

**PENGEMBANGAN PENGELOMPOKAN KEILMUAN PENELITI JAWA  
TENGAH MENGGUNAKAN TEKNIK *SCRAPING* BERBASIS GOOGLE  
SCHOLAR**

**LAPORAN TUGAS AKHIR**

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana  
Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri  
Universitas Islam Sultan Agung Semarang



**DISUSUN OLEH :**

**GHUFRON DIPO SETIYONO**

**NIM 32601400921**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG**

**NOVEMBER 2021**

**FINAL PROJECT**

**DEVELOPMENT OF CENTRAL JAVA RESEARCH SCIENTIFIC  
GROUPING USING SCRAPING TECHNIQUES BASED ON GOOGLE  
SCHOLAR**

*Proposed to complete the requirement to obtain a bachelor's degree (S-1) at  
Informatics Engineering Departement of Industrial Technology Faculty Sultan  
Agung Islamic University*



**Arranged By:**

**GHUFRON DIPO SETIYONO**

**NIM 32601400921**

**MAJORING OF INFORMATICS ENGINEERING  
INDUSTRIAL TECHNOLOGY FACULTY  
SULTAN AGUNG ISLAMIC UNIVERSITY  
SEMARANG**

**NOVEMBER 2021**

## LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul **“PENGEMBANGAN PENGELOMPOKAN KEILMUAN PENELITI JAWA TENGAH MENGGUNAKAN TEKNIK SCRAPING BERBASIS GOOGLE SCHOLAR”** ini disusun oleh :

Nama : Ghufron Dipo Setiyono

NIM : 32601400921

Program Studi : Teknik Informatika

Telah disahkan oleh dosen pembimbing pada :

Hari : .....

Tanggal : .....

Mengesahkan,

Pembimbing I

Pembimbing II

  
Imam Much Ibnu Subroto, ST, M.Sc, Ph.D  
NIDN. 0613037301

  
Sam Farisa Chaery Hawiana, ST, M.Kom  
NIDN. 0628028602

Mengetahui,

Ketua Program Studi Teknik Informatika  
Fakultas Teknologi Industri  
Universitas Islam Sultan Agung



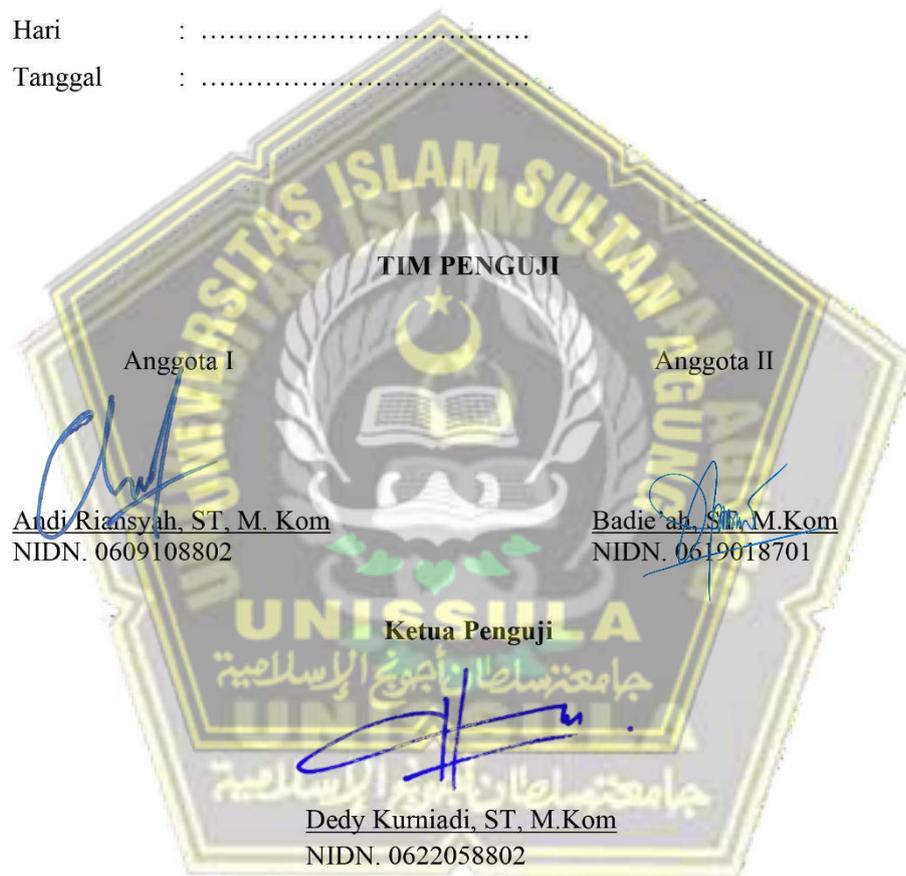
Andi Riarysyah, ST, M.Kom  
NIDN. 0609108802

## LEMBAR PENGESAHAN PENGUJI

Laporan tugas akhir dengan judul **“PENGEMBANGAN PENGELOMPOKAN KEILMUAN PENELITI JAWA TENGAH MENGGUNAKAN TEKNIK SCRAPING BERBASIS GOOGLE SCHOLAR”** ini telah dipertahankan di depan dosen penguji Tugas Akhir pada :

Hari : .....

Tanggal : .....



## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Ghufron Dipo Setiyono

NIM : 32601400921

Judul Tugas Akhir : **PENGEMBANGAN PENGELOMPOKAN KEILMUAN  
PENELITI JAWA TENGAH MENGGUNAKAN  
TEKNIK SCRAPING BERBASIS GOOGLE SCHOLAR**

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 30 November 2021

Yang Menyatakan,



Ghufron Dipo Setiyono

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Ghufron Dipo Setiyono

NIM : 32601400921

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : Sitangkil RT 01/08 Ds. Susukan Kec. Ungaran Timur Kab.  
Semarang

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :  
**PENGEMBANGAN PENGELOMPOKAN KEILMUAN PENELITI JAWA  
TENGAH MENGGUNAKAN TEKNIK SCRAPING BERBASIS GOOGLE  
SCHOLAR**

Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung.

Semarang, 30 November 2021

Yang menyatakan,



Ghufron Dipo Setiyono

## KATA PENGANTAR

Dengan mengucapkan puji syukur atas kehadiran Allah SWT berkat rahmatnya penulis dapat menyelesaikan Tugas Akhir dengan berjudul **Pengembangan Pengelompokan Keilmuan Peneliti Jawa Tengah Menggunakan Teknik Scraping Berbasis Google Scholar**. Penulis menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak baik moril maupun material. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada :

1. Ibunda dan Ayahanda yang telah banyak memberikan semangat, doa, keselamatan dan keberhasilan selama menempuh semua ujian.
2. Senior saya yang telah memberikan semangat, motivasi, doa dalam mengerjakan Tugas Akhir.
3. Bapak Imam Much Ibnu Subroto, ST, M.Sc., Ph.D selaku Dosen Pembimbing I yang telah meluangkan waktu dan memberikan ilmu kepada penulis.
4. Bapak Sam Farisa Chaerul Haviana, ST, M.Kom selaku Dosen Pembimbing II yang telah meluangkan waktu dan memberikan ilmu kepada penulis.
5. Para Dosen FTI Unissula yang telah memberikan banyak ilmu yang bermanfaat.

Dengan segala kerendahan hati, penulis menyadari masih banyak terdapat banyak kekurangan dari segi kualitas atau kuantitas maupun dari ilmu pengetahuan dalam menyusun laporan, sehingga penulis mengharapkan adanya saran dan kritikan bersifat membangun demi kesempurnaan laporan ini di masa mendatang.

Semarang, November 2021

Yang Menyatakan



Ghufron Dipo Setiyono

## DAFTAR ISI

LEMBAR PENGESAHAN PEMBIMBING .....	iii
LEMBAR PENGESAHAN PENGUJI .....	iv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR .....	v
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	x
DAFTAR GAMBAR .....	xi
ABSTRAK .....	xiv
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.2. Batasan Tugas Akhir .....	2
1.3. Tujuan Tugas Akhir .....	2
1.4. Manfaat .....	2
1.5. Sistematika Penulisan .....	2
<b>BAB II TINJAUAN PUSTAKA DAN DASAR TEORI .....</b>	<b>4</b>
2.1. Tinjauan Pustaka .....	4
2.2. Dasar Teori .....	5
2.2.1. Definisi <i>Web Scraping</i> .....	5
2.2.2. Pengembangan Sistem menggunakan <i>Waterfall</i> .....	7
<b>BAB III METODE PENELITIAN .....</b>	<b>8</b>
3.1. Flowchart Penelitian .....	8
3.2. Pengumpulan Data .....	9
3.3. Pengembangan Sistem <i>Scraping</i> .....	9
3.4. Pengelompokan Bidang Kepakaran .....	11
3.5. Perancangan Sistem Informasi Kepakaran Jawa Tengah .....	12
<b>BAB IV HASIL DAN ANALISIS PENELITIAN .....</b>	<b>20</b>
4.1. Pengumpulan Data Peneliti .....	20
4.2. Sistem <i>Scraping</i> Data Peneliti .....	24

4.3. Pengelompokan Kepakaran Peneliti.....	26
4.4. Implementasi Sistem.....	28
4.5. Pengujian Sistem.....	43
<b>BAB V KESIMPULAN DAN SARAN .....</b>	<b>45</b>
5.1. Kesimpulan .....	45
5.2. Saran .....	45
<b>DAFTAR PUSTAKA</b>	
<b>LAMPIRAN</b>	



## DAFTAR TABEL

Tabel 3.1 Rumpun Ilmu Level 1 .....	11
Tabel 4.1 Pengujian menu dan submenu halaman .....	43



## DAFTAR GAMBAR

Gambar 2.1	Metode <i>Waterfall</i> .....	7
Gambar 3.1	Diagram alir penelitian.....	8
Gambar 3.2	Proses <i>scraping</i> GS data peneliti.....	10
Gambar 3.3	Diagram Konteks.....	12
Gambar 3.4	DFD level 1 .....	13
Gambar 3.5	Rancangan menu Home .....	14
Gambar 3.6	Rancangan menu Peneliti .....	14
Gambar 3.7	Rancangan submenu Profil.....	15
Gambar 3.8	Rancangan menu Kepakaran .....	15
Gambar 3.9	Rancangan menu Bidang Fokus .....	16
Gambar 3.10	Rancangan menu Peta Kepakaran .....	17
Gambar 3.11	Rancangan <i>database</i> .....	18
Gambar 4.1	<i>sinta local DB</i> .....	20
Gambar 4.2	File SQL .....	20
Gambar 4.3	Serangkaian perintah <i>import data</i> MySQL .....	21
Gambar 4.4	Tabel Provinsi .....	21
Gambar 4.5	Tabel Kota .....	22
Gambar 4.6	Tabel Afiliasi.....	22
Gambar 4.7	Tabel Author .....	23
Gambar 4.8	Total peneliti di Provinsi Jawa Tengah .....	23
Gambar 4.9	Penyebaran peneliti menurut wilayah kabupaten / kota .....	24
Gambar 4.10	Eksekusi kode <i>scraping</i> .....	25

Gambar 4.11 Tabel Kepakaran .....	25
Gambar 4.12 Tabel Rumpun Ilmu DIKBUD.....	26
Gambar 4.13 Tabel Kepakaran 2 .....	27
Gambar 4.14 Tabel Peta Kepakaran.....	28
Gambar 4.15 Menu Home.....	28
Gambar 4.16 Menu Peneliti .....	29
Gambar 4.17 Contoh pencarian nama peneliti.....	29
Gambar 4.18 Contoh pencarian nama afiliasi.....	30
Gambar 4.19 Contoh pencarian nama kepakaran .....	30
Gambar 4.20 Menu Kepakaran .....	31
Gambar 4.21 Contoh pencarian nama kepakaran dengan Bahasa Indonesia...	31
Gambar 4.22 Contoh pencarian nama kepakaran dengan Bahasa Inggris.....	32
Gambar 4.23 Menu Bidang Fokus .....	32
Gambar 4.24 Kategori rumpun ilmu MIPA.....	33
Gambar 4.25 Kategori rumpun ilmu Tanaman.....	33
Gambar 4.26 Kategori rumpun ilmu Hewan.....	34
Gambar 4.27 Kategori rumpun ilmu Kedokteran .....	35
Gambar 4.28 Kategori rumpun ilmu Kesehatan .....	35
Gambar 4.29 Kategori rumpun ilmu Teknik.....	36
Gambar 4.30 Kategori rumpun ilmu Bahasa.....	36
Gambar 4.31 Kategori rumpun ilmu Ekonomi .....	37
Gambar 4.32 Kategori rumpun ilmu Sosial Humaniora .....	38
Gambar 4.33 Kategori rumpun ilmu Agama dan Filsafat.....	38
Gambar 4.34 Kategori rumpun ilmu Seni, Desain dan Media.....	39

Gambar 4.35 Kategori rumpun ilmu Pendidikan .....	39
Gambar 4.36 Menu Peta Kepakaran .....	40
Gambar 4.37 Jumlah peneliti tertinggi kota madya se Jawa Tengah.....	40
Gambar 4.38 Jumlah peneliti terendah kota madya se Jawa Tengah.....	41
Gambar 4.39 Jumlah peneliti tertinggi kabupaten se Jawa Tengah.....	42
Gambar 4.40 Jumlah peneliti terendah kabupaten se Jawa Tengah.....	42



## ABSTRAK

Perlunya pembangunan daerah menjadi perhatian saat ini. Pembangunan wilayah Jawa Tengah memerlukan para Ilmuan yang ahli di bidangnya dan Jawa Tengah sendiri belum mempunyai peta kepakaran ilmuan yang komprehensif yang terdiri dari para peneliti se Jawa Tengah. Salah satu permasalahan pertama yang dihadapi adalah bagaimana mengumpulkan data kepakaran penelitian secara menyeluruh Jawa Tengah. Beberapa sumber data dapat dijadikan sebagai referensi antara lain SINTA dan Google Scholar. Web *scraping* merupakan salah satu cara pendekatan untuk mengambil data dari halaman google scholar secara otomatis berbasis DOM yang telah dilakukan pada penelitian ini. Jumlah data terkumpul dalam *database* maka pengelompokan kepakaran bisa dilakukan. Hasil data yang di-*scraping* mencapai 20.973 peneliti Jawa Tengah dan dikelompokkan menjadi 7599 total bidang kepakaran. Setiap bidang kepakaran dikelompokkan lagi menurut 12 rumpun ilmu utama berdasarkan kategori yang dikeluarkan Kementerian Pendidikan dan Kebudayaan Indonesia.

