

**SKRIPSI**

**PEMBUATAN GRAPH DARI DOKUMEN  
PERATURAN/PERUNDANG-UNDANGAN DAN  
ANALISISNYA DENGAN TEKNOLOGI BIG DATA**



**Upekkha Lau**

**NPM: 6181801023**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2022**



**UNDERGRADUATE THESIS**

**GRAPH CONSTRUCTION FROM LEGISLATION  
DOCUMENTS AND ITS ANALYSIS USING BIG DATA  
TECHNOLOGY**



**Upekkha Lau**

**NPM: 6181801023**

**DEPARTMENT OF INFORMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2022**



# LEMBAR PENGESAHAN

## PEMBUATAN GRAPH DARI DOKUMEN PERATURAN/PERUNDANG-UNDANGAN DAN ANALISISNYA DENGAN TEKNOLOGI BIG DATA

Upekkha Lau

NPM: 6181801023

Bandung, 22 Juli 2022

Menyetujui,

Pembimbing

Digitally signed  
by Veronica Sri  
Moertini

Dr. Veronica Sri Moertini

Ketua Tim Penguji

Digitally signed  
by Natalia

Natalia, M.Si.

Anggota Tim Penguji

Digitally signed  
by Keenan  
Adiwijaya Leman

Keenan Adiwijaya Leman, M.T.

Mengetahui,

Ketua Program Studi

Digitally signed  
by Mariskha Tri  
Adithia

Mariskha Tri Adithia, P.D.Eng



## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

### **PEMBUATAN GRAPH DARI DOKUMEN PERATURAN/PERUNDANG-UNDANGAN DAN ANALISISNYA DENGAN TEKNOLOGI BIG DATA**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
Tanggal 22 Juli 2022



Upekkha Lau  
NPM: 6181801023





## ABSTRAK

Negara Indonesia adalah negara hukum. Sebagai negara hukum setiap tahunnya negara Indonesia melahirkan undang-undang baru. Dari tahun 1945 di saat Indonesia merdeka hingga saat ini sudah ada lebih dari 1000 undang-undang yang sudah diciptakan. Undang-undang tidak dibuat begitu saja, akan tetapi memiliki dasar tertentu, seperti yang akan diteliti pada skripsi ini yaitu berdasarkan undang-undang lain. Hampir seluruh undang-undang yang dibuat merujuk kepada undang-undang yang sudah dibuat sebelumnya baik untuk memperbaharui, mengganti, atau merevisi. Begitu banyaknya undang-undang beserta rujukannya tentu akan membuat para praktisi hukum kesulitan untuk mencari undang-undang yang relevan untuk pekerjaannya, maka dari itu pada skripsi ini dibuat sebuah perangkat lunak yang mampu membantu melakukan pencarian undang-undang terhadap topik atau nomor dan tahun undang-undang beserta rujukannya. Selain itu rujukan-rujukan ini akan dimodelkan dalam bentuk graf yang kemudian dianalisis dengan memanfaatkan teknologi *big data* yaitu *Apache Spark*.

Pemodelan dan penyiapan graf dilakukan dengan memanfaatkan bahasa pemrograman *python*, dokumen PDF undang-undang yang telah dikumpulkan dilakukan ekstraksi teks dengan menggunakan *library python* yaitu *MyMuPDF*. Kemudian dilakukan ekstraksi kata kunci dari teks menggunakan *Regular Expression* yaitu dengan mencari kata kunci seperti nomor dan tahun undang-undang atau topik undang-undang tersebut. Terakhir kata kunci dimodelkan menjadi graf dengan membuat *data node* yang berisikan seluruh nomor dan tahun undang-undang beserta topiknya dan *data edge* yang berisikan rujukan-rujukan yang dihasilkan oleh undang-undang.

