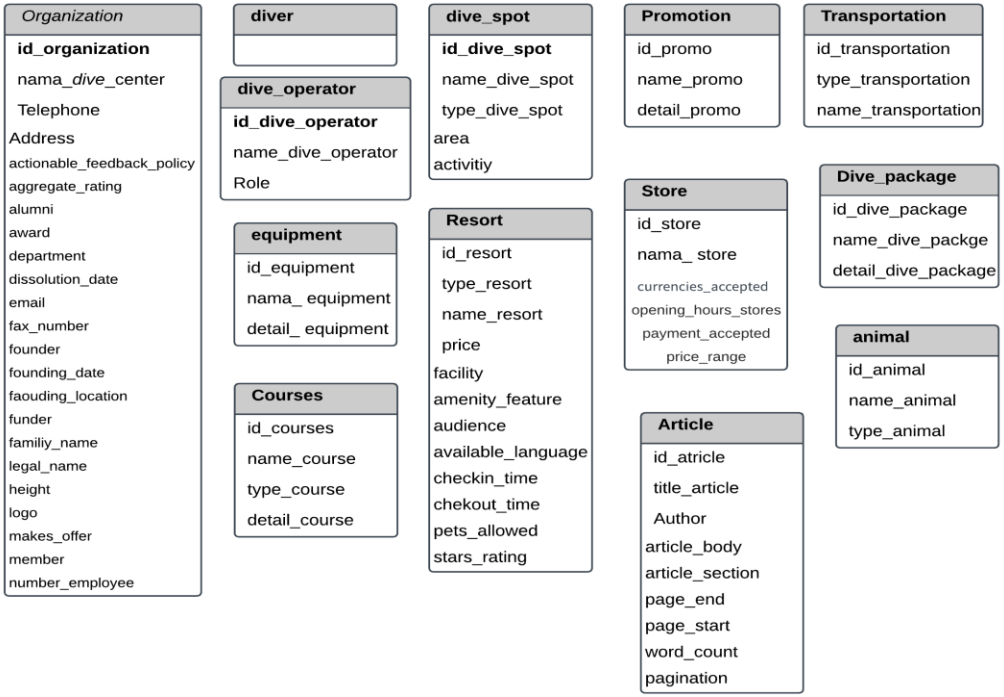
5	146187	Diving Day Trips Day Trips	846	0.1093%
6	146188	Day Trips Day Trips Cairns	846	0.1093%
7	146189	Trips Day Trips Cairns Day	846	0.1093%
8	146190	Day Trips Cairns Day Trips	846	0.1093%
9	146191	Trips Cairns Day Trips Port	846	0.1093%
10	146192	Cairns Day Trips Port Douglas	845	0.1092%
11	146193	Day Trips Port Douglas Liveaboard	845	0.1092%
12	146194	Trips Port Douglas Liveaboard Dive	845	0.1092%
13	146195	Port Douglas Liveaboard Dive Trips	845	0.1092%
14	146196	Douglas Liveaboard Dive Trips Outer	845	0.1092%
15	146197	Liveaboard Dive Trips Outer Barrier	845	0.1092%
16	146198	Dive Trips Outer Barrier Reef	845	0.1092%
17	146199	Trips Outer Barrier Reef Cod	845	0.1092%
18	146200	Outer Barrier Reef Cod Hole	845	0.1092%
19	146201	Barrier Reef Cod Hole &amp	845	0.1092%
20	146205	Ribbon Reefs Coral Sea Liveaboard	845	0.1092%

21	146235	Reef About Nth Queensland Visitor	844	0.1091%
22	146236	About Nth Queensland Visitor Information	844	0.1091%
23	39125	Micronesia Niue Papua New Guinea	821	0.1061%
24	39126	Niue Papua New Guinea Philippines	821	0.1061%
25	39110	Destinations Antarctica Australia Borneo Christmas	815	0.1053%

Dari tabel diatas, menunjukkan 25 kata kunci teratas berdasarkan frekuensi. 25 kata kunci ini akan dilakukan identifikasi mengenai entitas data. Dan dari analisa frekuensi pada 4 suku kata, didapatkan kata *' New Guinea Philippines Solomon Islands'* yang mempunyai frekuensi paling tinggi dengan presentase kemunculan 0.1972% dari keseluruhan kata kunci 4 suku kata

6.2 Rancangan Entitas Data

Pada sub bab kali ini, akan dijelaskan hasil dari perancagan model data yang dihasilkan dari analisa terhadap *website* mengenai industri *diving*. Didalam model data ini akan terdapat entitas dan atribut yang didapatkan dari analisa terhadap *website* tersebut. Beberapa entitas dan atribut akan mengikuti standarisasi dari *schema.org* dalam pemberian nama entitas dan juga atribut yang ada pada entitas tersebut. Penggunaan standarisasi *schema.org* digunakan sesuai dengan kebutuhan dan dilakukan penyesuaian.



Gambar 6.1 ***Entity Relational Diagram***

6.1.1 Entitas *organization*

Entitas *organization* mendeskripsikan suatu tempat segala informasi mengenai industri *diving*. Didalam *schema.org*,

Tabel 6.6 Entitas *dive_center*

Organization
id_organization
nama_dive_center
Telephone
Address

6.1.1.1 Atribut berdasarkan Standar

Terdapat beberapa atribut yang akan ditambahkan sesuai dengan standar yang digunakan, *schema.org*. Standar *schema.org* yang digunakan pada entitas *organzation* adalah tipe *organization* dengan atribut berikut.

Tabel 6.7 Atribut berdasarkan Standarisasi Pada Enttias Organization

No	Atribut *(property)*	Keterangan
1	*actionable_feedback_policy*	Publikasi media terhadap organisasi
2	*aggregate_rating*	Nilai kesulurhan
3	*Alumni*	Orang orang yang sudah menjadi alumni
4	*Award*	Penghargaan yang pernah diraih
5	*Department*	Bagian bagian yang ada pada organisasi

6	*dissolution_date*	Tanggal organisasi bubar
7	*Email*	Alamat organisasi
8	*fax_number*	Nomor fax
9	*Founder*	Pendiri organisasi
10	*founding_date*	Tanggal didirikan
11	*founding_location*	Tempat didirikan
12	*Funder*	Yang mendukung dana
13	*family_name*	Nama Keluarga
14	*legal_name*	Nama organisasi
15	*Height*	Tinggi dari *person*
16	*Logo*	Logo organisasi
17	*makes_offer*	Penawaran / layanan yang diberikna
18	*Member*	Anggota dari organisasi
19	*number_employee*	Jumlah karyawan
20	*Parent*	Orang tua dari *person*
21	*telephone*	Nomor telfon

6.1.2 Entitas *Diver*

Entitas *diver* mendeskripsikan orang yang melakukan aktivitas di industri diving. *Diver* memiliki arti penyelam. Untuk atribut dari entitas *diver* belum bisa diidentifikasi.

6.1.3 Entitas *dive_operator*

Entitas *dive_operator* mendeskripsikan para pengerak dari industri *diving*. Didalamnya terdapat orang – orang yang membantu para penyelam dalam melakukan aktivitas menyelam. Entitas *dive_operator* terdapat 3 atribut, yaitu

id_dive_operator, *name_dive_operator*, dan *role_dive_operator*.

Tabel 6.8 Entitas *dive_operator*

Dive_operator
id_dive_operator
name_dive_operator
role_dive_operator

6.1.4 Entitas *destination*

Entitas *destination* mendeksripsikan tempat tempat yang bisa dikunjungi oleh para *diver* atau pengunjung. Terdapat berbagai macam tempat yang dapat dikunjungi oleh para penyelam. Tetapi tempat tersebut dapat melakukan banyak aktivitas lainnya selain menyelam. Entitas *destination* mempunyai 5 atribut, yaitu *id_destinatio, name_destination, type_destination, area,* dan *activity*.

Tabel 6.9 Entitas *destination*

Destination
id_destination
name_destination
type_destination
Area
Activitiy

6.1.5 Entitas *equipment*

Entitas *equipment* mendeksripsikan peralatan yang dapat digunakan oleh para *diver* atau pengunjung saat melakukan aktivitas di industri *diving*. Peralatan ini dapat membantu para penyelam dalam melakukan ativtias didalam industri *diving*.