Kata kunci: Web *Scraping*, Peneliti Jawa Tengah, SINTA, Google Scholar

## ABSTRACT

*The need for regional development is a concern at this time. The development of the Central Java region requires scientists who are experts in their fields and Central Java itself does not yet have a comprehensive map of scientific expertise consisting of researchers throughout Central Java. One of the first problems faced was how to collect comprehensive research expertise data in Central Java. Several data sources can be used as references, including SINTA and Google Scholar. Web scraping is one approach to retrieve data from Google undergraduate pages automatically based on the DOM that has been carried out in this research. The amount of data that is collected in the database, then the grouping of expertise can be done. The results of the scraped data reached 20,973 Central Java researchers and became 7599 total fields of expertise. Each field of expertise has been further grouped according to 12 main science clusters based on categories issued by the Indonesian Ministry of Education and Culture.*

*Keywords: Web Scraping, Central Java Researcher, SINTA, Google Scholar*

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Jawa tengah mempunyai luas wilayah 32.800 km<sup>2</sup> dan penduduk 34 juta yang termasuk dalam kategori padat penduduk. Dan semakin banyak pula sarana dan prasanara yang dibutuhkan masyarakat. Perlunya pembangunan daerah menjadi perhatian saat ini. Pembangunan wilayah Jawa Tengah memerlukan para Ilmuan yang ahli di bidangnya dan Jawa Tengah sendiri belum mempunyai peta kepakaran ilmuan yang komprehensif yang terdiri dari para peneliti se Jawa Tengah. Salah satu permasalahan pertama yang dihadapi adalah bagaimana mengumpulkan data kepakaran penelitian secara menyeluruh Jawa Tengah.

Beberapa penelitian sebelumnya oleh Nana Nerina Nasution yang berjudul “Sistem Pengumpulan Data Publikasi Ilmiah Menggunakan Metode Web Crawling” berhubungan dengan masalah pengumpulan data kepakaran untuk dapat membuat publikasi ilmiah yang dapat bermanfaat bagi pemecahan masalah di masyarakat. Permasalahan yang dihadapi adalah sulitnya mengumpulkan data publikasi dari civitas akademika secara manual. Metode *Web Crawling* menjadi salah satu cara yang dapat digunakan untuk mengotomasi pencarian dan ekstraksi data di dalam Website.

Perlunya pengembangan sistem informasi tentang kepakaran para ilmuan di Jawa Tengah dalam rangka mengoptimalkan potensi alam dan potensi para pakar yang dimiliki menjadi fokus dalam topik pembahasan kali ini. Untuk bisa mengumpulkan data para peneliti beserta kepakarannya memerlukan sumber data yang selalu update yang salah satunya dari internet diantaranya adalah Google Scholar atau Google Cendekia. Saat ini google scholar merupakan database peneliti paling besar dan paling cepat *up to date* di dunia sehingga bisa digunakan sebagai rujukan untuk kepakaran seseorang. Dengan mengumpulkan seluruh data tentang kepakaran inilah nanti yang akan menjadi target utama. Pengumpulan data dengan teknik

*scraping* merupakan salah satu cara pendekatan untuk mengambil data dari halaman google scholar secara otomatis yang akan dilakukan padapenelitian ini. Setelah data terkumpul dalam database maka pengelompokan kepakaran bisa dilakukan berdasarkan bidang ilmu dan berdasarkan wilayah (kabupaten / kota) di Jawa Tengah.

## 1.2 Rumusan Masalah

Bagaimana mengumpulkan data kepakaran para ilmuan di Jawa Tengah secara cepat dan menyeluruh menggunakan Teknik *scraping* serta bagaimana mengelompokkan kepakaran berdasarkan bidang ilmu peneliti.

## 1.3 Batasan Tugas Akhir

1. Data ilmuan/peneliti yang diolah hanya khusus Jawa Tengah
2. *Scraping* data hanya bersumber dari Google Scholar
3. Pengelompokan kepakaran dilakukan berdasarkan pengelompokan rumpun Ilmu berdasarkan Kemetrian RISTEKDIKTI yaitu meliputi MIPA, Tanaman, Hewan, Kedokteran, Kesehatan, Teknik, Bahasa, Ekonomi, Sosial Humaniora, Agama dan Filsafat, Seni, Desain dan Media, dan Pendidikan.

## 1.4 Tujuan Tugas Akhir

Penelitian ini bertujuan untuk :

1. Mengumpulkan data peneliti Jawa Tengah beserta kepakarannya dengan teknik *scraping*.
2. Membangun sistem informasi pengelompokan berdasarkan bidang ilmu / kepakaran Jawa Tengah.

## 1.5 Manfaat

Mendapatkan peta kepakaran para ilmuan di Jawa Tengah sehingga bisa menggali potensi sumber daya manusia (SDM) untuk pertimbangan pembangunan wilayah Jawa Tengah.

## 1.6 Sistematika Penulisan

### Bab I : Pendahuluan

Memuat tentang latar belakang masalah, alasan pemilihan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan Tugas Akhir.

### Bab II : Tinjauan Pustaka

Memuat penjelasan tentang konsep dan prinsip dasar yang diperlukan untuk memecahkan masalah Tugas Akhir.

### Bab III : Metode Penelitian

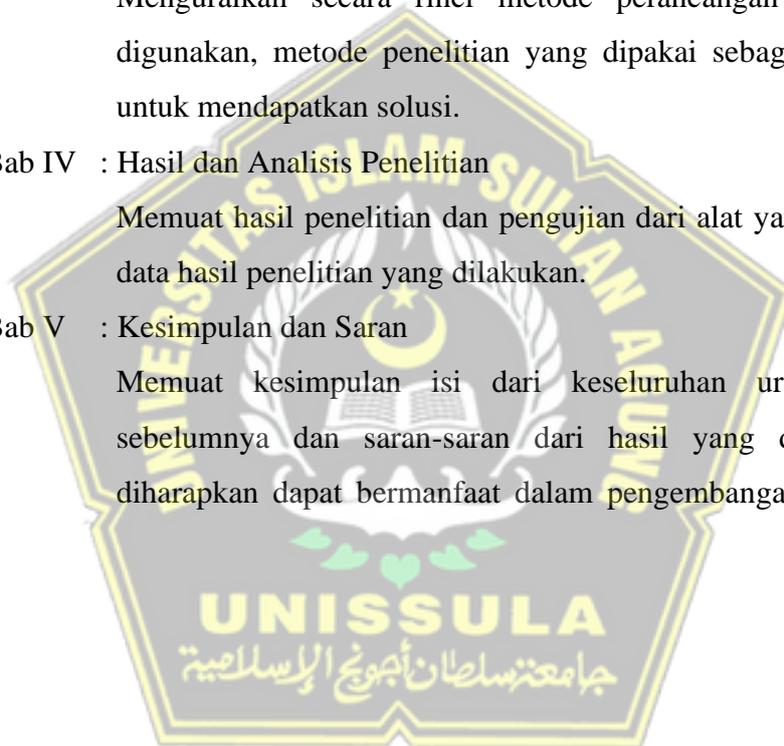
Menguraikan secara rinci metode perancangan sistem yang digunakan, metode penelitian yang dipakai sebagai pendekatan untuk mendapatkan solusi.

### Bab IV : Hasil dan Analisis Penelitian

Memuat hasil penelitian dan pengujian dari alat yang dibuat atau data hasil penelitian yang dilakukan.

### Bab V : Kesimpulan dan Saran

Memuat kesimpulan isi dari keseluruhan uraian bab-bab sebelumnya dan saran-saran dari hasil yang diperoleh dan diharapkan dapat bermanfaat dalam pengembangan selanjutnya.



## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Penelitian sebelumnya telah berhasil mengembangkan sistem yang menyajikan semua statistik penelitian personal menggunakan teknik *scraping* menggunakan URL Google Scholar menggunakan modul Python. (Pratiba et al., 2018)

Pengumpulan data dari Google Scholar telah dilakukan dengan menerapkan teknologi web *scraping* oleh peneliti lain. Eksperimen web *scraping* dari Google Scholar telah berhasil mengambil 238 data peneliti dan 2.523 file artikel. Data yang telah diunduh disimpan ke dalam basis data, kemudian digunakan untuk rekapitulasi publikasi artikel ilmiah, yang dapat menampilkan daftar profil peneliti, daftar afiliasi, daftar kutipan, dan daftar judul artikel. (Rahmatulloh & Gunawan, 2020)

Dalam implementasi aplikasi web *scraping* pada sejumlah portal *search engines* (Garuda, ISJD, Google Scholar), berhasil mengekstrak informasi mengenai artikel jurnal ilmiah dari sejumlah portal akademik baik yang berasal dari Indonesia maupun luar negeri. (Josi & Abdillah, 2014a)

Teknik *scraping* dapat dilakukan teknik regular ekspresi, regular ekspresi ditentukan pola yang mengawali dan mengakhiri suatu konten utama pada halaman situs. Salah satu situs yang mempunyai berbagai macam informasi yang akan digunakan sebagai obyek *scraping* adalah Wikipedia. Wikipedia sendiri adalah proyek ensiklopedia multibahasa dalam jaringan yang bebas dan terbuka. (Utomo, 2013)

Dengan melakukan implementasi Web *Scraping* dapat membentuk sebuah model pengumpulan data cuaca yang termonitoring dengan baik dan efisien. Identifikasi keadaan cuaca menjadi lebih akurat dengan memanfaatkan data yang besar dari hasil *scraping* dan dikombinasikan dengan K-Nearest Neighbor sehingga menghasilkan akurasi identifikasi yang baik. (Margaret, 2018)

## 2.2 Dasar Teori

### 2.2.1 Definisi Web Scraping

Menurut beberapa peneliti pengertian web *scraping* dapat didefinisikan sebagai berikut:

Web *scraping* atau disebut juga dengan web *harvesting* atau web *data extraction* merupakan sebuah teknik untuk mengambil informasi dari halaman website. Aplikasi web *scraping* juga disebut *intelligent, automated*, atau *autonomous agents* yang hanya memiliki fokus pada cara memperoleh data melalui pengenalan struktur semantik dan ekstraksi data dengan ukuran data yang bervariasi. (Josi & Abdillah, 2014b)

*Scraping* Web (juga disebut panen Web atau Web ekstraksi data) adalah sebuah perangkat lunak komputer teknik penggalian informasi dari situs web. Biasanya, program perangkat lunak tersebut mensimulasikan eksplorasi manusia dari Web salah satunya menerapkan Hypertext Transfer Protocol (HTTP), atau *embedding* Web browser tertentu seperti Internet Explorer (IE) dan Mozilla Web browser. (Utomo, 2013)

Web *scraping* merupakan teknik yang digunakan untuk mengekstrak sejumlah besar data dari situs web dimana data yang sudah diekstraksi disimpan ke sebuah file lokal di komputer atau ke database dalam format tabel (*spreadsheet*). Inilah yang memungkinkan user untuk mengeksplorasi isi dari situs web tanpa mengunjungi situs web yang bersangkutan, sehingga *user* bisa melakukan berbagai bentuk analisis tanpa mengganggu *resource* situs web yang bersangkutan. (Rizaldi & Putranto, 2017)

Ada beberapa teknik web *scraping* yang antara lain adalah:

#### 1. Parsing HTML

*Parsing* HTML adalah salah satu teknik yang paling banyak digunakan dalam web *scraping*. Biasanya *parsing* HTML dilakukan melalui JavaScript dan menarget halaman HTML *linear* dan *nested*. Teknik yang cepat ini mengidentifikasi *script* HTML dari website, yang mungkin saja dilakukan secara manual sebelumnya. *Script* ini kemudian digunakan untuk mengekstraksi *text*, *links*, dan *data*. (Patil et al., 2012)

## 2. Parsing DOM

Konten, *style*, dan struktur file XML didefinisikan dalam DOM, kependekan dari Document Object Model. *Scrapers* yang ingin mengetahui cara kerja internal halaman web dan mengekstrak skrip yang berjalan di dalamnya biasa memilih untuk melakukan web *scraping* melalui parsing DOM. *Node* spesifik dikumpulkan menggunakan *parser* DOM dan alat-alat seperti XPath membantu proses *scraping* sebuah halaman web. (Keith et al., 2005)

## 3. XPath

XML *Path Language* atau lebih dikenal dengan istilah XPath adalah bahasa query yang bekerja di dokumen XML. Karena dokumen XML biasa disusun dengan struktur pohon (*tree structure*), XPath bisa digunakan untuk menavigasi struktur dokumen tersebut dengan memilih *nodes* berdasarkan berbagai parameter. XPath juga bisa digunakan bersamaan dengan DOM *parsing* dalam mengesktrasi seluruh halaman website dan menampilkannya di website lain. (Blasco, 2007)

## 4. Google Docs

Google Sheets dapat digunakan untuk membantu web *scraping* dan ini adalah salah satu alat *scraping* yang populer. Di Google Sheets sendiri, bisa memanfaatkan fungsi *IMPORTXML* untuk melakukan *scraping* data dari website. Menggunakan *command* ini untuk melihat apakah website aman dari *scraping*. (Collins, 2015)