*Apache Spark* adalah teknologi *big data* yang mampu memproses graf dengan jutaan *node* dan *edge*. Komponen *Apache Spark* yang digunakan untuk analisis graf adalah *graphX* dan *graphFrames*. *GraphX* dan *graphFrames* menyediakan beberapa algoritma analisis graf, yaitu *PageRank*, *Connected Components*, *Strongly Connected Components*, *Breadth First Search*, dan *Motif Finding*. *PageRank* digunakan untuk mencari undang-undang yang penting, *Connected Components* digunakan untuk mencari sekumpulan undang-undang yang memiliki topik yang sama, *Strongly Connected Components* tidak ditemukan karena graf undang-undang tidak mungkin menciptakan sebuah graf *cyclic*, *Breadth First Search* digunakan untuk menemukan apakah dua undang-undang memiliki hubungan dan melalui apa saja, dan *Motif finding* yang tidak digunakan karena tidak ada pola yang relevan untuk dianalisis.

Perangkat lunak yang dibangun menggunakan bahasa pemrograman PHP, HTML, dan JavaScript. Menggunakan *Bootstrap* sebagai *framework CSS*-nya, dan menerapkan struktur *Model, View, Controller* sebagai arsitekturnya. Perangkat lunak ini dapat menampilkan teks dan graf dari undang-undang beserta menampilkan rujukan dan status revisinya.

**Kata-kata kunci:** undang-undang, rujukan, *Apache Spark*, *GraphFrames*, *PageRank*, *Connected Components*, *Strongly Connected Components*, *Breadth First Search*, *Motif finding*



## ABSTRACT

Indonesia is a country of law, as a country of law this state create new law every year. From 1945 when the start of Indonesian independence until this day there have been more than 1000 law created by this country. The laws are not made just like that, but it's also still has a some basis, as will be research by this thesis which is base by others law. Almost all law are made by referencing other laws that are already been made such as for update, replace, or revision. There are so many laws and their references that will certainly make it difficult for law practitioners to find laws that are relevant for their work, therefore in this thesis a software is made to help them search for laws by their topics or numbers and years of the laws and also their references. These references will also be modeled as graph and analyzed using big data technology that is Apache Spark.

The modeling dan pre-processing are conducted using the python programming language, the text extraction from the law documents are done using one of the libraries known as MyMuPDF. Then the keywords will be extracted using Regular Expression by finding the number and the year of the law or their topics. Lastly, the keywords are modeled into graph by creating data node that contains the number and the year of the law and also their topics, and data edge that contains the reference of the laws.

Apache Spark is a big data technology that help in processing graph with millions of nodes and edges. Apache Spark components used for graph analysis are GrpahX and GraphFrames, graphX and graphFrames provides many graph analysis algorithm, such as PageRank, Connected Components, Strongly Connected Components, Breadth First Search, and Motif Finding. PageRank is used for finding important laws, Connected Components is used for searching a set of rules that have the same topic, Strongly Connected Components will not be found because graph from law will never make a cyclic graph, Breadth First Search is used to find whether two specific laws have a relation and through what laws, and Motif Finding that will not be used because there is no relevant pattern to be analyze.

The software will be built using PHP, HTML, and JavaScript. Using Bootstrap as the framework CSS, and Model, View, Controller structure as the architecture. This software will be able to display the text and the graph of laws along with showing their references and revision status.

**Keywords:** laws, reference, *Apache Spark*, *GraphFrames*, *PageRank*, *Connected Components*, *Strongly Connected Components*, *Breadth First Search*, *Motif finding*



*Skripsi ini dipersembahkan untuk seluruh para pembaca, semoga  
bermanfaat bagi kita semua.*



# KATA PENGANTAR

**Sabbe Satta Bhavantu Sukhitatta**  
Semoga Semua Makhluk Hidup Berbahagia

Pada kesempatan ini penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah membantu dalam melancarkan pengerjaan skripsi ini, terutama kepada.

1. Orang tua dan saudara penulis yang selalu memberi dukungan dan selalu yakin bahwa penulis dapat menyelesaikan skripsi ini;
2. Ibu Dr. Veronica Sri Moertini selaku pembimbing yang selalu bersemangat dalam membimbing penulis;
3. Ibu Natalia, M.Si. dan Bapak Keenan Adiwijaya Leman, M.T. selaku penguji, terima kasih atas masukan dan saran yang diberikan;
4. Indra Permana Sugianto, terima kasih atas seluruh pedoman yang telah diberikan;
5. Dan kepada seluruh pihak yang tidak tercantumkan namanya.