Entitas *equipment* mempunyai atribut *id_equipment*, *name_equipment*, dan *detail_equipment*.

Tabel 6.10 Entity *Equipment*

Equipment
id_equipment
nama_ equipment
detail_ equipment

6.1.6 Entitas *Animal*

Entitas animal mendeksripsikan hewan yang hidup di industri *diving* dan para *diver* dapat melihatnya. Hewan yang ada di industri *diving* merupakan objek yang menarik untuk di lihat dan dinikmati oleh para penyelam. Entitas animal mempunyai atribut *id_animal*, *name animal*, dan *type_animal*

Tabel 6.11 Entitas *Animal*

animal
id_animal
name_animal
type_animal

6.1.7 Entitas *Resort*

Entitas resort mendeksripsikan tempat penginapan yang bisa digunakan oleh para *diver* atau pengunjung. Para penyelam membutuhkan temat peristirahatan ketika mereka melakukan berbagai aktivitas. Dan penginapan yang ditawarkan dengan berbagai tipe. Entitas resort mempunyai atribut *id_resort*, *type_resort*, *name_resort*, *price*.

Tabel 6.12 Entitas *Resort*

Resort

id_resort
type_resort
name_resort
price

6.1.7.1 Atribut berdasarkan Standar

Terdapat beberapa atribut yang akan ditambahkan sesuai dengan standar yang digunakan, *schema.org.* Standar *schema.org* yang digunakan pada entitas *resort* adalah tipe *resort* dengan atribut berikut.

Tabel 6.13 Atribut berdasarkan Standar pada Entitas *Resort*

No	Atribut *(property)*	Keterangan
1	*amenity_feature*	Ketesedian fitur yang dapat memudahkan pelanggan
2	*Audience*	Untuk siapa fasilitas disediakan
3	*available_language*	
4	*Checkin_time*	Waktu tercepat saat pelanggan dapat melakukan *check in*
5	*Checkout_time*	Waktu terlama saat pelanggan dapat melakukan *check out.*
6	*Pets_allowed*	Peraturan untuk membawa binatang
7	*Stars_rating*	Bintang yang ada pada *resort*

6.1.8 Entitas *shop*

Entitas *shop* mendeksripsikan toko peratalan pendukung melakukan aktivitas di industri *diving.* Toko khusus industri *diving* ini menyediakan berbagai alat yang dapat digunakan oleh para penyelam. Entitas *shop* mempunyai atribut *id_shop* dan *name_shop.*

Tabel 6.14 Entitas *shop*

shop
id_shop
name_ shop

6.1.8.1 Atribut berdasarkan Standar

Terdapat beberapa atribut yang akan ditambahkan sesuai dengan standar yang digunakan, *schema.org*. Standar *schema.org* yang digunakan pada entitas *store* adalah tipe *store* dengan atribut berikut.

Tabel 6.15 Atribut berdasarkan Standar pada Entitas *shop*

No	Atribut *(property)*	Keterangan
1	*currencies_accepted*	Mata uang yang berlaku
2	*opening_hours_stores*	Jam Buka
3	*Payment_accepteed*	Pembyaran yang berlaku
4	*Price_range*	Rentang harga

6.1.9 Entitas *Article*

Entitas artcile mendeksripsikan informasi informasi yang berkaitan dengan industri *diving* yang dapat dibaca oleh para *diver*. Entitas *article* mempunyai atribut *id_article*, *title_atricle*, *author*, dan *article_body*.

Tabel 6.16 Entitas *article*

Article
id_atricle
title_article
Author

article_body

6.1.9.1 Atribut berdasarkan Standar

Terdapat beberapa atribut yang akan ditambahkan sesuai dengan standar yang digunakan, *schema.org*. Standar *schema.org* yang digunakan pada entitas *article* adalah tipe *resort* dengan atribut berikut.

Tabel 6.17Atribut berdasarkan Standar pada Entitas *article*

No	Atribut *(property)*	Keterangan
1	*Article_section*	Katgori dari artikel
2	*Page_end*	Bagian akhir artikel
3	*Page_start*	Bagian awal artikel
4	*Pagination*	
5	*wordCount*	Jumlah kata pada artikel.

6.1.10 Entitas *Transportation*

Entitas *transportation* merupakan suatu kumpulan kendaraan yang dapat digunakan oleh parah penyela atau pengunjung. Entitas *transportation* mempunyai atribut *id_transportation* dan *name_transportation*.

Tabel 6.18 Entitas *transportation*

Transportation
id_transportation
name_transportation

6.1.11 Entitas *Courses*

Entitas *courses* mendeksripsikan pelatihan atau ilmu yang bisa didapatkan oleh para *diver* yang diberikan oleh *dive center*. Pelatihan ini dapat diambil oleh para penyelam untuk

meningkatan kemampuan dalam menyelam. Entitas *Course* mempunyai atribut *id_course*, *name_course*, *type_course*, dan *detail_course*.

Tabel 6.19 Entitas *Course*

Courses
id_courses
name_course
type_course
detail_course

6.1.12 Entitas *Promotion*

Entitas *promotion* mendeskripsikan penawaran menarik kepada para penyelam. Promo ini bisa dalam bentuk potong harga atau keuntungan yang didapatkan oleh penyelam. Entitas *promotion* mempunyai atribut *id_promo*, *name_promo*, dan *detail_promo.*

Tabel 6.20 Entitas *Promotion*

Promotion
id_promo
name_promo
detail_promo

6.1.13 Entitas Information

Entitas *information* mendeskripsikan informasi informasi penting yang diberikan kepada penyelam. Informasi ini dapat memberikan para penyelam mengenai apa saja yang ada pada industri *diving*. Entitas information mempunyai atribut *id_information*

Tabel 6.21 Entitas Information

Information
id_information
name_information
detail_information

6.1.14 Entitas *day_trips*

Entitas *day_trips* mendeskripsikan paket yang dapat diambil oleh para penyelam. *Day trips* mempunyai maksud paket perjalanan yang dapat digunakan oleh para penyelam. Entitas *day_trips* mempunyai atribut *id_day_trips*, *name_day_trips*, dan *detail_day_trips*

Tabel 6.22 *Entitas day_trips*

Day_trips
id_day_trips
name_day_trips
detail_day_trips

6.2 Evaluasi *Quality of Assurance*

Pada penelitian kali ini, penilian kualitas dengan menggunakan *Quality of Assurance*, berfokus pada penilaian seberapa baik model data yang sudah dibuat pada penilitian. Dan kategori yang digunakan sebagai penilaian, antara lain [15]:

Tabel 6.23 Hasil *Quality of Assurance*

No.	Metriks	Tingkat Kepentingan	Y/N	Keterangan

1	Dapatkah Cakupan dari model yang akan dibangun terdefinisikan.	Esensoial	Y	Data model yang dirancang harus mempunyai keselarasan yang jelas dengan bisnis dan pengguna
2	Dapatkah kebutuhan pengguna terdefinisikan	Esensial	N	Kebutuhan pengguna harus bisa didefinisikan.
3	Apakah data model memenuhi kebutuhan?	Esensial	N	Data model harus dapat memenuhi kebutuhan
4	Dapat kah business rules yang lengkap terdefinisikan?	Esensial	N	Sangat penting business rules yang didapat adalah sah.
5	Apakah terdapat daftar istilah yang lengkap?	Esensial	Y	Sangat penting istilah yang didapat adalah sah
6	Pengecekan normalisasi 1. Dapatkah nilai tiap data item pada tabel hanya berasal dari primary key? (Y) 2. Dapatkah data item apapun berasal dari item lain? (N) 3. Apakah nama kolom terdapat perulangan didalam tabel yang sama? (N) 4. Apakah item yang sama muncul didalam	*Desirable*	N	Model data masih tergolong baik meskipun semua pengecekan gagal.

	lebih dari satu tabel? (N)			
7	Apakah model data relevan dengan Design Patterns?	*Desirable*	N	Model data masih tergolong baik jika design patterns tidak bisa di identifikasi.
8	Apakah dilakuakn penyesuaian dengan menggunakan standarisasi?	*Desirable*	Y	Tidak terlalu esensial
9	Apakah model mudah untuk dibaca dan dimengerti?	*Desirable*	Y	
10	Dapatkah grup data yang berulang teridentifikasi?	Tidak bersifat kritis	Y	Tidak bersifat kritis. Biasanya indikasi dari tahap normalisasi.dapat diterima untuk sebuah desain *database*. Contoh seperti, Address Lines seriing berulang.
11	Dapatkah data asal dapat teridentifikasi	Tidak bersifat kritis	Y	Tidak bersifat kritis
12	Apakah model data menggunakan penamaan yang standar?	Tidak bersifat kritis	Y	

6.2.1 Cakupan dari Model Data

Pada penelitian ini, pembuatan model data berfokus pada industri *diving*. Dalam pembuatan model data, data yang dijadikan sebagai sumber data berdasarkan *website – website* yang berkaitan dengan industri *diving*. Maka dari itu, cakupan

dari model data sudah sangat terdefinisikan, karena sudah didukung dengan data yang tepat.