### *Apakah melakukan web scraping itu legal?*

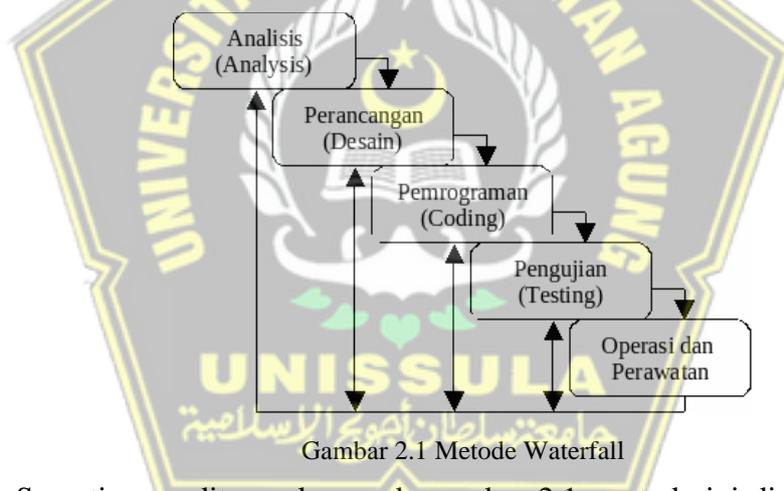
Sebelum melakukan web *scraping*, pelaku terlebih dahulu perlu mengetahui undang-undang spesifik di masing-masing Negara. Misalnya, pengguna harus berhati-hati dengan ketentuan hukum di situs web tertentu. Jika isinya berpemilik dan berorientasi pada tujuan komersial, *scraping* mungkin merupakan aktivitas ilegal. Dan berita dari surat kabar digital tidak dapat diekstraksi untuk diterbitkan di blog atau aplikasi tanpa menyebutkan sumbernya. Jika data yang akan discrap sudah jelas terbuka untuk umum maka *scrap* dianggap legal. (Castrillo-Fernández, 2015)

Perbedaan *Web Scraping* dan *Web Crawling* :

*Web crawling* sebenarnya juga melakukan *web scraping*. Namun sebaliknya ketika melakukan *web scraping* belum tentu melakukan *web crawling*. Kemudian untuk perbedaannya *crawling* biasanya digunakan untuk data dengan skala besar. Penerapannya menggunakan bot secara otomatis dan menggunakan *Application Programming Interface (API)*. Sedangkan *scraping* biasanya digunakan untuk data yang relative tidak terlalu besar dan proses pengambilan data pada *element HTML* ataupun *XML* menggunakan *protocol HTTP*. (vanden Broucke & Baesens, 2018)

### 2.2.2 Pengembangan Sistem Menggunakan *Waterfall*

Ada beberapa metode yang dapat digunakan para *developer* dalam membangun sistem salah satunya adalah metode *waterfall*



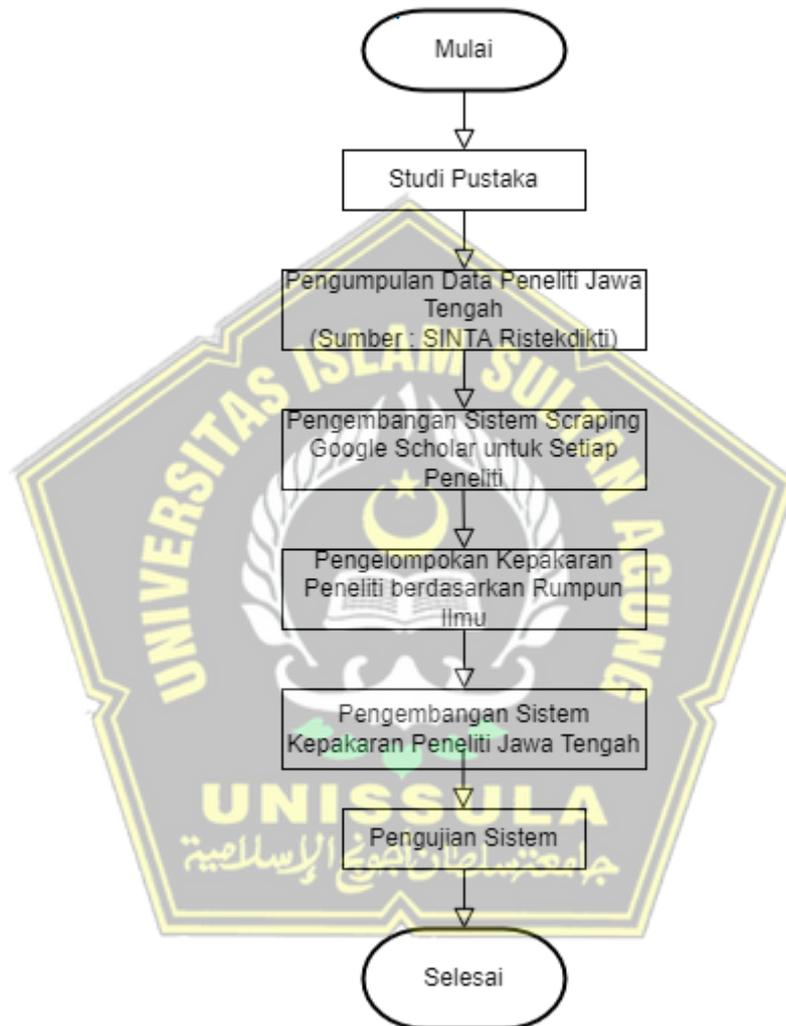
Seperti yang diterangkan pada gambar 2.1, metode ini digunakan untuk membagi tahap-tahap pengembangan sistem menjadi lima tahapan yaitu tahap Analisis, Perancangan, Implementasi, Pengujian, dan Pemeliharaan.

### BAB III

## METODE PENELITIAN

Pada Bab ini akan membahas tahapan yang dilakukan pada penelitian tugas akhir penulis dimulai dari tahap pengumpulan data dan perancangan sistem.

### 3.1 Alur Penelitian



Gambar 3.1 Diagram alir penelitian

Penjelasan pada gambar 3.1, penelitian diawali dengan studi pustaka dan pencarian bahan referensi untuk mencari cara yang efektif untuk mengumpulkan data peneliti serta bagaimana mengelompokkannya. Tahap selanjutnya adalah pengumpulan data para peneliti di Provinsi Jawa Tengah menggunakan data SINTA Ristekdikti. Langkah berikutnya mendapatkan data

tambahan dari Google Cendikia dengan mengembangkan sistem untuk *scrap* data setiap peneliti Jawa Tengah. Setelah mendapatkan data yang lengkap, data peneliti dikelompokkan berdasarkan dengan rumpun ilmunya sesuai pembagian rumpun ilmu dari Dikti. Data peneliti yang telah dikumpulkan dan disusun tadi akan digunakan untuk pengembangan sistem informasi kepakaran peneliti Jawa Tengah. Tahap terakhir adalah pengujian sistem dimana sistem informasi yang sudah dibuat akan diuji pada sisi fungsionalitas.

### 3.2 Pengumpulan Data

Penulis menggunakan SINTA sebagai sumber data peneliti karena disamping populer, penulis percaya bahwa data yang tercatat merupakan data yang *valid*. Saat ini SINTA merupakan *database* peneliti terbesar dan terlengkap di Indonesia. Dan karena yang dibutuhkan hanya wilayah Provinsi Jawa Tengah maka data akan diseleksi sesuai dengan peneliti khusus di Provinsi Jawa Tengah berdasarkan catatan SINTA.

### 3.3 Pengembangan Sistem Scraping

#### a. Menggunakan Bahasa Pemrograman PHP

Penulis menggunakan bahasa pemrograman PHP untuk sistem *scraping* dan untuk melakukannya di perlukan pengetahuan *basic* tentang pemrograman PHP dan beberapa tool pendukung web *scraping*.

#### b. Scraping Berbasis DOM HTML

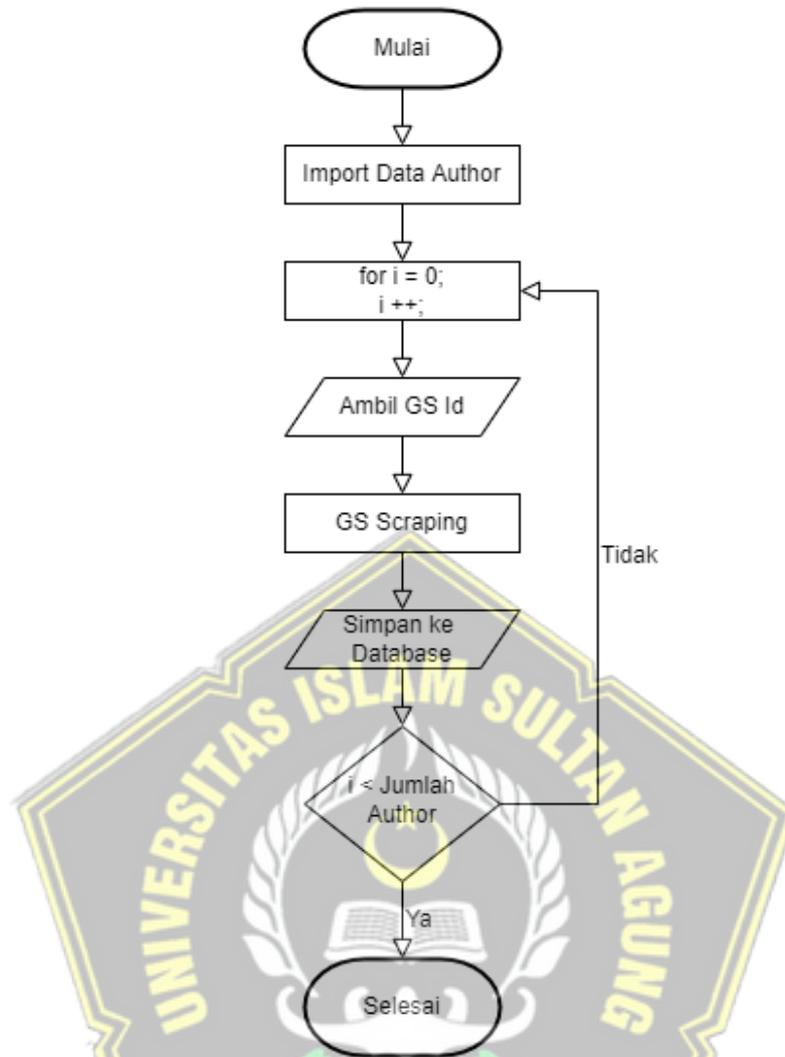
Teknik yang digunakan kali ini adalah *parsing* DOM dimana penulis akan menyeleksi *node* spesifik HTML yang memuat informasi tentang peneliti khusus Jawa Tengah melalui Google Cendikia.

#### c. Hasil Scraping disimpan di Database (DB)

Hasil *scraping* data peneliti pada google cendikia akan disimpan ke dalam MySQL DB yang kemudian akan dikelola lebih lanjut.

#### d. Data yang diambil dari Scraping Google Scholar (GS)

Data-data yang diambil meliputi gambar, nama peneliti, afiliasi dan *field area* / kepakaran.



Gambar 3.2 Proses *scraping* GS data peneliti

Apabila digabungkan proses pada poin bab 3.2 dan 3.3 akan diperoleh flowchart pada gambar 3.2. Data awal merupakan data peneliti yang di-*import* dari data SINTA ke dalam *database* berbasis MySQL menggunakan *console command line*. Kemudian data tersebut dikelompokkan sesuai jumlah *author*/peneliti yang bertugas di Jawa Tengah. Untuk data kedua berasal dari data peneliti pada google scholar. Dalam proses mendapatkan data dari google scholar, diperlukan fungsi perulangan yang memungkinkan untuk pencarian data peneliti yang jumlahnya begitu besar dimulai dari area Jawa Tengah dengan menandai Id google scholar yang dimiliki oleh setiap *author* yang bertugas di Jawa Tengah. Data yang diperoleh dari pencarian tadi akan disimpan ke dalam database. Proses pencarian akan terus berlangsung hingga

jumlah *author* yang bertugas di Jawa Tengah terpenuhi dan fungsi akan keluar dengan sendirinya menandakan pencarian telah selesai.

### 3.4 Pengelompokan Bidang Kepakaran

Pengelompokan kepakaran peneliti Jawa Tengah pada penelitian ini mengacu pada klasifikasi rumpun ilmu menurut Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan Nasional 2012 pada level satu dan berikut tahapannya.

1. Analisa hasil *scraping*

Hasil dari *scraping* yang terdiri dari data peneliti Jawa Tengah dengan kepakarannya masing-masing dan setiap peneliti dapat memiliki lebih dari satu kepakaran.

2. *Grouping*

Mengelompokkan data peneliti sesuai kepakarannya

3. *Labeling*

Memberian kode pada *group* kepakaran dengan kode rumpun level satu

4. Rumpun ilmu kepakaran

Mengelompokkan *group* kepakaran sesuai dengan 12 nama rumpun ilmu menurut Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan Nasional 2012 seperti yang ditunjukkan pada tabel 3.1.

Tabel 3.1 Rumpun Ilmu Level 1

Rumpun Ilmu	Kode	Level
Matematika Dan Ilmu Pengetahuan Alam (MIPA)	100	1
Ilmu Tanaman	140	1
Ilmu Hewani	200	1
Ilmu Kedokteran	260	1
Ilmu Kesehatan	340	1
Ilmu Teknik	410	1
Ilmu Bahasa	500	1
Ilmu Ekonomi	550	1
Ilmu Sosial Humaniora	580	1
Ilmu Agama dan Filsafat	630	1

Ilmu Seni, Desain dan Media	660	1
Ilmu Pendidikan	710	1

### 3.5 Perancangan Sistem Informasi Kepakaran Jawa Tengah

#### 1. Analisa Kebutuhan

Adapun beberapa alat yang diperlukan untuk menunjang perancangan sistem adalah :

##### a. Web Server

Karena sistem Informasi berbasis web maka diperlukan web server untuk melayani *user*. Penulis menggunakan web server Apache (Xampp)

##### b. Database

Untuk penyimpanan data, penulis menggunakan MySQL (Maria DB)

##### c. Web Browser (Chrome)

##### d. Internet

#### 2. Perancangan Sistem

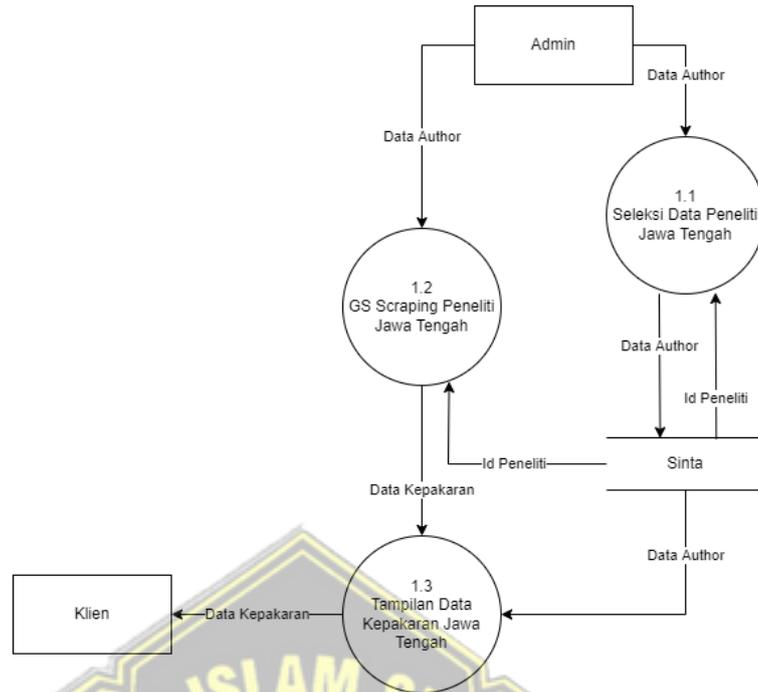
##### a. Diagram DFD



Gambar 3.3 Diagram Konteks

Gambar 3.3 merupakan diagram konteks Sistem Informasi Kepakaran Jawa Tengah dimana Admin menyediakan data kepakaran ke sistem dan diteruskan ke *user* dalam bentuk tampilan web.