4 tahun sudah berlalu dalam perjuangan menyelesaikan pendidikan S1. Untuk rekan-rekan yang lulus bersamaan dengan penulis, maju terus perjalanan hidup baru saja dimulai. Untuk rekan-rekan yang akan lulus, semangat terus dan pantang menyerah dalam menyelesaikan pendidikannya. Akhir kata, penulis memohon maaf jika terdapat kesalahan pada penulisan maupun teori yang dipaparkan pada skripsi ini.

Bandung, Juli 2022

Penulis





# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>xv</b>
<b>DAFTAR ISI</b>	<b>xvii</b>
<b>DAFTAR GAMBAR</b>	<b>xix</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	2
1.4 Batasan Masalah . . . . .	2
1.5 Metodologi . . . . .	3
1.6 Sistematika Pembahasan . . . . .	3
<b>2 LANDASAN TEORI</b>	<b>5</b>
2.1 Peraturan Perundang-undangan Indonesia . . . . .	5
2.2 Struktur Peraturan Perundang-undangan di Indonesia . . . . .	6
2.3 Penyiapan Data Teks . . . . .	9
2.4 Spark . . . . .	10
2.5 Scala . . . . .	11
2.6 <i>Dataframe</i> . . . . .	12
2.6.1 <i>Dataframe</i> dengan Scala . . . . .	13
2.6.2 <i>Dataframe</i> dengan python . . . . .	13
2.7 Graf . . . . .	14
2.8 GraphX dan Graphframes . . . . .	16
2.8.1 Algoritma Analisis Graf . . . . .	17
<b>3 PENGUMPULAN, PENYIAPAN, DAN EKSPLORASI DATA</b>	<b>21</b>
3.1 Pengumpulan Data Teks PDF Perundang-undangan Indonesia . . . . .	21
3.2 Penyiapan Data . . . . .	23
3.2.1 Ekstraksi Teks PDF Perundang-undangan Indonesia . . . . .	24
3.2.2 Ekstraksi Kata Kunci . . . . .	25
3.2.3 Ekstraksi Topik Undang-undang . . . . .	26
3.2.4 Menyiapkan <i>Node</i> dan <i>Edge</i> . . . . .	28
3.2.5 Ekstraksi Undang-undang Revisi . . . . .	30
3.3 Eksplorasi Data . . . . .	31
3.3.1 Memodelkan Kata Kunci Rujukan Menjadi Graf . . . . .	31
3.3.2 Memvisualisasikan Graf dengan Memanfaatkan Cytoscape . . . . .	35
<b>4 ANALISIS GRAF</b>	<b>37</b>
4.1 <i>PageRank</i> . . . . .	37
4.2 <i>Connected Components</i> . . . . .	40
4.3 <i>Strongly Connected Components</i> . . . . .	43

4.4	<i>Breadth First Search</i> . . . . .	44
4.5	Kesimpulan dan Ringkasan Insight . . . . .	47
<b>5</b>	<b>PEMBANGUNAN PERANGKAT LUNAK</b>	<b>49</b>
5.1	Fitur-fitur perangkat lunak . . . . .	49
5.2	Diagram Use Case . . . . .	51
5.3	Perancangan Antarmuka . . . . .	54
5.3.1	Halaman <i>search</i> . . . . .	54
5.3.2	Halaman <i>searchTentang</i> . . . . .	55
5.3.3	Halaman <i>searchPohon</i> . . . . .	56
5.3.4	Halaman <i>detailPDF</i> . . . . .	56
5.3.5	Halaman <i>detailPohon</i> . . . . .	57
5.4	Perancangan Basis Data . . . . .	58
5.4.1	Tabel <i>list_undang</i> . . . . .	58
5.4.2	Tabel <i>graph_edge</i> . . . . .	58
5.4.3	Tabel revisi . . . . .	58
5.5	<i>Diagram Kelas</i> . . . . .	58
5.5.1	Kelas <i>Model</i> . . . . .	59
5.5.2	Kelas <i>Controller</i> . . . . .	60
5.6	Implementasi . . . . .	60
5.6.1	Halaman <i>home</i> . . . . .	61
5.6.2	Halaman <i>tentang</i> . . . . .	62
5.6.3	Halaman <i>pdfHome</i> . . . . .	64
5.6.4	Halaman <i>Pohon</i> . . . . .	64
5.6.5	Halaman <i>pdfPohon</i> . . . . .	65
<b>6</b>	<b>KESIMPULAN DAN SARAN</b>	<b>69</b>
6.1	Kesimpulan Penelitian . . . . .	69
6.2	Saran Pengembangan . . . . .	69
	<b>DAFTAR REFERENSI</b>	<b>71</b>
	<b>A KODE PROGRAM</b>	<b>73</b>