6.2.2 Pendefinisian Kebutuhan

Kebutuhan dari suatu system dapat definisikan dari kebutuhan dari calon dari pengguna system tersebut. Pengguna akan memberikan apa saja yang mereka butuhkan dari suatu system untuk menunjang aktivitas pada suatu bidang. Pada penilitian kali ini, untuk kebutuhan dari system tidak dapat didefinisikan. Kebutuhan dari pengguna tidak didapatkan pada penelitian ini. Data yang digunakan berasal dari *website* yang sesuai dengan bidang nya.

6.2.3 Model Data Memenuhi Kebutuhan

Model data yang telah dirancang akan dibandingkan dengan kebutuhan dari pengguna yang sudah di definisikan sebelumnya. Terpenuhi nya kebutuhan pengguna, menandakan model data sudah mengimplementasikan seluruh kebutuhan menjadi sebuah rancangan model data. Pada penelitian ini, model data tidak bisa dibandingkan dengan kebutuhan. Karena kebutuhan pengguna tidak dapat didefinisikan.

6.2.4 Kelengkapan *Business Rules*

Business rules atau aturan bisnis merupakan sebuah kebijakan atau aturan aturan yang ada pada bidang tertentu, yang nantinya akan diterapkan pada system yang akan dibangun. Pendefinisian *business rules* berguna untuk mengetahui batasan dan hubungan antar entitas pada Idenftifikasi Entitas Data. Dalam penelitian ini, *business rules* tidak dapat didefinisikan.

6.2.5 Kelengkapan Daftar Istilah

Dalam Idenftifikasi Entitas Data yang telah dilakukan dalam penelitian ini, dihasilkan entitas – entitas yang mewakili industri *diving*. Entitas – entitas tersebut didapatkan dari kata – kata kunci pada *website* yang dijadikan sebagai sumber data. Entitas tersebut merupakan bagian dari industri *diving*. Sangat penting untuk mengetahui arti dari entitas yang digunakan. Hal tersebut

dilakukan agar pihak lain yang melihat hasil rancangan penelitian ini dapat memahami dengan mudah. Pada sub bab sebelumnya, telah dijelaskan maksud dan tujuan tiap entitas yang telah ditetapkan.

6.2.6 Normalisasi

Normalisasi merupakan suatu proses untuk meminimalkan redundansi pada suatu database yang berguna agar database tersebut dapat digunakan secara optimal. Dilakukannya normalisasi agar memastikan terjadinya dpendensi pada data. Pada penelitian kali ini, proses normalisasi tidak lakukan. Karena proses Idenftifikasi Entitas Data hanya hingga pembentukan entitas dan atribut.

6.2.7 Relevansi *Design Pattern*

Design Pattern dilakukan dengan cara hubungan (*relationship*) antar entitas. Hubungan antar entitas dapat dalam bentuk *one-to-many*, *one-to-one*, dan *many-to-many*. Dalam penelitian kali ini, Idenftifikasi Entitas Data hanya hingga analisa entitas dan atribut. Untuk *relationship* belum dapat didefinisikan.

6.2.8 Penyesuaian Standarisasi

Dalam Idenftifikasi Entitas Data, standarisasi berguna untuk pengoptimalan dalam penggunaan entitas dan atribut, agar sistem yang akan dibangun dapat menjadi sistem yang lebih maksimal. Dalam penelitian kali ini, standarisasi yang digunakan adalah standarisasi metadata yang disediakan oleh *schema.org*. Penggunaan *schema.org* digunakan untuk menentukan metadata apa saja yang dapat digunakan pada bidang yang digunakan. Penggunaan metadata yang tepat dapat mengoptimalkan kemudahaan dalam melakukan pencarian terhadap konten yang ada pada *website* pada *search engine* yang dilakukan oleh pengguna [9].

6.2.9 Kemudahan Memahami Model Data

Kemudahan dalam memahami model data dapat membantu dalam memudahkan perancangan pada sistem yang akan dibangun. Pemahaman yang dilakukan terhadap tiap entitas yang ada dan juga alur dari logika model data dapat dipahami dengan baik oleh orang lain. Alur model data harus bermulai dari bagian atas-kiri menuju ke bawah-kanan [15]. Untuk penilaian pemahaman terhadap alur model tidak dapat dilakukan, karena alur model belum didefinisikan.

6.2.10 Item yang Berulang dapat di Identifikasi

Perulangan pada item model data dapat terjadi. Hal itu terjadi karena adanya denormalisasi pada model data [15]. Item berulang ini berkaitan terhadap proses normalisasi database. Pada penelitian ini, tidak terdapat item perulangan pada model data yang dirancang.

6.2.11 Data Asal dapat di Identifikasi

Data asal berguna untuk data pada model data yang sudah dirancang. Pada penelitian ini, data asal dijadikan sebagai sumber data untuk melakukan Idenftifikasi Entitas Data. Jika sumber data tidak dapat teridentifikasi maka model data tidak dapat dirancang dengan baik. Sumber data yang akan digunakan berasal dari data yang ada pada *website* yang berkaitan dengan bidang industri *diving.*

6.2.12 Penggunaan Nama yang Terstandar

Penggunaan nama yang sudah standar akan memudahkan dalam memahami model data dan model data dapat terlihat lebih konsisten. Penamaan standar merupakan memberikan penamaan yang konsisten pada kesuluruhan entitas dan atribut. Standar yang digunakan pada penilitian ini adalah.

1. Penamaan entitas dan atribut menggunakan bahasa inggris
2. Seluruh huruf entitas dan atribut menggunakan huruf kecil
3. Jika nama entitas dan atribut mempunyai lebih dari 1 kata, garis bawah (_) dijadikan sebagai pemisah antar kata.

6.2.13 Pemilihan Web

Dalam metode ini, *web* berperan penting dalam Idenftifikasi Entitas Data, yaitu sebagai sumber data. Pemilihan web dapat dilakukan dengan melakukan pencarian menggunakan mesin pencari, salah satunya adalah *Google Search Engine.* Dalam pencarian *web* yang sesuai dengan bidangnya, dibutuhkan *keyword* yang sesuai dan tepat. Setelah melakukan pencarian website, akan dipilih website yang akan digunakan.

Dalam pemilihan web di pencarian *Google*, penelitian ini menggunakan 10 pencarian teratas pada masing masing *keyword.* Hal ini dilakukan karena dalam penelitian ini berfokus pada membuat suatu rancangan model data untuk industri *diving* berdasarkan data yang ada dan dibutuhkan oleh pihak yang terlibat dalam industri *diving*. Dan dengan melakukan pencarian mengguanakn *search engine*, daftar teratas adalah yang paling relevan dengan keyword yang dipilih [19].

6.2.14 Pengambilan Konten pada Web

Setelah didapatkannya *website – website* yang berkaitan dengan bidangnya, akan dilakukan pengambilan konten yang ada pada *website – website* terpilih. Metode yang digunakan untuk pengambilan konten adalah metode *web scrapping.* Metode web scrapping akan melakukan secara keseluruhan dari halaman web, dan mengambilnya dalam pentuk dokumen atau teks. Dalam melakukan proses metode ini, akan dibagi menjadi dua bagian proses, yaitu proses pengambilan link dan pengambilan konten.