##### b. Diagram DFD Level 1



Gambar 3.4 DFD level 1

Gambar 3.4 menerangkan alur data dimulai dari Admin yang membuat *coding* SQL untuk menyeleksi data peneliti (1.1) dan *coding scrap* PHP ke google scholar (1.2). Melalui *database* Sinta, admin menggunakan Id peneliti untuk dijadikan kunci utama dalam menandai data peneliti khusus Jawa Tengah yang kemudian diproses lebih lanjut untuk ditampilkan ke dalam web (1.3) dan dibaca oleh *user*.

### c. Perancangan *User Interface*

*User Interface* sangat memengaruhi daya tarik *user* dalam berselancar dimulai dari tampilan web yang menarik hingga kemudahan operasinya. Sistem Informasi Kepakaran Jawa Tengah memiliki lima bagian halaman utama dimulai dari home dan empat halaman utama lain yang memuat informasi tentang peneliti Jawa Tengah. Berikut *mock up* yang di desain oleh penulis.

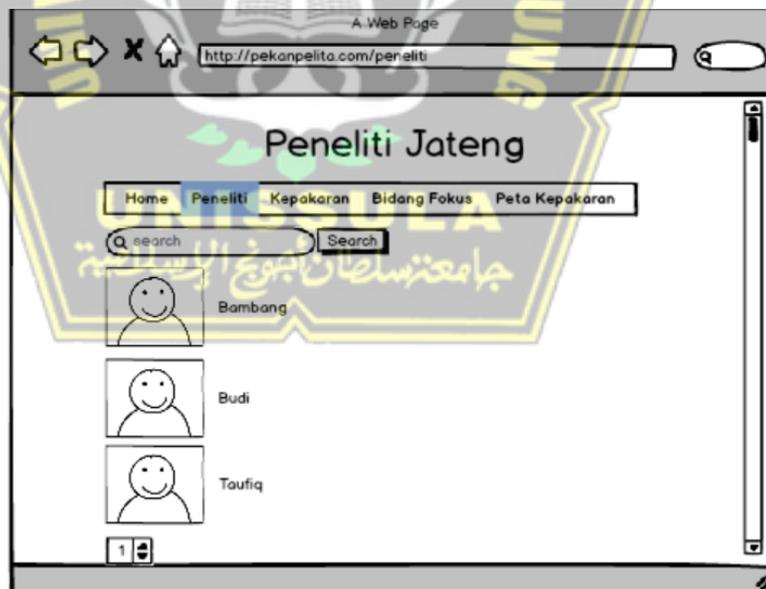
#### 1. Halaman Home



Gambar 3.5 Rancangan menu Home

Dilihat pada gambar 3.5, menu Home merupakan halaman awal dan memuat gambar *background* dan latar belakang perlu adanya sistem pengelompokan kepakaran peneliti

## 2. Halaman Peneliti



Gambar 3.6 Rancangan menu Peneliti

Seperti pada gambar 3.6, menu Peneliti merupakan halaman dimana gambar wajah peneliti beserta nama terangnya yang ditampilkan secara berurutan berdasarkan Id peneliti



Gambar 3.7 Rancangan submenu Profil

Ditinjau pada gambar 3.7, merupakan submenu Profil dimana apabila nama peneliti di klik akan menampilkan profil peneliti pada google scholar.

### 3. Halaman Kepakaran



Gambar 3.8 Rancangan menu Kepakaran

Rancangan pada gambar 3.8 adalah menu Kepakaran dimana menu ini memuat segala bentuk keilmuan atau kepakaran yang berada di Jawa Tengah

#### 4. Halaman Bidang Fokus



Gambar 3.9 Rancangan menu Bidang Fokus

Dari gambar 3.9 terdapat 12 rumpun ilmu yang tercantum di menu Bidang Fokus yaitu MIPA, Ilmu Tanaman, Ilmu Hewani, Ilmu Kedokteran, Ilmu Kesehatan, Ilmu Teknik, Ilmu Bahasa, Ilmu Ekonomi, Ilmu Sosial Humaniora, Ilmu Agama dan Filsafat, Ilmu Seni, Desain dan Media, dan Ilmu Pendidikan.

#### 5. Halaman Peta Kepakaran

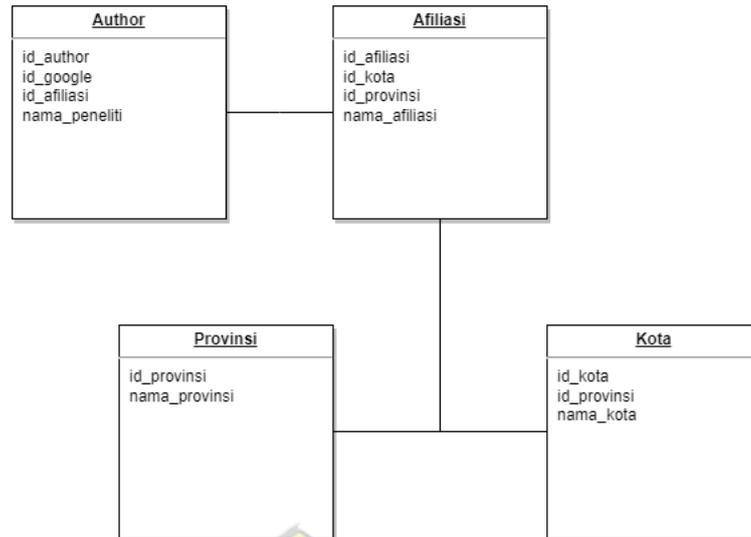


Gambar 3.10 Rancangan menu Peta Kepakaran

Menu Peta Kepakaran memberikan informasi tentang penyebaran lokasi para peneliti di Provinsi Jawa Tengah yang kurang lebih seperti yang dirancang pada gambar 3.10.

d. Perancangan *Database*

Sebelum memulai penyeleksian dan *scraping* terhadap data peneliti, penulis terlebih dahulu meng-*import data* SINTA ke dalam database MySQL yang memuat data peneliti di seluruh Indonesia yang di dalamnya terdapat empat tabel yang bernama afiliasi, author, kota, dan provinsi. Tabel tersebut berisi tentang informasi Id peneliti yang nanti akan digunakan lebih lanjut untuk menyeleksi data peneliti khusus Jawa Tengah. Berikut rancangan *database*-nya yang ditunjukkan pada gambar 3.11

Gambar 3.11 Rancangan *database*

### 3. Implementasi Sistem

#### a. Berbasis Web

Agar sistem *support*, penulis menggunakan web browser chrome versi 95. Sistem Informasi Kepakaran Jawa Tengah akan menggunakan desain web yang menarik bagi *user* dengan menambahkan *css template* yang disediakan oleh w3schools.com secara gratis.

#### b. Bahasa pemrograman PHP

Penulis menggunakan PHP versi 8 demi kelancaran eksekusi program. Pengetahuan *basic* tentang pemrograman PHP diperlukan karena kode program *scraping* akan disusun melalui bahasa pemrograman php untuk *function* dan penampilan web.

#### c. *Database MySQL*

Untuk keamanan dan kemudahan dalam mengolah data, penulis menggunakan *Database MySQL* mengingat data yang akan diolah sangat besar. Pengetahuan *basic* tentang pemrograman SQL dibutuhkan untuk menyusun kode program bahasa SQL yang akan digunakan untuk menyeleksi data yang dibutuhkan.

### 4. Pengujian/Testing

Tahap terakhir adalah pengujian sistem menggunakan Black Box *testing*. Black box adalah metode pengujian sistem yang hanya menguji sistem dari sisi fungsionalitasnya saja atau pengujian sistem dari luarnya saja tanpa perlu mengetahui isi kode program yang ada di dalam sistem.



## BAB IV

### HASIL DAN ANALISIS PENELITIAN

Pada Bab ini akan membahas hasil penelitian dari Bab III.

#### 4.1 Pengumpulan Data Peneliti

##### 1. *Import Data*

Langkah awal dalam mengumpulkan data peneliti adalah membuat *database* lokal terlebih dahulu untuk menampung data peneliti. Dan penulis memberi nama *database* tersebut dengan “sinta”.



Gambar 4.1 *sinta local DB*

DB *sinta* pada gambar 4.1 ini nanti akan digunakan untuk referensi pencarian identitas peneliti khusus Jawa Tengah setelah diisi dengan sumber data yang valid.

A screenshot of a file explorer window showing four SQL files. The files are listed in a table with columns for filename, date and time, file type, and size. The files are: afiliasi.sql (10,057 KB), author.sql (291,663 KB), kota.sql (21 KB), and provinsi.sql (3 KB).

afiliasi.sql	4/28/2021 10:16 AM	SQL File	10,057 KB
author.sql	4/28/2021 10:15 AM	SQL File	291,663 KB
kota.sql	4/28/2021 10:17 AM	SQL File	21 KB
provinsi.sql	4/28/2021 10:17 AM	SQL File	3 KB

Gambar 4.2 File SQL

Gambar 4.2 merupakan sumber data yang dimaksud dan sumber ini diperoleh dari SINTA Ristekdikti. Data ini berisi tentang data peneliti dari seluruh Indonesia. Terdiri dari empat file SQL.

Pada file “*afiliasi.sql*” berisi informasi tentang universitas / sekolah tinggi yang ada di Indonesia dan file “*author.sql*” berisi informasi tentang identitas para peneliti disetiap afiliasi di Indonesia. Pada file “*provinsi.sql*” memuat informasi tentang total Provinsi yang ada di Indonesia dan file “*kota.sql*” memuat informasi tentang jumlah kabupaten / kota disetiap Provinsi di Indonesia. Masing-masing tabel memiliki Id kunci utama / *Primary Key*.

Kemudian file-file tersebut akan di-*import* ke *database* lokal. Dan demi kelancaran, pemrosesan *import data* dikerjakan melalui *console* command prompt (CMD) karena data yang akan di-*import* cukup besar terutama file *author*.

```

MariaDB [sinta]> source c:\path\to\afiliasi.sql;
MariaDB [sinta]> source c:\path\to\author.sql;
MariaDB [sinta]> source c:\path\to\kota.sql;
MariaDB [sinta]> source c:\path\to\provinsi.sql;

```

Gambar 4.3 Serangkaian perintah *import data* MySQL

Setelah masuk ke dalam CMD, gunakan *database* *sinta* untuk target *import data*. Ada beberapa cara untuk meng-*import data* MySQL salah satunya menggunakan perintah *source* seperti gambar 4.3 dimana “*path\to*” diisi sesuai direktori penyimpanan empat file SQL tadi. *Import data* dilakukan satu persatu dan mungkin akan memakan waktu yang lama tergantung besarnya ukuran file SQL serta spesifikasi komputer yang digunakan.

## 2. Hasil *Import Data*

Setelah proses *import data* selesai, database akan berisi empat buah tabel yaitu tabel provinsi, kota, afiliasi dan author. Masing-masing tabel memiliki Id khusus yang akan menjadi kunci utama untuk menyeleksi data peneliti khusus wilayah Jawa Tengah. Id khusus tersebut meliputi id author dan id google pada tabel author, id afiliasi pada tabel afiliasi, id provinsi pada tabel provinsi, dan id kota pada tabel kota.

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
1	provinsi_id	int(11)		Tidak	Tidak ada	AUTO_INCREMENT			Ubah Hapus Lainnya
2	map_code	varchar(20)	latin1_swedish_ci	Ya	NULL				Ubah Hapus Lainnya
3	kode	varchar(20)	latin1_swedish_ci	Ya	NULL				Ubah Hapus Lainnya
4	nama_provinsi	varchar(255)	latin1_swedish_ci	Ya	NULL				Ubah Hapus Lainnya

Gambar 4.4 Tabel Provinsi

Tabel Provinsi ada gambar 4.4 memuat informasi tentang jumlah provinsi di Indonesia yang berjumlah 35 provinsi. Setiap provinsi memiliki id provinsi masing-masing.

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
1	kota_id	int(11)			Tidak	Tidak ada		AUTO_INCREMENT	Ubah Hapus Lainnya
2	provinsi_id	int(11)			Ya	NULL			Ubah Hapus Lainnya
3	kode	varchar(20)	latin1_swedish_ci		Ya	NULL			Ubah Hapus Lainnya
4	nama_kota	varchar(255)	latin1_swedish_ci		Ya	NULL			Ubah Hapus Lainnya

Gambar 4.5 Tabel Kota

Tabel Kota ada gambar 4.5 memuat informasi tentang jumlah kabupaten / kota yang di Indonesia yang berjumlah 515 kabupaten / kota. Setiap kabupaten / kota memiliki id kota dan id provinsi masing-masing.