## DAFTAR GAMBAR

2.1	Judul UU Nomor 1 tahun 2020	6
2.2	Frasa dan Jabatan	6
2.3	Konsideran dalam bentuk Menimbang	7
2.4	Dasar Hukum dalam bentuk Mengingat	8
2.5	Diktum UU nomor 1 tahun 2020	8
2.6	Contoh Batang Tubuh	9
2.7	Tanda Tangan Menteri dan Presiden	9
2.8	Arsitektur Spark	11
2.9	Contoh Graf	14
2.10	<i>Undirected Graph</i>	14
2.11	<i>Directed Graph</i>	14
2.12	<i>SelfLoop dan Parallel Edge</i>	15
2.13	<i>SimpleGraph</i>	15
2.14	<i>In-Degree</i>	15
2.15	<i>Out-Degree</i>	16
2.16	<i>Cycle</i>	16
2.17	Ilustrasi <i>PageRank</i>	17
2.18	Ilustrasi <i>Connected Components</i>	18
2.19	Ilustrasi <i>Strongly Connected Components</i>	18
2.20	Ilustrasi <i>Breadth First Search</i>	19
2.21	Ilustrasi <i>Motif Finding</i>	19
3.1	Jumlah undang-undang tiap tahunnya	22
3.2	undang-undang nomor 1 tahun 2020	23
3.3	undang-undang nomor 12 tahun 1969	23
3.4	Flow-chart Penyiapan Data	24
3.5	Hasil Ekstraksi Teks dari PDF	25
3.6	Hasil Ekstraksi Kata Kunci	26
3.7	Posisi Topik Undang-undang	27
3.8	Hasil Ekstraksi Topik Teks Undang-undang	28
3.9	Hasil Node	29
3.10	Hasil Edge	30
3.11	Undang-undang Revisi	30
3.12	Undang-undang Revisi	31
3.13	Error dalam menjalankan <code>graphFrames</code> di lokal	32
3.14	Spark yang berhasil dijalankan di lab	32
3.15	Menyiapkan Node dan Edge	33
3.16	Membuat graph	33
3.17	<code>node</code>	34
3.18	<code>edges</code>	34
3.19	<i>In-Degree</i>	34
3.20	<i>Out-Degree</i>	34