Proses dibagi menjadi dua agar dapat menyaring dokumen atau halaman yang tidak diperlukan. Pada proses pengambilan *link* dilakukan pada seluruh *web* yang sudah dipilih. *Link* yang tidak berkaitan dengan bidang dan berbentuk skrip akan tersaring oleh sistem yang sudah dibangun. *Link – link* yang sudah didapatkan akan diambil seluruh konten dan akan ditampilkan dalam bentuk teks. Penyaringan dilakukan kembali dengan membuang skrip yang ada pada *web tersebut.*

6.2.15 Melakukan Frekuensi Kata

Setelah didapatkannya konten pada *website* yang dijadikan sebagai sumber data, akan dicatat frekuensi tiap kata pada konten tersebut. Frekuensi dilakukan pada kata – kata dan melihat kata – kata apa yang sering muncul pada konten tersebut. Sebelum melakakukan penghitungan teks frekuensi, akan dilakukan pemecahan konten menjadi beberapa suku kata, yaitu 1 suku kata, 2 suku kata, 3 suku kata, 4 suku kata, dan 5 suku kata. Hal ini dilakukan untuk melihat kemungkinan frasa frasa yang muncul, dan tidak hanya satu kata saja.

6.2.16 Melakukan Analisa Tiap Frasa

Frasa – frasa yang sudah didapatkan, akan dilakukan proses analisa untuk mengetahui pengkategorian dalam model data. Pengkatagorian tersebut antara lain entitas, atribut, dan data. Dalam penelitian ini, proses dalam Idenftifikasi Entitas Data hanya sampai dengan proses pengkatagorian sumber data. Proses analisa dilakukan dengan cara pengambilan kata atau frasa, dan mengambil 5 kata sebelum dan 5 kata sesudah yang ada pada konten *web* dan akan berbentuk sebuah kalimat.

Kalimat tersebut akan membatu dalam melakukan pengkatagoirian entitas, atribut, dan data. Dan untuk penelitian kali ini cukup hingga analisa entitas.

6.3 Evaluasi Metode

Pada penelitian kali ini, dalam pembuatan model data menggunakan metode pengambilan data *scrapping* dan juga pemecahan konten menjadi kata kunci yang akan dianalisa. Dalam menjalankan metode – metode tersebut, tentunya terdapat kasus – kasus yang terjadi. Dengan mengetahui nya kasus tersebut, bisa menjadi evaluasi untuk penelitian selanjutnya atau penelitian yang serupa.

6.3.1 Pemilihan *Web*

Data yang dijadikan sebagai sumber data dalam penelitan kali ini berasal dari *website* yang berisikan tentang industri *diving*.

Penentuan *website* yang akan dijadikan sumber data, dilakukan dengan cara melakukan pencarian *website* pada mesin pencari *Google*. Saat melakukan pencarian tentunya diperlukan kata kunci. Pemilihan kata kunci yang tepat, maka akan mendapatkan *website* yang berkaitan dengan industri *diving*. Kata kunci yang digunakan pada penelitian ini adalah *diving* Philipina, *diving* Malaysia, *diving* Indonesia, *diving* Australia, *snorkling* Philipina, *snorkling* Malaysia, *snorkling* Indonesia, dan *snorkling* Australia.

Website yang akan diambil melalui mesin *Google* adalah yang termasuk dalam daftar 10 teratas hasi pencarian.

Saat melakukan pencarian didapatkan beberapa *website* yang serupa dalam 10 urutan teratas tiap kata kunci yang digunakan. Maka dari itu, *website* yang akan diambil hanya salah satu saja. Hasil dari pencarian dengan menggunakan kata kunci tersebut, didapatkan 66 *website* yang berbeda. Untuk memperkecil cakupan data, dilakukan pemilihan *website* yang berfokus pada industri *diving*. Dan dari hasil analisa yang dilakukan, didapatkan 27 *website* yang terpilih. *Website* tersebut merupakan *website* yang secara keseluruhan membahas industri *diving*.

6.3.2 Pengambilan *Link* dan Konten

Dalam pengambilan *Link* dan konten yang ada pada website dilakukan dengan metode *scrapping*. Semua halaman atau *link* yang ada pada *website* tersebut akan diambil. Tetapi didalam website tersebut terdapat halaman – halaman script dan juga halaman sosial media yang tidak berkaitan dengan industri *diving*. Maka dari itu diperlukan cara untuk menghindari pengambilan *link* tersebut. Cara yang digunakan adalah menggunakan fungsi `strpos` pada sistem yang telah dibuat.

Dan setelah link link tersebut didapatkan, proses scrapping juga dilakukan pengambilan konten pada tiap link sudah didapatkan sebelumnya. Untuk konten yang akan diambil adalah yang berbentuk text, sedangkan untuk script tidak diambil. Hal itu dilakukan dengan menggunakan `strip_tags.`

6.3.3 Pemecahan Konten dan Melakukan Text Frequency

Dalam pemecahan konten dilakukan dengan cara menggunakan fungsi `explode.` Fungsi ini dilakukan dengan memecahkan konten menjadi beberapa kata kunci, dan pengambilan kata kunci ini dilakukan per suku kata. Pada penilitian ini, dilakukan dengan mengguna 1 hingga 5 suku kata. Untuk proses dari membutuhkan waktu yang sangat lama dan juga kemampuan perangkat yang mumpuni untuk melakukan proses ini.

Penghitungan frekuensi dilakukan saat 1 kata kunci mempunyai lebih dari satu kali muncul pada konten yang ada pada link yang ada pada website tersebut. Untuk kemunculan frekuensi tertinggi yaitu didominasi oleh website yang mempunya halaman yang lebih banyak.

6.3.4 Analisa tiap Kata Kunci

Setelah didapatkannya kata kunci yang dijadikan menjadi beberapa suku kata, akan dilakukan analisa pada tiap kata kunci tersebut. Analisa dilakukan dengan mengambil kata kunci dan menggabungkannya dengan 3 kata sebelumnya dan 3 kata sesudahnya pada konten yang ada pada masing masing *link website*. Hal itu dilakukan agar dapat membantu dalam melakukan analisa.

Untuk data yang digunakan dalam analisa ini adalah 10 frekuensi teratas pada masing – masing kata kunci. Setelah dilakukannya analisa, 10 kata kunci teratas tersebut ternyata berada pada 1 *website* yang sama terjadi pada beberapa kata kunci 3 suku kata, 4 suku kata, dan 5 suku kata. Hal ini dapat terjadi karena frekuensi tertinggi didominasi oleh website yang mempnyai halaman yang banyak.

Untuk menentukan jumlah data yang diambil, digunakan dengan metode slovin.

Analisa dilakukan dengan 3 cara inti, yaitu dengan cara menganalisa kata kunci menggunakan teks, menganalisa kata kunci menggunakan halaman *website* terkait, dan menganalisa kata kunci menggunakan mesin pencari *Google*. Dalam

melakukan analisa melalui teks mengalami kendala, dikarenakan untuk mengetahui karakteristik dari entitas dan atribut pada model data. Tiap teks memiliki struktur kalimat yang berbeda dan pendefinisian tiap kata berbeda juga. Jadi adanya kesulitan dalam menganalisa pengkatagorian dengan menggunakan teks.

Untuk analisa menggunakan halaman *website* yang terkait, entitas dan atribut dapat di definisikan. Halaman *website* terkait dapat membantu dalam pendefenisian model data. Hal itu dapat dilakukan dengan cara mengetahui letak dari kata kunci tersebut. Jika kata kunci tersebut terletak pada *menu bar* dari halaman, maka dapat dikatagorikan sebagai entitas, karena menghimpun beberapa data penting terkait dengan industri *diving*. Dan jika suatu kata kunci berada pada *submenu bar*, maka dapat definisikan sebagai data dari suatu entitas dan dapat dijabarkan juga sebagai atribut dari entitas tersebut.

Dan untuk analisa dengan menggunakan bantuan mesin pencari *Google*, dapat mendefinisikan pengkatagorian dari suatu entitas, atribut ataupun data. Kata kunci yang dijadikan sebagai bahan analisa akan dimasukkan pada mesin pencari *Google*. Dan *Google* akan memberikan informasi terkait dengan kata kunci tersebut. Sehingga dapat mempermudah dalam pengkatagorian kata kunci.