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
1	afiliasi_id	int(11)			Tidak	Tidak ada		AUTO_INCREMENT	Ubah Hapus Lainnya
2	country	varchar(3)	latin1_swedish_ci		Tidak	ID			Ubah Hapus Lainnya
3	category	varchar(100)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
4	wilayah	int(11)			Tidak	Tidak ada			Ubah Hapus Lainnya
5	province	int(11)			Tidak	Tidak ada			Ubah Hapus Lainnya
6	city	int(11)			Tidak	Tidak ada			Ubah Hapus Lainnya
7	latitude	varchar(50)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
8	longitude	varchar(50)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
9	kode_pt	varchar(20)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
10	afiliasi_name	varchar(150)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
11	afiliasi_abbrev	varchar(50)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya

Gambar 4.6 Tabel Afiliasi

Tabel Afiliasi pada gambar 4.6 memuat informasi tentang jumlah afiliasi di Indonesia yang berjumlah 8153 afiliasi. Setiap afiliasi memiliki id afiliasi, id kota, dan id provinsi masing-masing.

#	Nama	Jenis	Penyortiran	Atribut	Tak Ternilai	Bawaan	Komentar	Ekstra	Tindakan
1	author_id	bigint(20)			Tidak	Tidak ada		AUTO_INCREMENT	Ubah Hapus Lainnya
2	afiliasi_id	int(11)			Tidak	0			Ubah Hapus Lainnya
3	country	varchar(3)	latin1_swedish_ci		Tidak	ID			Ubah Hapus Lainnya
4	NIDN	varchar(20)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
5	role	varchar(15)	latin1_swedish_ci		Tidak	Tidak ada	dosen atau peneliti selain dosen		Ubah Hapus Lainnya
6	idcard	varchar(25)	latin1_swedish_ci		Tidak	Tidak ada	no ktp		Ubah Hapus Lainnya
7	academic_grade	varchar(40)	latin1_swedish_ci		Tidak	Tidak ada	jabatan fungsional		Ubah Hapus Lainnya
8	jabatan_fungsional	varchar(50)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
9	prodi	varchar(100)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya
10	jenjang_prodi	varchar(25)	latin1_swedish_ci		Tidak	Tidak ada			Ubah Hapus Lainnya

Gambar 4.7 Tabel Author

Tabel Author ada gambar 4.7 memuat informasi tentang jumlah peneliti di Indonesia yang berjumlah 630669 peneliti. Setiap peneliti memiliki id author, id google, dan id afiliasi masing-masing.

Untuk penyeleksian data, penulis menggunakan pengetahuan tentang bahasa pemrograman SQL untuk menyusun sintaks SQL seperti contoh mengumpulkan data peneliti khusus Provinsi Jawa Tengah pada gambar 4.8

Menampilkan baris 0 - 24 (total 18198). Pencarian dilakukan dalam 0.0005 detik.

```
SELECT a.author_id, a.fullname, af.afiliasi_name FROM author a join afiliasi af on af.afiliasi_id = a.afiliasi_id WHERE af.province = 13;
```

author_id	fullname	afiliasi_name
1	ANIS MASHDUROHATUN	Universitas Islam Sultan Agung
14	AMIN PURNAWAN	Universitas Islam Sultan Agung
28	HERU SULISTYO	Universitas Islam Sultan Agung
31	WIDIYANTO	Universitas Islam Sultan Agung
39	MUHAMMAD QOMARUDDIN	Universitas Islam Sultan Agung
41	ANDRE SUGIYONO	Universitas Islam Sultan Agung
42	ARIEF MARWANTO	Universitas Islam Sultan Agung
44	IMAM MUCH IBNU SUBROTO	Universitas Islam Sultan Agung
45	ANTONIUS	Universitas Islam Sultan Agung
47	PRATIJKSO	Universitas Islam Sultan Agung
50	SLAMET IMAM WAHYUDI	Universitas Islam Sultan Agung
74	SRI HARTONO	Universitas Islam Sultan Agung
	NANIEK WIDYANINGRUM	Universitas Islam Sultan Agung

Gambar 4.8 Total peneliti di Provinsi Jawa Tengah

Terdapat sebanyak 18.198 total peneliti di Jawa Tengah beserta afiliasinya yang tercatat oleh SINTA dan berikut contohnya menurut wilayah kabupaten / kota seperti yang ditunjukkan pada gambar 4.9.

```
select a.author_id, a.fullname, a.google_id, af.afiliasi_name, af.city, k.nama_kota from author a JOIN afiliasi af ON af.afiliasi_id = a.afiliasi_id JOIN kota k ON k.kota_id = af.city where af.province = 13 order BY rand() limit 1,50;
```

Profil [ Edit dikotak ] [ Ubah ] [ Jelaskan SQL ] [ Buat kode PHP ] [ Segarkan ]

author_id	fullname	google_id	afiliasi_name	city	nama_kota
6650968	RIRIN FEBRIANA ANGGRAENI	dpTW008AAAAJ	Akademi Kesehatan Asih Husada	220	KOTA SEMARANG
6012491	AGUS SUPARDI	QubriGIAAAAAJ	Universitas Muhammadiyah Surakarta	198	KAB. SUKOHARJO
5975219	EVA YOVITA DWI UTAMI	HyBvKyYAAAAJ	Universitas Kristen Satya Wacana	219	KOTA SALATIGA
6662028	ARI SARWANTO	zEHX47QAAAAJ	Poltekkes Kemenkes Surakarta	218	KOTA SURAKARTA
6032479	AGUS SANTOSA	MBm9r9AAAAAJ	Universitas Muhammadiyah Purwokerto	189	KAB. BANYUMAS
6643636	DYAH CORRY KUSWARDANI	NbJMq0AAAAAJ	Universitas Semarang	220	KOTA SEMARANG
6654183	MAMUN MUMIN	sKKeT7QAAAAJ	INSTITUT AGAMA ISLAM NEGERI KUDUS	206	KAB. KUDUS
6707056	AHMAD MUSTOFA	emQBZYAAAAAJ	STAI Al-Husain Magelang (STI AlH)	195	KAB. MAGELANG
6673748	YUSNITA SYLVIA NINGRUM	b-xy9IsAAAAJ	Universitas Negeri Semarang	220	KOTA SEMARANG
6684156	SITI ROBIAH ADAWIYAH	wyVGTNQAAAAJ	Universitas Sains Al-Quran	194	KAB. WONOSOBO
6081022	SLAMET BUDI PRAYITNO	Owu_AnMAAAAAJ	Universitas Diponegoro	220	KOTA SEMARANG
6069446	ANDEKA ROCKY TANAAMAH	kc0Syt4AAAAJ	Universitas Kristen Satya Wacana	219	KOTA SALATIGA
6161411	MAULIDA DWI KARTIKASARI	EyhmCasAAAAJ	Universitas Pancasakti	222	KOTA TEGAL
6647438	LAHAN ADI PURWANTO	3oBd82YAAAAJ	Universitas Muhammadiyah Purwokerto	189	KAB. BANYUMAS
6657505	NUZUL WIDYAS	YJP-Hb8AAAAJ	Universitas Sebelas Maret	218	KOTA SURAKARTA
6022598	FAJAR ARI WIDIYATMOKO	_oFc5dkAAAAJ	Universitas PGRI Semarang	220	KOTA SEMARANG
6664468	SALIYO	s3yXq3IAAAAAJ	INSTITUT AGAMA ISLAM NEGERI KUDUS	206	KAB. KUDUS
6664468	TRISNINIK RATHI WILANDARI	EsnbHUIAAAAJ	Universitas Sebelas Maret	218	KOTA SURAKARTA

Gambar 4.9 Penyebaran peneliti menurut wilayah kabupaten / kota

Gambar 4.9 adalah contoh pencarian data peneliti di Jawa Tengah beserta sebaran wilayahnya secara acak.

## 4.2 Sistem Scraping Data Peneliti

### 1. Scraping dengan Code Program PHP

Untuk mempermudah proses *scraping*, penulis menggunakan dua *script*, *script* yang pertama bernama “simple\_html\_dom.php” sebagai referensi *function*. Untuk yang *script* kedua, penulis menggunakan *function curl* sebagai fungsi *scraping*-nya dalam pencarian data peneliti dari google scholar. File ini disimpan dengan nama “gs\_lib.php” (lihat lampiran). Kedua *script* tersebutlah yang akan digunakan untuk meringkas kode program.

Selanjutnya barulah penulis membuat program untuk mendapatkan data peneliti khusus area Jawa Tengah dengan menandai Id peneliti (id author dan id google) dan afliasinya (dengan id provinsi 13) dengan menyusun sintaks bahasa SQL. Kode provinsi 13 adalah kode provinsi milik Jawa Tengah pada DB sinta di tabel provinsi. Setiap peneliti akan dikategorikan menurut afliasinya yang berada di Provinsi Jawa Tengah. Selanjutnya barulah proses *scraping* data peneliti pada google scholar dilakukan satu persatu untuk diambil data kepakarannya. Data yang telah didapatkan akan disimpan ke dalam DB sinta dengan tabel baru yang bernama tabel Kepakaran. Dan dengan menggunakan fungsi perulangan, proses *scraping* akan terus berlanjut sampai seluruh data peneliti se Jawa Tengah terhitung. File ini disimpan

dengan nama “gs\_scrap\_pakar.php” (lihat lampiran) dan program ini yang akan dieksekusi.

## 2. Tabel Hasil *Scraping*

Agar lebih efektif, penulis menggunakan *console* command prompt untuk mengeksekusi program *scraping* tadi seperti berikut.

```
C:\>cd path\to\php gs_scrap_pakar.php
```

Gambar 4.10 Eksekusi kode *scraping*

Proses pada gambar 4.10 ini mungkin akan memakan waktu yang sangat lama tergantung dari koneksi internet dan spesifikasi komputer yang digunakan. Setelah proses *scraping* pada google scholar selesai, keluaran data tersebut berupa tabel yang berisi tentang informasi kepakaran yang ada di Jawa Tengah.

Menampilkan baris 0 - 24 (total 20973, Pencarian dilakukan dalam 0,0008 detik.)

SELECT \* FROM `kepakaran`

Profil [ Edit kotak ] [ Ubah ] [ Jelaskan SQL ] [ Buat kode PHP ] [ Segarkan ]

1 > >> | Jumlah baris: 25 | Saring baris: Cari di tabel ini

	kepakaran_id	author_id	kepakaran
<input type="checkbox"/> Ubah Salin Hapus	8704	260	Psikologi Pendidikan
<input type="checkbox"/> Ubah Salin Hapus	8705	261	Humaniora: Culture
<input type="checkbox"/> Ubah Salin Hapus	8706	261	society
<input type="checkbox"/> Ubah Salin Hapus	8707	261	and Edu-techno
<input type="checkbox"/> Ubah Salin Hapus	8708	265	Literary Studies
<input type="checkbox"/> Ubah Salin Hapus	8709	265	Cross Cultural Understanding
<input type="checkbox"/> Ubah Salin Hapus	8710	265	Language Teaching
<input type="checkbox"/> Ubah Salin Hapus	8711	271	sociology
<input type="checkbox"/> Ubah Salin Hapus	8712	271	culture

Gambar 4.11 Tabel Kepakaran

Hasil dari *data scraping* yang ditunjukkan pada gambar 4.11 berupa tabel Kepakaran yang di dalamnya memuat informasi id author beserta

kepakaran dan id kepakarannya yang didapatkan dari google scholar. Jumlah data peneliti yang diperoleh mencapai total 20.973 peneliti. Jumlah data peneliti yang didapatkan bertambah dari yang awalnya berjumlah 18.198 orang menjadi 20.973 orang karena setiap peneliti dapat memiliki lebih dari satu kepakaran sehingga menyebabkan redundansi data peneliti.

### 4.3 Pengelompokan Kepakaran Peneliti

Untuk proses pengelompokan data, penulis mengelompokkan nama kepakaran berdasarkan disiplin ilmunya dengan menggunakan acuan pada tabel rumpun ilmu dari Kemendikbud.

#### 1. Tabel Rumpun Ilmu

RUMPUN ILMU		
Kode	Rumpun	Level
100	MATEMATIKA DAN ILMU PENGETAHUAN ALAM (MIPA)	1
110	ILMU IPA	2
111	Fisika	3
112	Kimia	3
113	Biologi (dan Bioteknologi Umum)	3
114	Bidang Ipa Lain Yang Belum Tercantum	3
120	MATEMATIKA	2
121	Matematika	3

Gambar 4.12 Tabel Rumpun Ilmu DIKBUD

Ada total 425 rumpun ilmu beserta sub rumpunnya menurut Kementerian Pendidikan dan Kebudayaan yang masing masing memiliki level 1-3 seperti gambar 4.12. Level rumpun tersebut disusun dari rumpun utama yang memiliki level 1 dan sub rumpunnya yang diberi level 2 dan level 3 untuk sub rumpun dibawah level 2. Dan masing-masing rumpun memiliki kode tersendiri.

Penulis hanya menggunakan kode rumpun level 1 untuk pembagian setiap kepakaran peneliti. Jumlah rumpun level 1 terdiri dari 12 rumpun ilmu sesuai yang ditampilkan pada tabel 3.1 beserta kode rumpunnya.

## 2. Tabel Kepakaran Peneliti

Untuk pembagian rumpun ilmunya, penulis meng-*update* tabel kepakaran dengan menambahkan kode rumpun sebagai label dari setiap bidang ilmu / kepakaran peneliti.