3.21	Kumpulan Undang-undang yang saling berhubungan dari graf mengingat 1 . . . . .	35
3.22	Kumpulan Undang-undang yang saling berhubungan dari graf mengingat 2 . . . . .	35
3.23	Kumpulan Undang-undang yang saling berhubungan dari graf menimbang 1 . . . . .	36
3.24	Kumpulan Undang-undang yang saling berhubungan dari graf menimbang 2 . . . . .	36
4.1	Graf Undang-undang Nomor 6 Tahun 1969 . . . . .	38
4.2	Graf Undang-undang Nomor 15 Tahun 1969 . . . . .	38
4.3	Graf Undang-undang Nomor 14 Tahun 1970 . . . . .	39
4.4	Graf Undang-undang Nomor 5 tahun 1974 . . . . .	39
4.5	Graf Undang-undang Nomor 4 tahun 1999 . . . . .	40
4.6	Graf Connected Components 12 <i>Node</i> . . . . .	41
4.7	Graf Connected Components 6 <i>Node</i> . . . . .	42
4.8	Graf Connected Components 3 <i>Node</i> . . . . .	43
4.9	Graf dari BFS 1 . . . . .	45
4.10	Hasil dari BFS 2 . . . . .	45
4.11	Hasil dari BFS 3 . . . . .	46
4.12	Graf dari BFS 4 . . . . .	47
5.1	<i>Screen Flow Diagram</i> . . . . .	51
5.2	<i>Diagram Use Case</i> . . . . .	52
5.3	<i>Diagram Rancangan halaman search</i> . . . . .	55
5.4	<i>Diagram Rancangan halaman searchTentang</i> . . . . .	55
5.5	<i>Diagram Rancangan halaman searchPohon</i> . . . . .	56
5.6	<i>Diagram Rancangan halaman detailPDF</i> . . . . .	57
5.7	<i>Diagram Rancangan halaman detailPohon</i> . . . . .	57
5.8	Diagram Kelas . . . . .	59
5.9	Halaman <i>home</i> . . . . .	61
5.10	Hasil Pencarian Undang-undang Nomor 24 . . . . .	62
5.11	Hasil Pencarian Undang-undang Tahun 2020 . . . . .	62
5.12	Halaman Tentang . . . . .	63
5.13	Hasil Pencarian Undang-undang dengan Topik "narkotika" . . . . .	63
5.14	Halaman pdfHome . . . . .	64
5.15	Halaman Pohon . . . . .	65
5.16	Hasil Pencarian Undang-undang nomor 23 tahun 2000 . . . . .	65
5.17	Halaman pdfPohon contoh 1 . . . . .	66
5.18	Halaman pdfPohon contoh 2 . . . . .	67
5.19	Halaman pdfPohon contoh 3 . . . . .	67

# BAB 1

## PENDAHULUAN

Bab ini akan membahas latar belakang penelitian yang dikerjakan, rumusan masalah yang dikemukakan, tujuan penelitian yang akan dicapai, batasan masalah penelitian, metodologi penelitian yang akan dilaksanakan dan sistematika pembahasan dokumen penulisan ilmiah ini.

### 1.1 Latar Belakang

Negara Indonesia adalah negara hukum, dengan sebutan sebagai negara hukum Indonesia memiliki banyak sekali aturan-aturan hukum yang disimpan dalam bentuk perundang-undangan. Dari tahun 1945 di saat indonesia merdeka hingga saat ini sudah ada lebih dari 1000 undang-undang yang sudah diciptakan. [1] Peraturan perundang-undangan ini berfungsi untuk mengatur masyarakat ke arah yang lebih baik. Peraturan perundang-undangan adalah aturan yang tertulis dan disahkan oleh negara untuk dipatuhi seluruh masyarakat. Undang-undang tidak dibuat begitu saja akan tetapi tetap memiliki dasar tertentu, seperti dari undang-undang lain yang kurang lengkap atau memiliki kelemahan, maka dari itu dibuat peraturan baru untuk menyokong peraturan yang lain. Hampir seluruh undang-undang yang dibuat merujuk kepada undang-undang yang sudah dibuat sebelumnya baik untuk memperbaharui, mengganti, atau merevisi. Begitu banyaknya undang-undang beserta rujukannya tentu akan membuat para praktisi hukum kesulitan untuk mencari undang-undang yang relevan untuk pekerjaannya.

Melihat hal tersebut muncul sebuah pertanyaan, apakah rujukan undang-undang ini memiliki manfaat bagi praktisi hukum? Maka dari itu akan dilakukan sebuah survei kepada pihak Fakultas Hukum UNPAR untuk menanyakan apakah hubungan rujukan-rujukan peraturan perundang-undangan dapat membantu mereka dalam bekerja dibidangnya. Rujukan-rujukan ini akan dimodelkan dalam bentuk graf yang kemudian dianalisis dengan memanfaatkan teknologi *big data* yaitu *Apache Spark*.