BAB VII
KESIMPULAN DAN SARAN

Pada bab ini akan dijelaskan mengenai kesimpulan dari hasil penelitian dan saran untuk pengembangan penelitian di kemudian hari.

7.1 Kesimpulan

Dari hasil penelitian yang dilakukan dalam Idenftifikasi Entitas Data berdasarkan data yang ada pada *website* yang berkitan dengan industri *diving*, berikut merupakan beberapa kesimpulan yang bisa didapatkan dari penelitian tugas akhir ini adalah :

1. Metode yang digunakan pada penelitian ini untuk mengidentifikasi entitas data berdasarkan data *website* sebagai berikut.
    a. Pemilihan *website* sumber data
    b. Pengambilan konten pada *website* sumber data
    c. Melakukan pemecahan kata dan frekuensi kata
    d. Melakukan analisa tiap kata menjadi model data.
2. *Website* yang dijadikan sumber data pada penelitian ini berjumlah 27 *website* berkaitan dengan industri *diving.*
3. Kata kunci yang mempunyai frekuensi kemunculan kata paling tinggi sebagai berikut.

	Keyword	Frekuensi	Persentase
1 Suku Kata	Dive	28090	1.931%
2 Suku Kata	Barrier Reef	4763	0.2604%
3 Suku Kata	Great Barrier Reef	2280	0.2342%
4 Suku Kata	Niue Papua New Guinea	1160	0.1152%
5 Suku Kata	New Guinea Philippines Solomon Islands	1526	0.1972%

4. Model data yang dirancang hanya sebatas hingga pendefinisian entitas dan atribut. Dan Entitas yang dapat didefinisikan berjumlah 14 entitas. Perancangan entitas dan atribut juga berdasarkan standarisasi *schema.org*
5. Kualitas dari model data tidak baik, dikarenakan terdapat beberapa item yang tidak dapat terpenuhi.

7.2 Saran

Saran penulis yang dapat dipertimbangkan untuk perbaikan penelitian dan pengembangan selanjutnya dengan topik serupa adalah sebagai berikut:

1. Metode Idenftifikasi Entitas Data berdasarkan data *website* dapat dilakukan pada bidang atau industri lainnya.
2. Jika suatu kata muncul lebih dari satu kali pada suatu *website*, maka penghitungan frekuensi kata hanya di hitung. Hal tersebut bisa dilakkan jika menggunakan banyak *website* sebagai sumber data
3. Dapat dilakukan evaluasi terhadap kualitas dari model data dan metode Idenftifikasi Entitas Data.
4. Untuk peneletian selanjutnya, melanjutkan perancangn model data hingga terbentu *relationship*
5. Dapat menggunakan *website* yang lainnya sebagai sumber data yang berkaitan dengan industri *diving*.

Daftar Pustaka

[1] Y. IKAWATI, "Segitiga Koral, Jantung Dunia," 12 September 2009. [Online]. Available: http://sains.kompas.com/read/2009/02/12/10425772/segitiga.koral.jantung.dunia. [Accessed 14 February 2017].

[2] F. Shafa, "Scuba Diving Atau Free Dive, Pilih Mana?," 05 July 2012. [Online]. Available: https://travel.detik.com/read/2012/07/05/083038/1958021/1383/scuba-diving-atau-free-dive-pilih-mana. [Accessed 14 February 2017].

[3] Bali Reef Divers, "Bali & Beyond," [Online]. Available: http://www.balireefdivers.com/en/bali-beyond. [Accessed 21 Maret 2017].

[4] N. Tryfona, F. Busborg and J. G. B. Christiansen, "starER: A Conceptual Model for Data Warehouse Design," Aalborg University, Aalborg Ost.

[5] D. L. Moody, "Measuring the Quality of Data Models: An Empirical Evaluation of the Use of Quality Metrics in Practice," Prague.

[6] J. WIjaya, "Analisis Metadata Halaman Productdisplay Pada Pasar Online Di Indonesia," Surabaya, 2016.

[7] C. B. Thomas Connoly, Database System : A Practical to Design, Implementation, and Management (sixth edition), England: Pearson Education Limited, 2015.

[8] schema.org, "Welcome to schema.org," schema.org, 2011. [Online]. Available: http://schema.org/. [Accessed 7 Maret 2017].

[9] L. M. C. Phil Barker, "Learning Resource Metadata Initiative: using schema.org to describe open educational resources," OpenCourseWare Consortium Global, Bolton, 2014.

[10] Research Optimus, "What is Frequency Analysis?," [Online]. Available: https://www.researchoptimus.com/article/frequency-analysis.php. [Accessed 17 January 2018].

[11] W. Trochim, "Descriptive Statistics," 2006. [Online]. Available: https://www.socialresearchmethods.net/kb/statdesc.php. [Accessed 17 January 2018].

[12] J. J. Tejada, "On the Misuse of Slovin's Formula," The Philippine Statiscation, 3023.

[13] F. B. T. Isip, "Encouraging the use of Slovin's Formula in computing sample sizes in," ISIP.

[14] J. Penman, Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, Indiana: IPCC National Greenhouse Cost Inventories programme, 2000.

[15] B. Williams, "Quality of Assurance Data Models," Database Answers, 2012.

[16] Unknown, "Afidated," [Online]. Available: http://www.afidated.com/2015/04/meaning-of-entity-its-characteristics.html. [Accessed 20 11 2017].

[17] R. A. Mata-Toledo and P. K. Cushman, Fundamental of Relational Database, New York: Schaum's Outline, 2000.

[18] T. Lonne, "How To Become A PADI Dive Instructor," [Online]. Available: https://www.divein.com/guide/how-to-become-a-padi-dive-instructor/. [Accessed 4 Januari 2018].

[19] D. Sullivan, "How Search Engines Work," 2002.

[20] Dublin Core, "Mission and Principles," Dublin Core, [Online]. Available: http://dublincore.org/about-us/. [Accessed 8 Maret 2017].

[21] Dublin Core, "Dublin Core Metadata Element Set, Version 1.1," Dublin Core, 14 Juni 2012. [Online].

Available: http://dublincore.org/documents/dces/. [Accessed 8 Maret 2017].

[22] PADI, "ABOUT PADI DIVE CENTERS AND RESORTS," PADI, [Online]. Available: https://www.padi.com/about-padi/about-padi-dive-centers-and-resorts. [Accessed 4 Januari 2018].

(Halaman sengaja dikosongkan)

BIODATA PENULIS

Penulis lahir di Pekanbaru pada tanggal 12 Mei 1995. Merupakan anak kedua dari 3 bersaudara. Penulis telah menempuh beberapa pendidikan formal, yaitu; SD Cendana Rumbai, MTsN Pekanbaru, MAN 2 Model Pekanbaru.

Pada tahun 2013 pasca kelulusan SMA, penulis melanjutkan pendidikan dengan jalur mandiri di Jurusan Sistem Informasi Institut Teknologi Sepuluh Nopember (ITS) Surabaya dan terdaftar sebagai mahasiswa dengan NRP 5213100164.

Sebagai mahasiswa penulis aktif dalam urusan akademik, non akademik, organisasi maupun menjalankan bisnis. Tercatat pada tahun kedua penulis pernah menjadi staff pada Departemen Sosial Masyarakakt Himpunan Mahasiswa Sistem Informasi . Dan pada tahun kedua, penulis pernah menjadi Kepala Divisi Pengabdian Masyarakat Himpunan Mahasiswa Sistem Informasi dan staff Hubungan Masyarakat Lembaga Minat Bakat ITS. Selain itu penulis juga berkecimpung di dunia bisnis. Bisnis pertama yang sudah dijalani adalah Heroes Cloth, dan bisnis kedua yang hingga saat ini berjalan adalah Resto Pasta Kangen.