			▼ kepakaran_id	author_id	kepakaran	kode_rumpun		
<input type="checkbox"/>	Ubah	Salin	Hapus	8704	260	Psikologi Pendidikan	710	
<input type="checkbox"/>	Ubah	Salin	Hapus	8705	261	Humaniora: Culture	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8706	261	society	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8707	261	and Edu-techno	710	
<input type="checkbox"/>	Ubah	Salin	Hapus	8708	265	Literary Studies	500	
<input type="checkbox"/>	Ubah	Salin	Hapus	8709	265	Cross Cultural Understanding	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8710	265	Language Teaching	710	
<input type="checkbox"/>	Ubah	Salin	Hapus	8711	271	sociology	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8712	271	culture	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8713	271	social humanities	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8714	274	Signal Processing	410	
<input type="checkbox"/>	Ubah	Salin	Hapus	8715	274	Adaptive System	410	
<input type="checkbox"/>	Ubah	Salin	Hapus	8716	275	internet of things	410	
<input type="checkbox"/>	Ubah	Salin	Hapus	8717	275	radio over fiber	410	
<input type="checkbox"/>	Ubah	Salin	Hapus	8718	275	digital television	660	
<input type="checkbox"/>	Ubah	Salin	Hapus	8719	275	smart city	580	
<input type="checkbox"/>	Ubah	Salin	Hapus	8720	275	digital creative industry	550	
<input type="checkbox"/>	Konsol	Ubah	Salin	Hapus	8721	279	Kecerdasan Buatan	410

Gambar 4.13 Tabel Kepakaran 2

Gambar 4.13 merupakan tampilan tabel kepakaran setelah di *update* dengan menambahkan kode rumpun. Dan data inilah yang nanti akan ditampilkan ke halaman Kepakaran.

## 3. Tabel Peta Kepakaran

Selanjutnya adalah mengelompokkan data peneliti berdasarkan sebaran wilayahnya di Provinsi Jawa Tengah menurut wilayah kabupaten / kota.

Menampilkan baris 0 - 24 (total 34, Pencarian dilakukan dalam 0,0667 detik.)

```
SELECT af.city,k.nama_kota, count(a.author_id) jumlah_peneliti FROM afiliasi af join author a on a.afiliasi_id = af.afiliasi_id join kota k on k.kota_id = af.city WHERE af.province = 13 GROUP by af.city;
```

Profil [ Edit dikotak ] [ Ubah ] [ Jelaskan SQL ] [ Buat kode PHP ] [ Segarkan ]

1 > >> |  Tampilkan semua | Jumlah baris: 25 | Saring baris: Cari di tabel ini

+ Opsi

city	nama_kota	jumlah_peneliti
188	KAB. CILACAP	244
189	KAB. BANYUMAS	2116
191	KAB. BANJARNEGARA	30
192	KAB. KEBUMEN	111
193	KAB. PURWOREJO	175
194	KAB. WONOSOBO	135
195	KAB. MAGELANG	61
196	KAB. BOYOLALI	52
197	KAB. KLATEN	185
198	KAB. SUKOHARJO	1166
199	KAB. WONOGIRI	38
200	KAB. KARANGANYAR	94
Konsol	SRAGEN	8

Gambar 4.14 Tabel Peta Kepakaran

Pada gambar 4.14 terdapat 34 total kabupaten / kota di Jawa Tengah beserta jumlah penelitiannya pada setiap kabupaten / kota. Data ini nanti akan digunakan untuk tampilan halaman Peta Kepakaran.

#### 4.4 Implementasi Sistem

Setelah seluruh data tersusun rapi, barulah penulis membangun antarmuka pengguna. Tampilan Sistem kepakaran Jawa Tengah yang dibangun oleh penulis dibuat dengan menggunakan CSS *template* dari w3schools.com untuk kecantikan penampilan web agar lebih menarik.

##### 1. Tampilan Halaman Home



Gambar 4.15 Menu Home

Menu halaman Home memiliki tampilan seperti gambar 4.15 dengan gambar *background* peta Jawa Tengah dan berisi tentang latar belakang dibuatnya sistem pengelompokan kepakaran di Jawa Tengah.

## 2. Tampilan Halaman Peneliti



The screenshot shows a web interface with a navigation bar at the top containing 'HOME', 'Peneliti', 'Kepakaran', 'Bidang Fokus', and 'Peta Kepakaran'. Below the navigation bar is a section titled 'Profil Peneliti'. It features a search bar with the placeholder text 'Cari Peneliti / Afiliasi / Kepakaran' and a 'Cari' button. Below the search bar is a table with the following data:

No	Id Author	Nama	Afiliasi	Kepakaran	Email
1	261	 DIDIK MURWANTONO	Universitas Islam Sultan Agung	Humaniora: Culture	didik_025@yahoo.com
			Universitas		

Gambar 4.16 Menu Peneliti

Gambar 4.16 merupakan tampilan menu halaman Peneliti yang berisi deretan profil peneliti se Jawa Tengah beserta id author, gambar, afiliasi, *field area* / kepakaran, dan emailnya. Setiap peneliti dapat memiliki lebih dari satu kepakaran. Terdapat fitur pencarian di atasnya untuk mencari nama peneliti / afiliasi / kepakaran yang ada di Jawa Tengah.

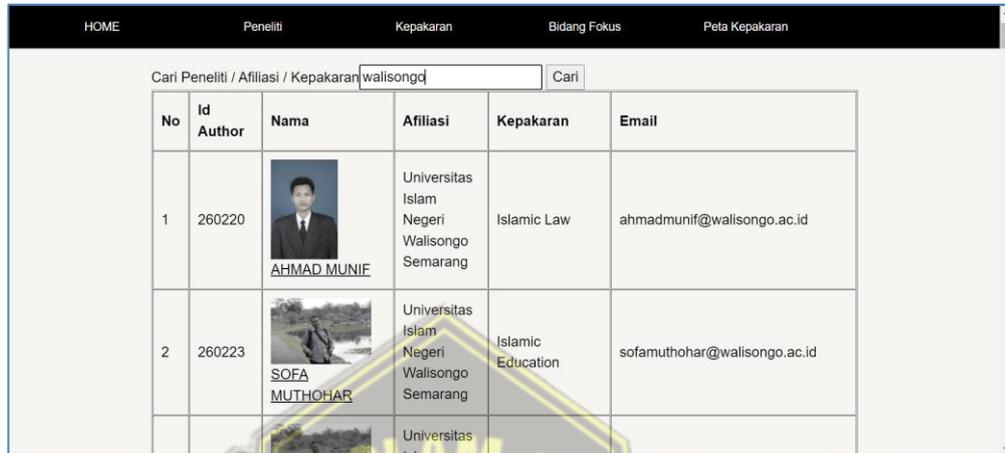


The screenshot shows the same web interface as Gambar 4.16, but with the search bar containing the name 'farhan'. The search results table is as follows:

No	Id Author	Nama	Afiliasi	Kepakaran	Email
1	6197161	 FARHAN SAEFUDIN WAHID	Universitas Muhadi Setiabudi	Ilmu Pendidikan	farhansaeudinwahid@gmail.com
2	6644558	 MOH. FARHAN	Universitas Islam Sultan Agung	Pendidikan Islam	moh.farhan@unissula.ac.id

Gambar 4.17 Contoh pencarian nama peneliti

Gambar 4.17 adalah contoh pencarian nama peneliti di Jawa Tengah yang memiliki nama “farhan” atau nama peneliti yang mengandung nama “farhan” di dalamnya.



No	Id Author	Nama	Afiliasi	Kepakaran	Email
1	260220	 AHMAD MUNIF	Universitas Islam Negeri Walisongo Semarang	Islamic Law	ahmadmunif@walisongo.ac.id
2	260223	 SOFA MUTHOHAR	Universitas Islam Negeri Walisongo Semarang	Islamic Education	sofamuthohar@walisongo.ac.id

Gambar 4.18 Contoh pencarian nama afiliasi

Gambar 4.18 adalah contoh pencarian nama afiliasi di Jawa Tengah yang memiliki nama “walisongo” atau nama afiliasi yang mengandung nama “walisongo” di dalamnya.



No	Id Author	Nama	Afiliasi	Kepakaran	Email
1	494	 MULAWARMAN	Universitas Negeri Semarang	Internet Addiction	mulawarman@mail.unnes.ac.id
2	8071	 ARI SRIYANTO NUGROHO	Politeknik Negeri Semarang	Internet of Things	ari.sriyanto@gmail.com

Gambar 4.19 Contoh pencarian nama kepakaran

Gambar 4.19 adalah contoh pencarian nama kepakaran di Jawa Tengah yang memiliki nama “internet” atau nama kepakaran yang mengandung nama “internet” di dalamnya.

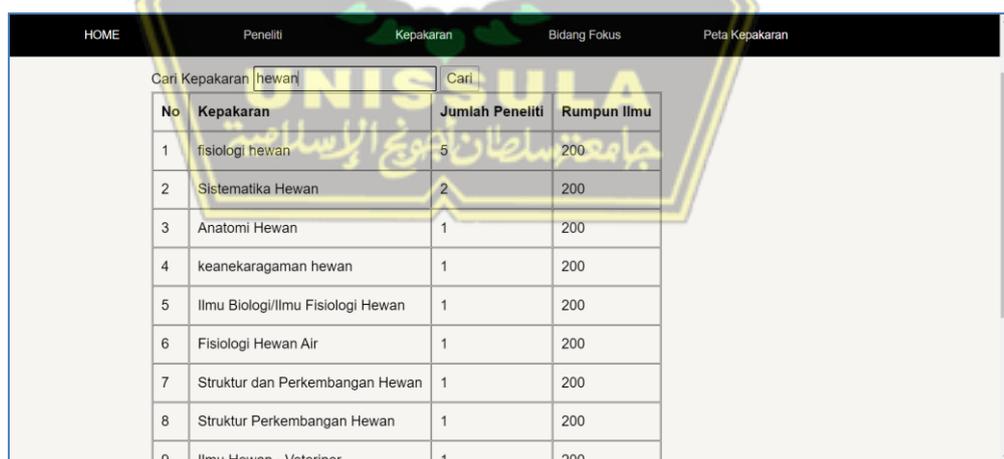
### 3. Tampilan Halaman Kepakaran



No	Kepakaran	Jumlah Peneliti	Rumpun Ilmu
1	Manajemen	325	550
2	kesehatan	236	340
3	keperawatan	210	340
4	Akuntansi	209	550
5	Kebidanan	181	340
6	education	164	710

Gambar 4.20 Menu Kepakaran

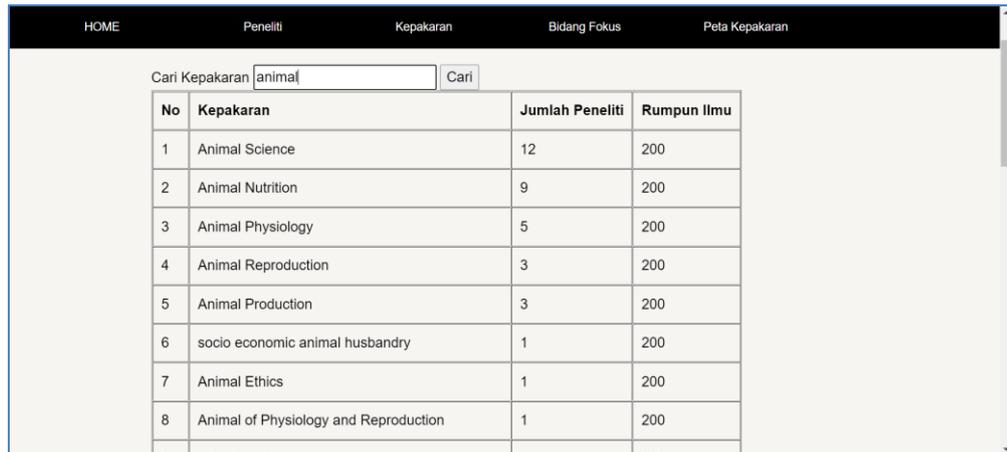
Tampilan menu halaman Kepakaran pada gambar 4.20 memuat informasi mengenai jenis-jenis keilmuan / kepakaran yang ada di Jawa Tengah beserta jumlah penelitiannya. Setiap kepakaran dikategorikan menurut rumpun ilmunya dengan kode rumpun level satu. Jumlah jenis bidang ilmu pada halaman ini mencapai 7559 bidang. Jumlah keilmuan di Jawa Tengah sangat banyak karena informasi kepakaran yang ditulis oleh *author* ada yang ditulis menggunakan ejaan bahasa selain Bahasa Indonesia. Untuk jumlah peneliti terbanyak adalah kategori ilmu manajemen yang mencapai 325 orang.



No	Kepakaran	Jumlah Peneliti	Rumpun Ilmu
1	fisiologi hewan	5	200
2	Sistematika Hewan	2	200
3	Anatomi Hewan	1	200
4	keanekaragaman hewan	1	200
5	Ilmu Biologi/Ilmu Fisiologi Hewan	1	200
6	Fisiologi Hewan Air	1	200
7	Struktur dan Perkembangan Hewan	1	200
8	Struktur Perkembangan Hewan	1	200
9	Ilmu Hewan - Veteriner	1	200

Gambar 4.21 Contoh pencarian nama kepakaran dengan Bahasa Indonesia

Gambar 4.21 merupakan tampilan pencarian nama kepakaran dengan menggunakan Bahasa Indonesia dengan kata kunci “hewan”.



No	Kepakaran	Jumlah Peneliti	Rumpun Ilmu
1	Animal Science	12	200
2	Animal Nutrition	9	200
3	Animal Physiology	5	200
4	Animal Reproduction	3	200
5	Animal Production	3	200
6	socio economic animal husbandry	1	200
7	Animal Ethics	1	200
8	Animal of Physiology and Reproduction	1	200

Gambar 4.22 Contoh pencarian nama kepakaran dengan Bahasa Inggris

Gambar 4.22 merupakan tampilan pencarian nama kepakaran dengan menggunakan Bahasa Inggris dengan kata kunci “animal”.

#### 4. Tampilan Halaman Bidang Fokus



No	Kode Rumpun	Rumpun Ilmu	Jumlah Peneliti
1	100	MATEMATIKA DAN ILMU PENGETAHUAN ALAM (MIPA)	1857
2	140	ILMU TANAMAN	822
3	200	ILMU HEWANI	382
4	260	ILMU KEDOKTERAN	465
5	340	ILMU KESEHATAN	2681
6	410	ILMU TEKNIK	3392
7	500	ILMU BAHASA	1021
8	550	ILMU EKONOMI	2954

Gambar 4.23 Menu Bidang Fokus

Gambar 4.23 memperlihatkan tampilan menu halaman Bidang Fokus. Pembagian bidang fokus peneliti dibagi menjadi 12 rumpun ilmu menurut tabel 3.1 beserta jumlah peneliti di setiap rumpun ilmu. Pada kolom Jumlah Peneliti memuat informasi jenis-jenis kepakaran mana yang termasuk dalam kategori ilmu MIPA, Tanaman, Hewan, Kedokteran, Kesehatan, Teknik, Bahasa, Ekonomi, Sosial Humaniora, Agama dan Filsafat, Seni, Desain dan Media, atau Pendidikan.