Pemodelan dan penyiapan graf dilakukan dengan memanfaatkan bahasa pemrograman *python*, dokumen PDF undang-undang yang telah dikumpulkan dilakukan ekstraksi teks dengan menggunakan *library python* yaitu *MyMuPDF*. Kemudian dilakukan ekstraksi kata kunci dari teks menggunakan *Regular Expression* yaitu dengan mencari kata kunci seperti nomor dan tahun undang-undang atau topik undang-undang tersebut. Terakhir kata kunci dimodelkan menjadi graf dengan membuat *data node* yang berisikan seluruh nomor dan tahun undang-undang beserta topiknya dan *data edge* yang berisikan rujukan-rujukan yang dihasilkan oleh undang-undang.

Graf adalah cara merepresentasikan koneksi antar *node* yang kemudian akan mempermudah proses analisis ataupun visualisasi. *Node* merepresentasikan peraturan undang-undang dan *edges* merepresentasikan rujukan undang-undang. *Node* didapatkan dari mengekstrak kata kunci dari dokumen peraturan perundang-undangan dengan menggunakan salah satu teknik *Natural Language Processing* yaitu *Regular Expressions*. *Regular Expressions* atau yang sering disebut sebagai *regex* adalah sebuah teknik *pre-processing* untuk menentukan *string* pencarian teks, mencari sebuah pola yang ada dari teks.

*Apache spark* adalah teknologi *big data* yang mampu memproses graf dengan jutaan *node* dan *edge*. Komponen *apache spark* yang digunakan untuk analisis graf adalah *graphX* dan *graphFrames*, *graphX* dan *graphFrames* menyediakan beberapa algoritma analisis graf, yaitu *pageRank*, *connected*

*components*, *strongly connected components*, *breadth first search*, dan *motif finding*. Selain itu juga analisis akan dilakukan dengan cara memvisualisasikan graf untuk dicari rumpunan aturan yang terbentuk dengan memanfaatkan tools seperti cytoscape.

Pada Skripsi ini, akan dibangun sebuah perangkat lunak yang mampu mempermudah penggunaannya untuk melihat teks dan status revisi undang-undang yang ingin dicarinya, perangkat lunak ini juga mampu menampilkan rujukan dari undang-undang yang dicari, pengguna dapat dengan mudah menavigasikan antar undang-undang dengan menekan rujukan undang-undang yang dipaparkan, aplikasi berbasis web ini juga akan dirancang untuk dapat memvisualisasikan rujukan dari undang-undang yang dicarinya. Pencarian dapat dilakukan dengan dua cara, yaitu dengan mencari menggunakan nomor atau tahun undang-undang yang dicari dan mencari berdasarkan topik dari undang-undangnya. Pengguna juga dapat dengan mudah berpindah dari melihat teks ke halaman yang menampilkan graf rujukannya.

## 1.2 Rumusan Masalah

Berikut ini adalah rumusan masalah dari latar belakang yang telah dipaparkan.

1. Bagaimana cara mengumpulkan data teks PDF peraturan perundang-undangan Indonesia?
2. Bagaimana cara menyiapkan data teks PDF peraturan perundang-undangan Indonesia?
3. Bagaimana cara mengekstraksi kata kunci rujukan dari teks peraturan perundang-undangan?
4. Bagaimana cara mengkonstruksi graf dari kata kunci rujukan peraturan perundang-undangan?
5. Bagaimana cara menganalisis graf rujukan peraturan perundang-undangan?
6. Bagaimana cara menggali *insight* dari graf yang dihasilkan?
7. Bagaimana membangun perangkat lunak yang mampu melakukan penyiapan dan analisis data peraturan perundang-undangan?