Penulis memiliki ketertarikan pada bidang pengolahan data, manajemen bisnis dan perusahaan, oleh karena itu penulis memutuskan untuk mengambil minat Sistem Enterprise dengan topik penelitian berkaitan dengan pengolahan data. Untuk kepentingan penelitian selanjutnya, penulis dapat dihubungi melalui email di *robbighf@gmail.com*

(Halaman sengaja dikosongkan)

LAMPIRAN A *SOURCE CODE* PENGAMBILAN LINK DAN KONTEN

```
<?php
// Library from intip.in/fw
require( 'csc_curl.php' ); //library crawl
require( 'csc_pdo.php' ); //library db

ini_set('max_execution_time', 0);

/* Database config*/
$_db = new stdClass;
$_db->name = 'cobalagi'; // YOUR DB name HERE
$_db->user = 'root'; // YOUR DB username HERE
$_db->pass = ''; // YOUR DB password HERE
$_db->pfx = ''; // YOUR DB prefix HERE
[optional]

$db = new csc_pdo( $_db->user, $_db->pass,
$_db->name );

 //$file = fopen("php://output", 'w');

// header('Content-Disposition: attachment;
filename="filename.csv";');

// Initiate class
$c = new curl;

function nyimpen( $url, $id, $n=0 ) {
      global $db, $c; //manggil db
      $domain1 = explode( '/', $url )[2];
//pemecah
      $c->bc = $c->get( $url );

      if ($url) $db->u( 'link_web', 'SET
Content=? WHERE link=?', array( strip_tags(
$c->bc ),$url  ) ); //pergantian flag menjadi
1
// Get text
      for ( $x=1; $x<=200;$x++){
          $link = $c->xp( 'href="', '"', $x
```

```
if ( $opt[0] != 'http:' && $opt[0] != 'https:' )
{
             //echo 'true';
             $domain2 = $domain1;
             $http = explode( '/', $url )[0];
             $link = $http . '//' . $domain1 .
'/' . ltrim( ltrim( $link, '.' ), '/' );
         } else $domain2 = null;

         if ( strpos( $link, str_replace(
['http://', 'https://', 'www'], '', $domain1 ) )
=== false) continue; //filter
         if ( strpos( $link, 'css' ) !== false)
continue; //filter
         if ( strpos( $link, '.js' ) !== false)
continue; //filter
         if ( strpos( $link, '/ko/' ) !== false)
continue; //filter
         if ( strpos( $link, '/nl/' ) !== false)
continue; //filter
         if ( strpos( $link, '/fr/' ) !== false)
continue; //filter
         if ( strpos( $link, '/de/' ) !== false)
continue; //filter
         if ( strpos( $link, '/it/' ) !== false)
continue; //filter
         if ( strpos( $link, '/jp/' ) !== false)
continue; //filter
         if ( strpos( $link, '/ru/' ) !== false)
continue; //filter
         if ( strpos( $link, '/es/' ) !== false)
continue; //filter
         if ( strpos( $link, '.ico' ) !== false)
continue; //filter
         if ( strpos( $link, '.jpg' ) !== false)
continue; //filter
         if ( strpos( $link, '.jpeg' ) !==
false) continue; //filter
         if ( strpos( $link, '.gif' ) !== false)
continue; //filter
         if ( strpos( $link, '.xml' ) !== false)
continue; //filter
         if ( strpos( $link, 'mailto' ) !==
false) continue; //filter
         if ( strpos( $link, '.png' ) !== false)
continue; //filter
```

```
if ( strpos( $link, 'json' ) !== false)
continue; //filter
         if ( strpos( $link, 'facebook' ) !==
false) continue; //filter
         if ( strpos( $link, '.pdf' ) !== false)
continue; //filter
         if ( strpos( $link, 'twitter' ) !==
false) continue; //filter
         if ( strpos( $link, 'google' ) !==
false) continue; //filter

         echo "\n $x $link $n \n<br>";
//menampilkan

         $f = $db->f( 'link_web', 'link', 'WHERE
link=?', $link );
         if ( ! $f ) $db->i('link_web',
'id_web,link,flag', array( $id, $link, 0 ) );
//menyimpan link

    }

    echo $url;
    $html =
preg_replace('/<script\b[^>]*>(.*?)<\/script>/is
', "", $c->bc); //mereplace content script
    $html =
preg_replace('/<style\b[^>]*>(.*?)<\/style>/is',
"", $html); //mereplace content css
    if ($url) $db->u( 'link_web', 'SET
flag=1,Content=? WHERE link=?', array(
strip_tags( $html ),$url ) ); //pergantian flag
menjadi 1 dan pengambilan konten
    $f = $db->f( 'link_web', '*', 'WHERE flag=0'
);
         if ($f) return nyimpen( $f->link, $id,
$n+1 ); //looping

}
$f = $db->f( 'website', '*', 'WHERE id_web=1' );
$url = $f->url; //sumber awal
nyimpen( trim($url),  $f->id_web  ); //trim
menghilangkan spasi atau enter

// Display text
```

(Halaman sengaja dikosongkan)

LAMPIRAN B *SOURCE CODE* PEMECAHAN KATA DAN FREKUENSI KATA

```
<?php
require( 'csc_curl.php' ); //library crawl
require( 'csc_pdo.php' ); //library db

$_db = new stdClass;
$_db->name = 'diveindustry'; // YOUR DB name
HERE
$_db->user = 'root'; // YOUR DB username HERE
$_db->pass = ''; // YOUR DB password HERE
$_db->pfx = ''; // YOUR DB prefix HERE
[optional]

$db = new csc_pdo( $_db->user, $_db->pass, $_db-
>name );

$diapus = [
         '(','.',',', ')',':','!','?','*','|','-
','/',

    '•','@','#','$','%','^','_','=','<','>',';'
         ];
$stopword =
['a','about','above','after','again','against','
all','am','an','and','any','are","aren't","as","
at','be','because','been','before','being','belo
w','between','both','but','by","can't","cannot",
'could',"couldn't","did","didn't","do","does","d
oesn't","doing","don't","down","during","each","
few","for","from","further","had","hadn't","has"
,"hasn't","have","haven't","having","he","he'd",
"he'll","he's","her","here","here's","hers","her
self","him","himself","his","how","how's","i","i
'd","i'll","i'm","i've","if","in","into","is","i
sn't","it","it's","its","itself","let's","me","m
ore","most","mustn't","my","myself","no","nor","
not","of","off","on","once","only","or","other",
'ought','our','ours','ourselves','out','over','o
wn','same',"shan't","she","she'd","she'll","she'
s","should","shouldn't","so","some","such","than
","that","that's","the","their","theirs","them",
'themselves','then','there',"there's",'these','t
```

```
];
$f = $db->r( 'link_web', '*', 'WHERE
flag_konten=0 and id_web=21 '  ); //get tabel
tujuan
foreach ( $f as $r ) { //looping, satu satu
   $r['Content'] = str_replace( "\n", ' ',
$r['Content'] ); //menghilangkan enter
   $r['Content'] = preg_replace('!\s+!', ' ',
$r['Content'] ); //menghilangkan dobel
spasimenggunakan requler expression
   $pecah = explode( ' ', $r['Content'] ); //get
yang mau dipecah
   foreach ( $pecah as $p) {
         $kata = str_replace( $diapus, '', $p );
//menghilangkan karakter

         $kata = trim( $kata ); //menghilangkan
enter dan spasi
         if ( ! $kata ) continue; //jika kosng,
lanjut

         foreach ($stopword as $sw) { //sw satu
persatu
                     if ( $kata == $sw ) continue
2;
         }
         echo $kata . '<br>';

         $cek = $db->f( 'link_keyword', '*',
'WHERE keyword=?', $kata ); //mengecek
         if ( ! $cek ) $db->i('link_keyword',
'keyword,frekuensi', array( $kata, 0 ) );
//menambahkan keyword
         else $db->u( 'link_keyword', 'SET
frekuensi=frekuensi+1 WHERE id_key=? ', [$cek-
>id_key] ); //jika sudah ada keyword, tambahkan
frekuensi
         $db->u( 'link_web', 'SET flag_konten=1
WHERE id_link=?', $r['id_link'] );
   }

//var_dump( $pecah );
}
```