1854	DAN ILMU PENGETAHUAN ALAM (MIPA)	SLAMET IMAM WAHYUDI	water	Universitas Islam Sultan Agung
1855	MATEMATIKA DAN ILMU PENGETAHUAN ALAM (MIPA)	AZIZAH HIKMA SAFITRI	Biochemistry	Universitas Islam Sultan Agung
1856	MATEMATIKA DAN ILMU PENGETAHUAN ALAM (MIPA)	AZIZAH HIKMA SAFITRI	Biotechnology	Universitas Islam Sultan Agung
1857	MATEMATIKA DAN ILMU PENGETAHUAN ALAM (MIPA)	DINA FATMAWATI	cellular and molecular biology	Universitas Islam Sultan Agung

Gambar 4.24 Kategori rumpun ilmu MIPA

Gambar 4.24 adalah informasi kepakaran peneliti yang termasuk kategori ilmu MIPA. Jumlah peneliti pada kategori ilmu MIPA memiliki jumlah data sebanyak 1857 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

	TANAMAN			Semarang
817	ILMU TANAMAN	FAFA NURDYANSYAH	Pangan Fungsional	Universitas PGRI Semarang
818	ILMU TANAMAN	IFFAH MUFLIHATI	Teknologi Pangan	Universitas PGRI Semarang
819	ILMU TANAMAN	HASLINA	Teknologi Pangan	Universitas Semarang
820	ILMU TANAMAN	SRI BUDI WAHJUNINGSIH	Ilmu Pangan	Universitas Semarang
821	ILMU TANAMAN	BAMBANG DWILOKA	rekayasa pangan	Universitas Diponegoro
822	ILMU TANAMAN	SITI SUSANTI	Kimia dan Biokimia Pangan	Universitas Diponegoro

Gambar 4.25 Kategori rumpun ilmu Tanaman

Gambar 4.25 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Tanaman. Jumlah peneliti pada kategori ilmu Tanaman

memiliki jumlah data sebanyak 822 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

376	HEWANI	WIDODO	animal science	Soedirman
377	ILMU HEWANI	HERMAWAN SETYO WIDODO	dairy	Universitas Jenderal Soedirman
378	ILMU HEWANI	HERMAWAN SETYO WIDODO	milk protein	Universitas Jenderal Soedirman
379	ILMU HEWANI	HERMAWAN SETYO WIDODO	milk proteomics	Universitas Jenderal Soedirman
380	ILMU HEWANI	LUTOJO	Animal Production	Universitas Sebelas Maret
381	ILMU HEWANI	KASPRIJO	aquaculture	Universitas Jenderal Soedirman
382	ILMU HEWANI	FRIDA PURWANTI	Manajemen Sumber Daya Perairan	Universitas Diponegoro

Gambar 4.26 Kategori rumpun ilmu Hewan

Gambar 4.26 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Hewan. Jumlah peneliti pada kategori ilmu Hewan memiliki jumlah data sebanyak 382 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah. Jumlah pakar ilmu Hewan di Jawa Tengah adalah yang paling sedikit dari seluruh kategori rumpun ilmu.

460	ILMU KEDOKTERAN	SETIONO	neural network	Universitas Sebelas Maret
461	ILMU KEDOKTERAN	YUNIA HASTAMI	Anatomi	Universitas Sebelas Maret
462	ILMU KEDOKTERAN	SRI SULISTYOWATI	Spesialist Obstetri and Ginekologi	Universitas Sebelas Maret
463	ILMU KEDOKTERAN	AHMADI NUR HUDA	Ilmu penyakit jiwa	Universitas Islam Sultan Agung
464	ILMU KEDOKTERAN	ISMA YUNIAR	Critical Care	Sekolah Tinggi Ilmu Kesehatan Muhammadiyah Gombong
465	ILMU KEDOKTERAN	AZIZAH HIKMA SAFITRI	Degenerative Disorders	Universitas Islam Sultan Agung

Gambar 4.27 Kategori rumpun ilmu Kedokteran

Gambar 4.27 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Kedokteran. Jumlah peneliti pada kategori ilmu Kedokteran memiliki jumlah data sebanyak 465 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

2677	ILMU KESEHATAN	ATINA HUSSAANA	Medicine	Islam Sultan Agung
2678	ILMU KESEHATAN	ATINA HUSSAANA	Pharmacy	Universitas Islam Sultan Agung
2679	ILMU KESEHATAN	ATINA HUSSAANA	Traditional Medicine	Universitas Islam Sultan Agung
2680	ILMU KESEHATAN	INDAH NURAENI	Pangan Gizi Kesehatan	Universitas Jenderal Soedirman
2681	ILMU KESEHATAN	FAFA NURDYANSYAH	Mikrobiologi dan keamanan pangan	Universitas PGRI Semarang

Gambar 4.28 Kategori rumpun ilmu Kesehatan

Gambar 4.28 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Kesehatan. Jumlah peneliti pada kategori ilmu Kesehatan memiliki jumlah data sebanyak 2681 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

	TEKNIK			Sultan Agung
3387	ILMU TEKNIK	ARIEF MARWANTO	software defined radio	Universitas Islam Sultan Agung
3388	ILMU TEKNIK	IMAM MUCH IBNU SUBROTO	Artificial Intelligence	Universitas Islam Sultan Agung
3389	ILMU TEKNIK	ANTONIUS	Civil Engineering	Universitas Islam Sultan Agung
3390	ILMU TEKNIK	PRATIJKSO	Teknik Sipil	Universitas Islam Sultan Agung
3391	ILMU TEKNIK	SLAMET IMAM WAHYUDI	engineering	Universitas Islam Sultan Agung
3392	ILMU TEKNIK	AZIZAH HIKMA SAFITRI	Nutrients	Universitas Islam Sultan Agung

Gambar 4.29 Kategori rumpun ilmu Teknik

Gambar 4.29 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Teknik. Jumlah peneliti pada kategori ilmu Teknik memiliki jumlah data sebanyak 3392 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah. Jumlah pakar ilmu Teknik adalah yang paling banyak di Jawa Tengah dari seluruh kategori rumpun ilmu.

	BAHASA			Purwokerto
1016	ILMU BAHASA	NING YULIASTUTI	kajian bahasa	Universitas Sebelas Maret
1017	ILMU BAHASA	TEGUH KASPRABOWO	Sastra Inggris	Universitas Stikubank
1018	ILMU BAHASA	AGUS YULIANTORO	linguistik pragmatik	Universitas Widya Dharma
1019	ILMU BAHASA	SRI SUKASIH	Bahasa	Universitas Negeri Semarang
1020	ILMU BAHASA	DENIS EKA CAHYANI	natural language processing	Universitas Sebelas Maret
1021	ILMU BAHASA	DENIS EKA CAHYANI	information retrival	Universitas Sebelas Maret

Gambar 4.30 Kategori rumpun ilmu Bahasa

Gambar 4.30 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Bahasa. Jumlah peneliti pada kategori ilmu Bahasa memiliki jumlah data sebanyak 1021 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

2949	ILMU EKONOMI	IKA SEPTI ANGGRAENI	Manajemen Keuangan	Universitas Nahdlatul Ulama Surakarta
2950	ILMU EKONOMI	OKI KUNTARYANTO	Accounting	Universitas Widya Dharma
2951	ILMU EKONOMI	ANDRE SUGIYONO	Production System	Universitas Islam Sultan Agung
2952	ILMU EKONOMI	ANDRE SUGIYONO	Logistic Management	Universitas Islam Sultan Agung
2953	ILMU EKONOMI	SRI HARTONO	Feasibility Study and Finance	Universitas Islam Sultan Agung
2954	ILMU EKONOMI	OLIVIA FAHRUNNISA	Business Intelligence	Universitas Islam Sultan Agung

Gambar 4.31 Kategori rumpun ilmu Ekonomi

Gambar 4.31 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Ekonomi. Jumlah peneliti pada kategori ilmu Ekonomi memiliki jumlah data sebanyak 2954 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

2931	SOSIAL HUMANIORA	RUSENO ARJANGGI	adjustment	Universitas Islam Sultan Agung
2932	ILMU SOSIAL HUMANIORA	MUBAROK	Ilmu Komunikasi	Universitas Islam Sultan Agung
2933	ILMU SOSIAL HUMANIORA	OLIVIA FAHRUNNISA	Human Resource Management	Universitas Islam Sultan Agung
2934	ILMU SOSIAL HUMANIORA	OLIVIA FAHRUNNISA	Organizational Behavior	Universitas Islam Sultan Agung
2935	ILMU SOSIAL HUMANIORA	MARNO NUGROHO	HUMAN RESOURCE MANAGEMENT	Universitas Islam Sultan Agung

Gambar 4.32 Kategori rumpun ilmu Sosial Humaniora

Gambar 4.32 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Sosial Humaniora. Jumlah peneliti pada kategori ilmu Sosial Humaniora memiliki jumlah data sebanyak 2935 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

940	DAN FILSAFAT	ACHMAD MAIMUN	ISLAMIC STUDIES	IAIN Salatiga
941	AGAMA DAN FILSAFAT	ACHMAD MAIMUN	ISLAMIC RELIGIOUS EDUCATION	IAIN Salatiga
942	AGAMA DAN FILSAFAT	ACHMAD MAIMUN	PSYCHOLOGY OF RELIGION	IAIN Salatiga
943	AGAMA DAN FILSAFAT	MUH. HANIF	Pendidikan Agama Islam	Institut Agama Islam Negeri Purwokerto
944	AGAMA DAN FILSAFAT	MUH. HANIF	Studi Agama-Agama	Institut Agama Islam Negeri Purwokerto

Gambar 4.33 Kategori rumpun ilmu Agama dan Filsafat

Gambar 4.33 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Agama dan Filsafat. Jumlah peneliti pada kategori ilmu Agama

dan Filsafat memiliki jumlah data sebanyak 944 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

724	ILMU SENI, DESAIN DAN MEDIA	AJI WIYOKO	seni kriya logam	Institut Seni Indonesia Surakarta
725	ILMU SENI, DESAIN DAN MEDIA	NING YULIASTUTI	seni budaya	Universitas Sebelas Maret
726	ILMU SENI, DESAIN DAN MEDIA	DEVI AFRIYANTARI PUSPA PUTRI	game	Universitas Muhammadiyah Surakarta

Gambar 4.34 Kategori rumpun ilmu Seni, Desain dan Media

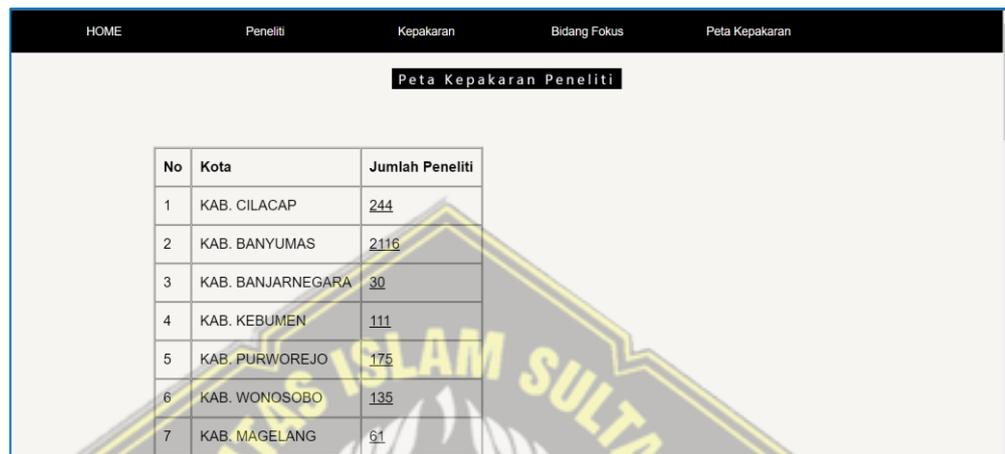
Gambar 4.34 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Seni, Desain dan Media. Jumlah peneliti pada kategori ilmu Seni, Desain dan Media memiliki jumlah data sebanyak 726 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

2776	ILMU PENDIDIKAN	IMAM MUCH IBNU SUBROTO	Machine Learning	Universitas Islam Sultan Agung
2777	ILMU PENDIDIKAN	IMAM MUCH IBNU SUBROTO	Education Technology	Universitas Islam Sultan Agung
2778	ILMU PENDIDIKAN	OLIVIA FAHRUNNISA	Knowledge Management	Universitas Islam Sultan Agung
2779	ILMU PENDIDIKAN	EVI CHAMALAH	Pendidikan Bahasa dan Sastra Indonesia	Universitas Islam Sultan Agung
2780	ILMU PENDIDIKAN	ENDANG LESTARI	Medical Education	Universitas Islam Sultan Agung

Gambar 4.35 Kategori rumpun ilmu Pendidikan

Gambar 4.35 adalah informasi kepakaran peneliti yang termasuk kategori ilmu Pendidikan. Jumlah peneliti pada kategori ilmu Pendidikan memiliki jumlah data sebanyak 2780 pakar beserta setiap nama peneliti dan afiliasinya di Jawa Tengah.

### 5. Tampilan Halaman Peta Kepakaran



No	Kota	Jumlah Peneliti
1	KAB. CILACAP	244
2	KAB. BANYUMAS	2116
3	KAB. BANJARNEGARA	30
4	KAB. KEBUMEN	111
5	KAB. PURWOREJO	175
6	KAB. WONOSOBO	135
7	KAB. MAGELANG	61

Gambar 4.36 Menu Peta Kepakaran

Menu Peta Kepakaran pada gambar 4.36 dibagi menjadi 34 kabupaten dan kota madya beserta jumlah peneliti di setiap kabupaten dan kota madya seperti yang ditunjukkan pada gambar 4.14. Pada kolom Jumlah Peneliti memuat informasi siapa saja peneliti.