## 1.3 Tujuan

Adapun tujuan penelitian dari rumusan masalah yang telah dipaparkan adalah:

1. Mengumpulkan teks PDF peraturan perundang-undangan Indonesia;
2. Memanfaatkan teknik *Natural Language Processing* yaitu *Regular Expression* untuk menyiapkan data dan mengekstraksi kata kunci rujukan dari teks peraturan perundang-undangan;
3. Mempelajari teknologi GraphX dan GraphFrame untuk mengkonstruksi graf dari kata kunci rujukan peraturan perundang-undangan;
4. Mempelajari algoritma *PageRank*, *Connected Components*, *Strongly Connected Components*, *Breadth First Search*, dan *Motif Finding* untuk menganalisis graf rujukan peraturan perundang-undangan;
5. Mencari algoritma analisis graf yang berpotensi untuk digunakan;
6. Menggali *insight* dari graf yang dihasilkan;
7. Membangun perangkat lunak yang mampu melakukan penyiapan dan analisis data peraturan perundang-undangan.

## 1.4 Batasan Masalah

Batasan untuk penelitian ini adalah sebagai berikut:

1. Teks Undang-undang yang berbentuk gambar tidak ditangani karena teks tidak dapat diekstraksi;
2. Teks Undang-undang dengan pembuka yang tidak lengkap tidak ditangani karena merusak proses penyiapan data;
3. Teks Undang-undang yang dianalisis hanya sampai undang-undang tahun 1969 karena lebih tua dari itu data sudah tidak lengkap.

## 1.5 Metodologi

Metodologi yang digunakan dalam pembuatan skripsi ini adalah sebagai berikut:

1. Melakukan studi literatur tentang peraturan perundang-undangan di Indonesia;
2. Melakukan studi literatur tentang *Natural Language Processing*;
3. Melakukan eksplorasi dan studi literatur tentang Spark;
4. Melakukan eksplorasi dan studi literatur tentang bahasa pemrograman Scala;
5. Melakukan eksplorasi dan studi literatur tentang cara kerja pembuatan graf menggunakan *library* GraphX dan GraphFrame;
6. Melakukan eksplorasi dan studi literatur tentang cara kerja *PageRank*, *Connected Components*, *Strongly Connected Components*, *Breadth First Search*, dan pencarian subgraf dengan pola tertentu;
7. Mengumpulkan data teks PDF peraturan perundang-undangan Indonesia;
8. Mengekstrak kata kunci rujukan dari teks PDF peraturan perundang-undangan Indonesia;
9. Memodelkan kata kunci rujukan dalam bentuk graf;
10. Menganalisis graf menggunakan algoritma *PageRank*, *Connected Components*, *Strongly Connected Components*, *Breadth First Search*, dan pencarian subgraf dengan pola tertentu;
11. Membangun perangkat lunak;
12. Menguji perangkat lunak yang telah dibuat;
13. Menulis dokumen skripsi.

## 1.6 Sistematika Pembahasan

Sistematika penulisan skripsi ini adalah sebagai berikut.

1. Pendahuluan
  - (a) Latar Belakang
  - (b) Rumusan Masalah
  - (c) Tujuan
  - (d) Batasan Masalah
  - (e) Metodologi
  - (f) Sistematika Pembahasan
2. Landasan Teori
  - (a) Peraturan Perundang-undangan Indonesia
  - (b) Struktur Peraturan Perundang-undangan di Indonesia
  - (c) Penyiapan Data Teks
  - (d) *Apache Spark*
  - (e) *Scala* dan *Dataframe*
  - (f) Graf
  - (g) *GraphX* dan *GraphFrames*
3. Pengumpulan, Penyiapan dan Eksplorasi Data
  - (a) Pengumpulan Data Teks PDF Perundang-undangan Indonesia
  - (b) Penyiapan Data
  - (c) Eksplorasi Data
4. Analisis Graf
  - (a) *PageRank*
  - (b) *Connected Components*
  - (c) *Strongly Connected Components*
  - (d) *Breadth First Search*
5. Pembangunan Perangkat Lunak
  - (a) Fitur-fitur Perangkat Lunak
  - (b) Diagram Use Case

- (c) Perancangan Antarmuka
  - (d) Perancangan Basis Data
  - (e) Diagram Kelas
  - (f) Implementasi
6. Kesimpulan dan Saran
- (a) Kesimpulan Penelitian
  - (b) Saran Pengembangan