LAMPIRAN C *SOURCE CODE* ANALISA KATA

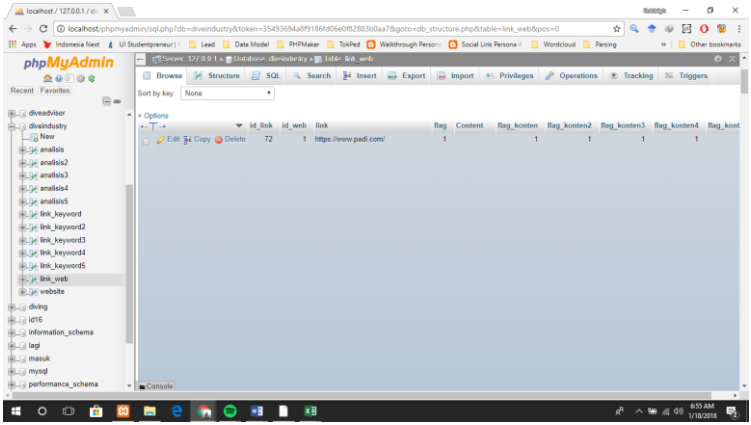
```
<?php
require( 'csc_curl.php' ); //library crawl
require( 'csc_pdo.php' ); //library db

set_time_limit(0);
//ini_set('memory_limit', '20000000000M');

$_db = new stdClass;
$_db->name = 'diveindustry'; // YOUR DB name
HERE
$_db->user = 'root'; // YOUR DB username HERE
$_db->pass = ''; // YOUR DB password HERE
$_db->pfx = ''; // YOUR DB prefix HERE
[optional]

$db = new csc_pdo( $_db->user, $_db->pass, $_db-
>name );

$f = $db->r( 'link_web', '*'  );

$fk = $db->r( 'link_keyword', '*', 'where
id_key=304


 ');


   foreach ($fk as $k){

foreach ($f as $x){

   $x['Content'] = str_replace( "\n", ' ',
$x['Content'] ); //menghilangkan enter
   $x['Content'] = preg_replace('!\s+!', ' ',
$x['Content'] ); //menghilangkan dobel
spasimenggunakan requler expression

   $kataLengkap = $x['Content'];
```

```
    $kataKunci = $k['keyword'];

    $kataSebelum = null;
    $kataSesudah = null;

    $kataExplode = explode($kataKunci,
$kataLengkap);

    if(count($kataExplode) >= 2)
    {
          if(!empty($kataExplode[0]))
          {
              $kataSebelum = $kataExplode[0];
          }
          if(!empty($kataExplode[1]))
          {
              $kataSesudah = $kataExplode[1];
          }
    }

    $explodeKataSebelum = explode(" ",
$kataSebelum);
    $kataSebelum = implode(" ",
array_slice($explodeKataSebelum, -7, 7,true));

    $explodeKataSesudah = explode(" ",
$kataSesudah);
    $kataSesudah = implode(" ",
array_splice($explodeKataSesudah, 0, 6));

    if(!empty($kataSesudah)){

    echo ''.$kataSebelum.' (' .$kataKunci.')
'.$kataSesudah.'<br>';

    $result = $kataSebelum.' ('.$kataKunci.')
'.$kataSesudah;
```

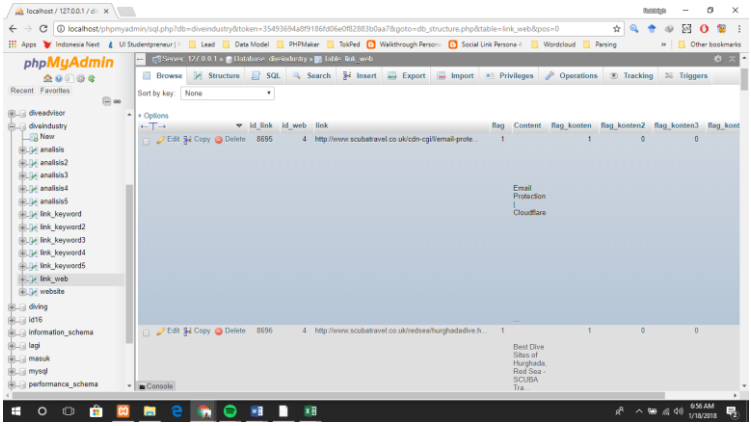
```
$cek = $db->f( 'analisis', '*', 'WHERE
analisis=?', $result ); //mengecek
         if ( ! $cek ) $db->i('analisis',
'id_key,id_link,analisis,frekuensi', array(
$k['id_key'], $x['id_link'],$result, 0 ) );
//menambahkan keyword
         else $db->u( 'analisis', 'SET
frekuensi=frekuensi+1  WHERE id_analisis=? ',
[$cek->id_analisis] );
         $db->u( 'link_keyword', 'SET
flag_analisis=1 WHERE id_key=?', $k['id_key']
);//jika sudah ada keyword, tambahkan frekuensi
         //$db->i('analisis', 'analisis', array(
$result) ); //menambahkan keyword
}
   }
}
?>
```

(Halaman sengaja dikosongkan)