	SEMARANG		Husada Semarang
7235	KOTA SEMARANG	SIGIT SUGIHARTO	Universitas Widya Husada Semarang
7236	KOTA SEMARANG	AGUNG SATRIO NUGROHO	Universitas Widya Husada Semarang
7237	KOTA SEMARANG	CHUSNUL ZULAIKA	Universitas Widya Husada Semarang
7238	KOTA SEMARANG	INTAN ANDRIANI	Universitas Widya Husada Semarang
7239	KOTA SEMARANG	NANIK SURANINGSIH	Universitas Widya Husada Semarang
7240	KOTA SEMARANG	IRAWAN WIBISONO	Universitas Widya Husada Semarang

Gambar 4.37 Jumlah peneliti tertinggi kota madya se Jawa Tengah

Gambar 4.37 adalah contoh sebaran peneliti di Jawa Tengah pada kota Semarang. Kota Semarang memiliki jumlah peneliti tertinggi kota madya sekaligus terbanyak di Provinsi Jawa Tengah yang hingga mencapai total 7240 orang.

	TEGAL		
301	KOTA TEGAL	HASBI ISKANDAR	Politeknik Baja Tegal
302	KOTA TEGAL	VITA NUR MILLATY	Politeknik Baja Tegal
303	KOTA TEGAL	ISMI KUSUMANINGROEM	Politeknik Baja Tegal
304	KOTA TEGAL	ATIEK NURINDRIANI	Politeknik Baja Tegal
305	KOTA TEGAL	SOFIYUDIN, M.PD	Politeknik Baja Tegal
306	KOTA TEGAL	SLAMET RIYADI	Politeknik Baja Tegal

Gambar 4.38 Jumlah peneliti terendah kota madya se Jawa Tengah

Gambar 4.38 adalah contoh sebaran peneliti di Jawa Tengah pada kota Tegal. Kota Tegal memiliki jumlah peneliti terendah kota madya yang hanya mencapai total 306 orang.

	BANYUMAS		
2111	KAB. BANYUMAS	FARIDATUN NIDA	Universitas Amikom Purwokerto
2112	KAB. BANYUMAS	EKO HIDAYATURROHMAN KHUMAENI	STIKES Ibnu Sina Ajibarang
2113	KAB. BANYUMAS	INDIRA PIPIT MIRANTI	STIKES Ibnu Sina Ajibarang
2114	KAB. BANYUMAS	WILDA PUTRI FESTIYANTI	STIKES Ibnu Sina Ajibarang
2115	KAB. BANYUMAS	NOVITA ENDANG FITRIYANI	STIKES Ibnu Sina Ajibarang
2116	KAB. BANYUMAS	IVA RINIA DEWI	STIKES Ibnu Sina Ajibarang

Gambar 4.39 Jumlah peneliti tertinggi kabupaten se Jawa Tengah

Gambar 4.39 adalah contoh sebaran peneliti di Jawa Tengah pada Kabupaten Banyumas. Kabupaten Banyumas memiliki jumlah peneliti tertinggi kabupaten yang hingga mencapai total 2116 orang.

No	Kota	Nama Peneliti	Afiliasi
1	KAB. BATANG	PRASETYAWAN AJI SUGIHARTO	Sekolah Tinggi Keguruan Dan Ilmu Pendidikan Muhammadiyah Batang
2	KAB. BATANG	MADYA GIRI ADITAMA	Sekolah Tinggi Keguruan Dan Ilmu Pendidikan Muhammadiyah Batang
3	KAB. BATANG	FAHMI SURYA ADIKARA	Sekolah Tinggi Keguruan Dan Ilmu Pendidikan Muhammadiyah Batang
4	KAB. BATANG	DWI CAHAYA NURANI	Sekolah Tinggi Keguruan Dan Ilmu Pendidikan Muhammadiyah Batang
5	KAB. BATANG	YANUAR AKHMAD	Sekolah Tinggi Keguruan Dan Ilmu Pendidikan Muhammadiyah Batang
6	KAB. BATANG	MAR'ATUL FAIDA	Sekolah Tinggi Keguruan Dan Ilmu Pendidikan Muhammadiyah Batang

Gambar 4.40 Jumlah peneliti terendah kabupaten se Jawa Tengah

Gambar 4.40 adalah contoh sebaran peneliti di Jawa Tengah pada Kabupaten Batang. Kabupaten Batang memiliki jumlah peneliti terendah

kabupaten sekaligus paling sedikit di Provinsi Jawa Tengah yang hanya mencapai total 6 orang.

#### 4.5 Pengujian Sistem

Untuk memahami fungsi-fungsi yang berada di dalam Sistem Kepakaran Jawa Tengah, penulis menggunakan *black box testing* sebagai media pengujian bahwa sistem sudah berjalan sesuai harapan atau tidak.

Tabel 4.1 Pengujian menu dan submenu halaman

<b>Aktifitas Pengujian</b>	<b>Ekspektasi</b>	<b>Hasil Uji</b>	<b>Kesimpulan</b>
Masuk disetiap menu halaman	Dapat menampilkan menu halaman yang di- <i>klik</i>	Halaman berganti sesuai menu yang dipilih	Valid
Pencarian nama peneliti / afiliasi / kepakaran pada menu Peneliti	Menampilkan nama peneliti / afiliasi / kepakaran sesuai nama yang diketik atau mendekati	Nama peneliti / afiliasi / kepakaran muncul seperti yang diketik pada box pencarian atau mendekati	Valid
Pencarian nama kepakaran pada menu Kepakaran	Menampilkan nama kepakaran sesuai nama yang diketik atau mendekati	Nama kepakaran muncul seperti yang diketik pada box pencarian atau mendekati	Valid
Meng- <i>klik</i> angka pada kolom Jumlah Peneliti dalam menu Bidang Fokus	Menampilkan Informasi tentang kepakaran peneliti sesuai rumpun ilmu yang dipilih	Informasi kepakaran peneliti dapat muncul sesuai yang dipilih	Valid

Meng- <i>klik</i> angka pada kolom Jumlah Peneliti dalam menu Peta Kepakaran	Menampilkan Informasi tentang lokasi peneliti sesuai kabupaten / kota yang dipilih	Informasi lokasi peneliti dapat muncul sesuai yang dipilih	Valid
--	--	--	-------



## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil pengkajian dari Bab I sampai Bab IV maka dapat disimpulkan:

1. Pengumpulan data peneliti Jawa Tengah beserta kepakarannya yang dilakukan dengan teknik *scraping* menggunakan *parsing* DOM dan *code* bahasa pemrograman PHP serta SQL telah berhasil mendapatkan 20.973 peneliti dengan masing-masing kepakarannya yang diperoleh dari google scholar / cendikia.
2. Pengelompokan berdasarkan bidang ilmu / kepakaran Jawa Tengah dapat dilakukan melalui pemberian kode rumpun pada setiap bidang ilmu peneliti dan dibagi menjadi 12 rumpun ilmu sesuai pembagian rumpun ilmu oleh Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

#### 5.2 Saran

Proses *import data* dan *scraping* yang dilakukan oleh penulis memakan waktu yang sangat lama hingga berjam-jam. Besarnya data yang diolah dan keterbatasan spesifikasi komputer menjadi faktor utama disini. Oleh karena itu penulis menyarankan satu hal dari dua solusi menurut penulis yaitu :

1. Menyediakan server yang mempunyai spesifikasi tinggi
2. Perlu diterapkan komputasi paralel sehingga waktu pemrosesan menjadi lebih efektif

## DAFTAR PUSTAKA

- Blasco, J. (2007). Introduction to xpath injection techniques. *Hakin9, Conference on IT Underground, Czech Republic*, 23–31.
- Castrillo-Fernández, O. (2015). Web *scraping*: applications and tools. *European Public Sector Information Platform*.
- Josi, A., & Abdillah, L. A. (2014a). Penerapan teknik web *scraping* pada mesin pencari artikel ilmiah. *ArXiv Preprint ArXiv:1410.5777*.
- Josi, A., & Abdillah, L. A. (2014b). Penerapan teknik web *scraping* pada mesin pencari artikel ilmiah. *ArXiv Preprint ArXiv:1410.5777*.
- Keith, J., Sambells, J., & DOM, S. (2005). Web Design with JavaScript and the Document Object Model. *Friends of ED Publishing*.
- Patil, P. C., Chawan, P. M., & Chauhan, P. M. (2012). Parsing of HTML Document. *IJAR CET In*.
- Pratiba, D., Abhay, M. S., Dua, A., Shanbhag, G. K., Bhandari, N., & SINGH, U. (2018). Web *Scraping* And Data Acquisition Using Google Scholar. *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, 277–281.
- Rahmatulloh, A., & Gunawan, R. (2020). Web *Scraping* with HTML DOM Method for Data Collection of Scientific Articles from Google Scholar. *Indonesian Journal of Information Systems*, 2(2), 95–104.
- Rizaldi, T., & Putranto, H. A. (2017). Perbandingan Metode Web *Scraping* Menggunakan CSS Selector dan Xpath Selector. *Teknika*, 6(1), 43–46.
- Teknik, M., Panjaitan, W. S., & Margaret, P. (n.d.). *Sistem Monitoring Cuaca dan Identifikasi Keadaan Cuaca*.  
<http://repositori.usu.ac.id/handle/123456789/4272>
- Utomo, M. S. (2013). Web *Scraping* pada Situs Wikipedia menggunakan Metode Ekspresi Regular. *Dinamik*, 18(2).
- vanden Broucke, S., & Baesens, B. (2018). From Web *Scraping* to Web Crawling. In *Practical Web Scraping for Data Science* (pp. 155–172). Springer.

## LAMPIRAN

### Coding pada file “gs\_lib.php”

```
<?php
function ambil_curl($url){
    $curl = curl_init($url);
    curl_setopt($curl, CURLOPT_SSL_VERIFYPEER, false);
    $dir = dirname(__FILE__);
    $config['cookie_file'] = $dir . '/cookies/' . md5($_SERVER['REMOTE_ADDR']) . '.txt';

    curl_setopt($curl, CURLOPT_COOKIEFILE, $config['cookie_file']);
    curl_setopt($curl, CURLOPT_COOKIEJAR, $config['cookie_file']);

    curl_setopt($curl, CURLOPT_FOLLOWLOCATION, true);
    curl_setopt($curl, CURLOPT_RETURNTRANSFER, true);

    $data = curl_exec($curl);
    curl_close($curl);
    return $data;
}

function google_profile($html){
    $html = str_get_html($html);
    //Biodata
    $pub["nama"] = trim($html->find('#gsc_prf_in div[]',0)->innertext);
    //print_r($nama);
    $pub["afiliasi"] = $html->find('div[class="gsc_prf_il"',0)->plaintext;
    for($i=0;$i<5;$i++){
        $bdg = $html->find('a[class="gsc_prf_inta gs_ibl"', $i)->plaintext;
        if(isset($bdg)) { $bid[] = $bdg; }
    }
    $pub["bidang"] = $bid;
    $pub["sitasi"] = $html->find('td[class="gsc_rsb_std"',0)->plaintext;
    $pub["h-index"] = $html->find('td[class="gsc_rsb_std"',2)->plaintext;
    $pub["i10"] = $html->find('td[class="gsc_rsb_std"',4)->plaintext;

    return $pub;
}
```

#001

#002

Penjelasan kode program :

Selain *library* simple\_html\_dom.php, penulis membuat *library* sendiri yang terdiri dari dua *function* yaitu ‘ambil\_curl’ dan ‘google\_profile’

- #001 : Merupakan serangkaian *function* ‘ambil\_curl’ yang digunakan untuk memindai data halaman html, fungsi CURL ini sendiri merupakan kependekan dari Client Url yang umumnya digunakan untuk melihat isi halaman web.
- #002 : Merupakan serangkaian *function* ‘google\_profile’ yang digunakan untuk menarget html pada halaman google scholar berupa profil peneliti untuk diambil data kepekarannya (maksimal 5 kepekararan).

## Coding pada file “gs\_scrap\_pakar.php”

```
$sql = "SELECT af.afiliasi_id,af.afiliasi_abbrev,af.afiliasi_name,author_id, a.fullname,a.google_id
FROM afiliasi af
join author a on a.afiliasi_id = af.afiliasi_id
WHERE af.province = 13
order by author_id asc
";
$result = $conn->query($sql);

if ($result->num_rows > 0) {
// output data of each row
while($row = $result->fetch_assoc()) {
//echo "id: " . $row["author_id"]. " - Name: " . $row["fullname"]. " <br> " . $row["google_id"]. "<br>";

//scraping satu per satu author
$gs_id = $row["google_id"];
$url = "https://scholar.google.co.id/citations?user=".$gs_id."&shl=enview_op=list_works";
$html = ambil_curl($url);
$pub = google_profile($html);
//print_r($pub);
//echo "<br>";
echo "\n".$row["author_id"]." ".$row["fullname"]." : ";

$conn->query("delete from kepakaran where author_id = ".$row["author_id"]);

for($i=0;$i<sizeof($pub['bidang']);$i++){
//masukkan setiap kepakaran ke dalam database
echo $qinsert ="insert into kepakaran (kepakaran_id,author_id,kepakaran) value ('".$row["author_id"]."','".$pub['bidang'][$i]."')";
$conn->query($qinsert);
echo "\n";
}
}
```

Penjelasan kode program :

Setelah membuat *function* pada file tersendiri, barulah pembuatan *code scraping* dilakukan agar lebih ringkas

- #001 : Merupakan serangkaian *query* SQL yang digunakan untuk menyeleksi Id peneliti khusus Jawa Tengah termasuk Id google scholar milik peneliti.
- #002 : Merupakan serangkaian proses *scraping* dimana proses ini menggunakan Id google scholar sebagai kunci utama untuk mendapatkan data peneliti satu persatu. Penggunaan *function* pada file “gs\_lib.php” akan dipanggil untuk menyelesaikan tugas ini.
- #003 : Merupakan tahap penyimpanan data peneliti ke dalam *database* setelah data berhasil di-*scrap*. Proses ini akan terus berlangsung sampai jumlah Id peneliti terhitung seluruhnya beserta kepakaran peneliti.