LAMPIRAN D DATA WEB LINK DAN KONTEN

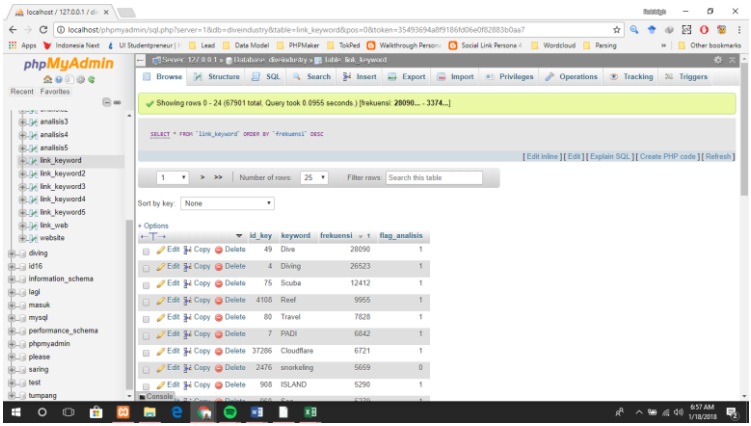


Gambar 11.1 Data Web Link dan Konten

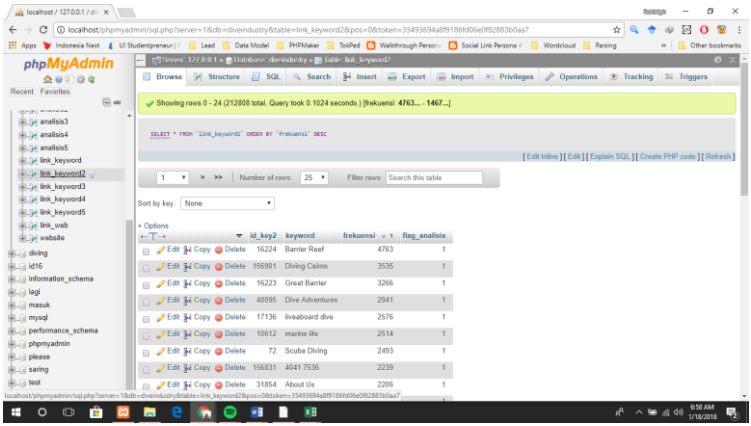


Gambar 11.2 Data Web Link dan Konten 2

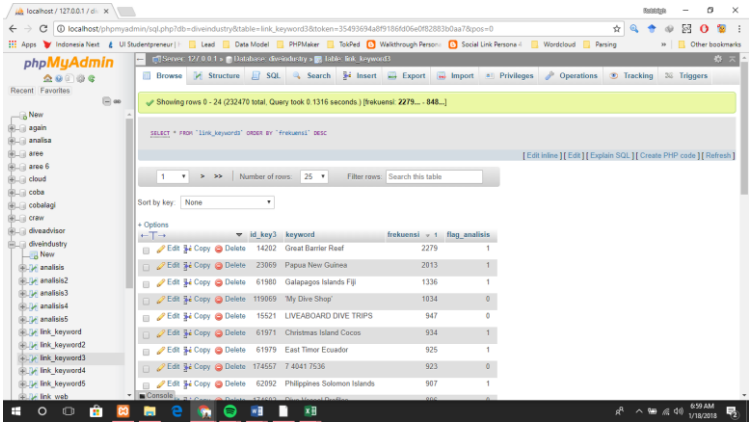
LAMPIRAN E DATA PEMECAHAN KATA



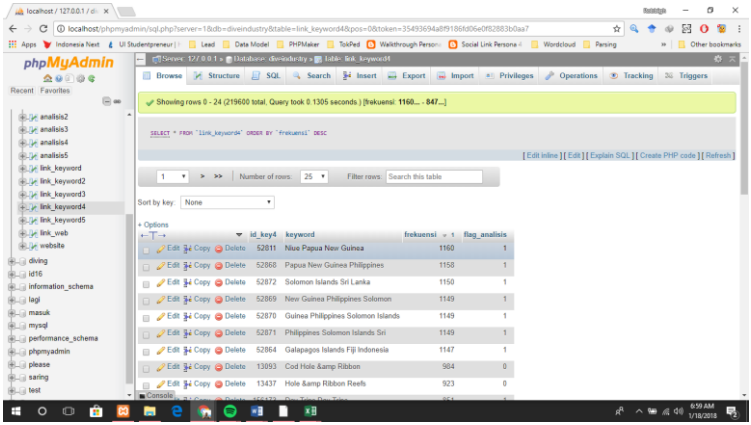
Gambar 12.1 Data Kata Kunci 1 Suku Kata



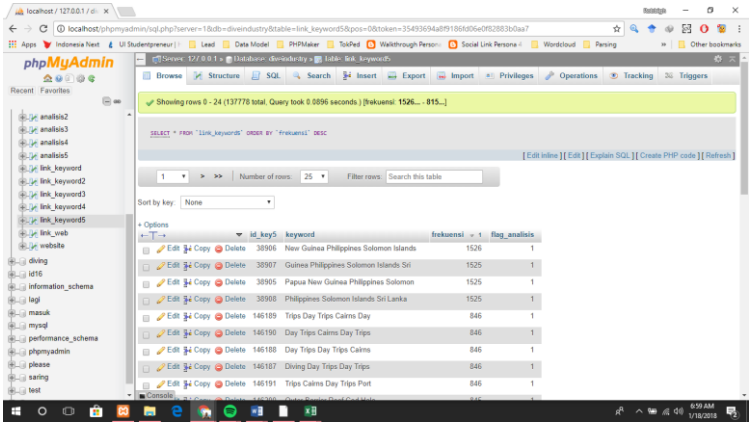
Gambar 12.2 Data Kata Kunci 2 Suku Kata



Gambar 12.3 Data Kata Kunci 3 Suku Kata

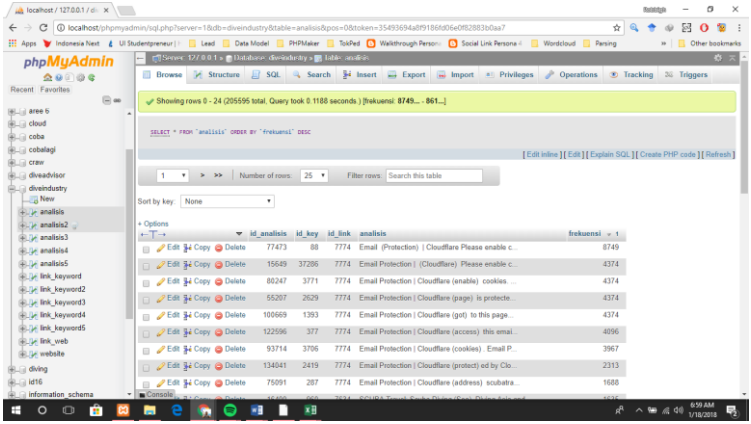


Gambar 12.4 Data Kata Kunci 4 Suku Kata

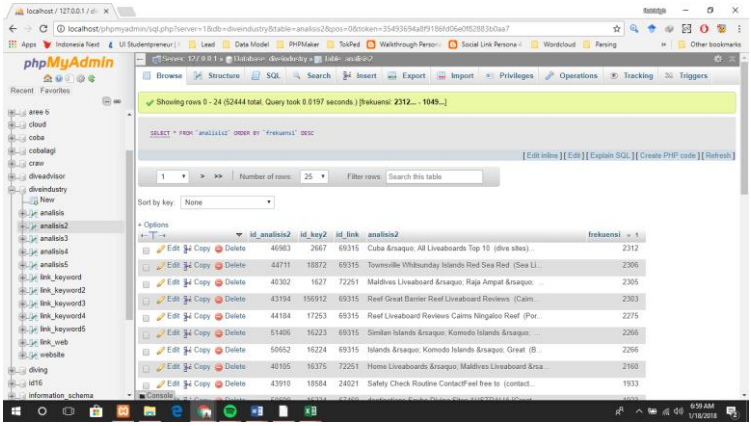


Gambar 12.5 Data Kata Kunci 5 Suku Kata

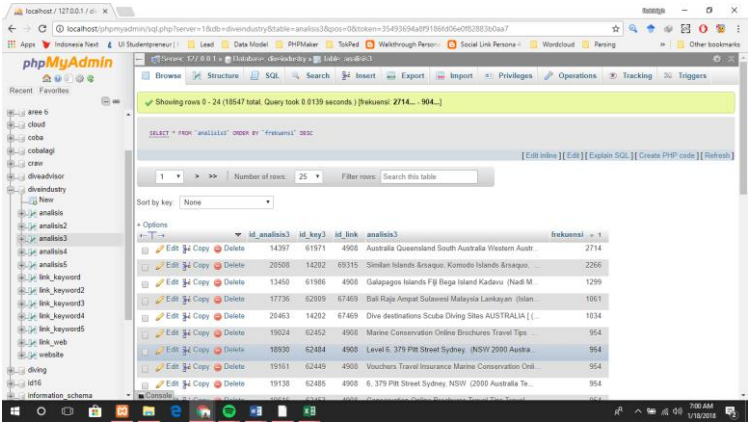
LAMPIRAN F DATA ANALISA KATA



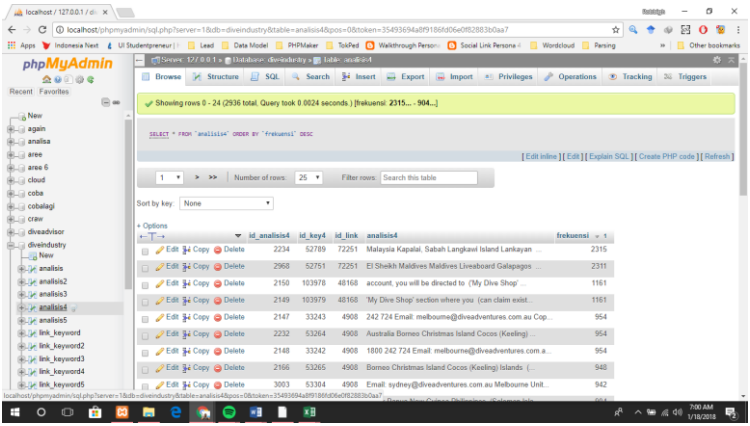
Gambar 13.1 Data Analisis 1 Suku Kata



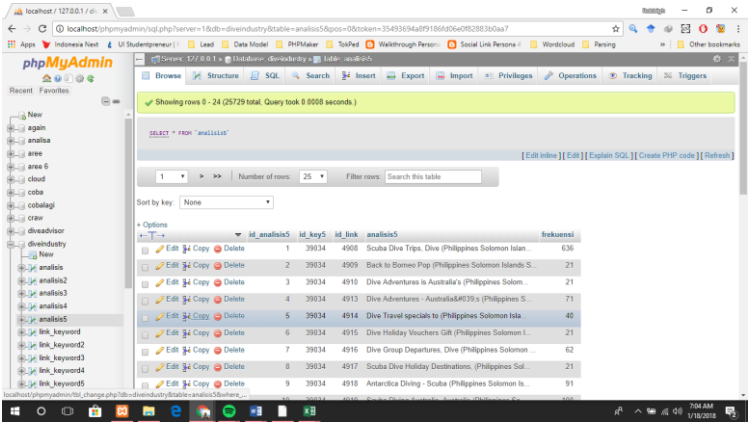
Gambar 13.2 Data Analisis 2 Suku Kata



Gambar 13.3 Data Analisis 3 Suku Kata



Gambar 13.4 Data Analisis 4 Suku Kata



Gambar 13.5 Data Analisis 5 Suku Kata