

**OWL KNOWLEDGE BASE GENERATOR DINAMIS DARI  
KONTEN HALAMAN WEBSITE DENGAN METODE  
GRAPH BREADTH FIRST SEARCH**

**SKRIPSI**

**Diajukan kepada:  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
untuk Memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Sarjana Komputer (S. Kom)**

**Oleh:  
KALVIN NIAM AUNILLAH  
NIM. 17650060**

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2021**

**LEMBAR PERSETUJUAN**

**OWL KNOWLEDGE BASE GENERATOR DINAMIS DARI  
KONTEN HALAMAN WEBSITE DENGAN METODE  
GRAPH BREADTH FIRST SEARCH**

**SKRIPSI**

**Oleh:  
KALVIN NIAM AUNILLAH  
NIM. 17650060**

Telah Diperiksa dan Disetujui untuk Diuji:  
Tanggal: 6 Agustus 2021

Pembimbing I,



A'la Syaqi, M.Kom  
NIP. 19771201 200801 1 007

Pembimbing II,



Yunifa Miftachul Arif, M.T  
NIP. 19830616 201101 1 004

Mengetahui,  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. Fachrul Kurniawan, M.MT  
NIP. 19771020 200912 1 001

## LEMBAR PENGESAHAN

### OWL KNOWLEDGE BASE GENERATOR DINAMIS DARI KONTEN HALAMAN WEBSITE DENGAN METODE GRAPH BREADTH FIRST SEARCH

#### SKRIPSI

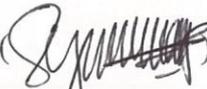
Oleh:  
**KALVIN NIAM AUNILLAH**  
NIM. 17650060

Telah Dipertahankan di Depan Dewan Penguji Skripsi  
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan  
untuk Memperoleh Gelar Sarjana Komputer (S. Kom)  
Tanggal: 6 Agustus 2021

#### Susunan Dewan Penguji

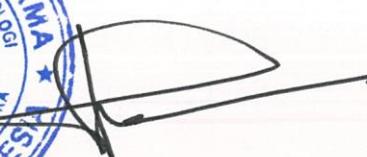
1. Penguji Utama : Dr. Cahyo Crysdiان  
NIP. 19740424 200901 1 008
2. Ketua Penguji : Puspa MiIadin, M.Kom  
NIP. 19930828 201903 2 018
3. Sekretaris Penguji : A'Ia Syauqi, M. Kom  
NIP. 19771201 200801 1 007
4. Anggota Penguji : Yunifa Miftachul Arif, M. T  
NIP. 19830616 201101 1 004

#### Tanda Tangan

()  
()  
()  
()

Mengetahui,  
Ketua Jurusan Teknik Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



  
Dr. Fachrul Kurniawan, M.MT  
NIP. 19771020 200912 1 001

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Calvin Niam Aunillah  
NIM : 17650060  
Jurusan : Teknik Informatika  
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 6 Agustus 2021

Yang membuat pernyataan,

A handwritten signature in black ink is written over a yellow and red revenue stamp. The stamp features the number '1000' and the text 'SEPULUH RIBU RUPIAH' on the left, 'TEL. 20' and 'METERAI TEMPEL' on the right, and the alphanumeric code 'CADCEAJX552797754' at the bottom.

Kalvin Niam Aunillah  
NIM. 17650060

## HALAMAN PERSEMBAHAN



Skripsi ini saya persembahkan untuk  
kedua orang tua, keluarga,  
seluruh guru, dosen dan  
teman-teman seperjuangan

Terima kasih.

## MOTTO

*“Bukan ilmu yang semestinya mendatangiimu, tapi kamulah yang seharusnya mendatangi ilmu itu”*

## KATA PENGANTAR

Segala puji dan syukur kehadirat Allah SWT karena atas berkat Rahmat dan karunia-Nya, penulis diberikan kemudahan dalam menyelesaikan skripsi ini. Penyusunan skripsi ini bertujuan untuk memenuhi syarat kelulusan bagi mahasiswa Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang. Keberhasilan penulisan skripsi ini tidak lepas dari dorongan dan bimbingan dari berbagai pihak. Untuk itu dalam kesempatan ini penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. M. Zainuddin, MA selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Sri Hariani, M.Si selalu dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Bapak Dr. Fachrul Kurniawan, M.MT selaku Ketua Jurusan Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Bapak A'la Syauqi, M. Kom selaku dosen pembimbing I yang telah bersedia meluangkan waktunya dalam membimbing dan memberikan dukungan kepada penulis sehingga dapat menyelesaikan skripsi ini.
5. Yunifa Miftachul Arif, M. T selaku dosen pembimbing II yang juga bersedia meluangkan waktunya dalam membimbing dan memberi arahan kepada penulis sehingga dapat menyelesaikan skripsi ini.
6. Bapak dan Ibu beserta keluarga yang telah memberikan dukungan baik moral maupun spiritual sehingga penulis diberi kemudahan dalam menyelesaikan skripsi ini.
7. Seluruh dosen dan staff Jurusan Teknik Informatika yang telah memberikan ilmu dan pengalaman yang berharga.
8. Teman-teman seangkatan, kakak tingkat, adik tingkat dan teman-teman komunitas yang telah memberikan dukungan dan pengalaman yang berharga.
9. Semua pihak yang telah membantu dalam penyelesaian skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa masih banyak kekurangan dari laporan ini. Oleh karena itu, penulis memohon maaf atas segala kekurangan yang terjadi selama proses penyusunan skripsi ini. Semoga tugas akhir ini dapat memberikan kontribusi yang bermanfaat bagi penulis dan khususnya pembaca.

Malang, 6 Agustus 2021

Penulis.

## DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PERSETUJUAN.....	<b>Error! Bookmark not defined.</b>
LEMBAR PENGESAHAN .....	<b>Error! Bookmark not defined.</b>
PERNYATAAN KEASLIAN TULISAN .....	iii
HALAMAN PERSEMBAHAN .....	v
MOTTO .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
ABSTRAK .....	xiii
ABSTRACT.....	xiv
المخلص .....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang .....	1
1.2. Pernyataan Masalah.....	4
1.3. Tujuan Penelitian.....	5
1.4. Batasan Penelitian .....	5
1.5. Manfaat Penelitian.....	5
1.6. Sistematika Penulisan.....	5
BAB II KAJIAN PUSTAKA .....	9
2.1. <i>Web Semantic</i> .....	10
2.2. <i>Knowledge Base</i> .....	12
2.3. <i>Resource Description Framework (RDF)</i> .....	12
2.4. <i>Web Scraping</i> .....	15
2.5. Node Js .....	18
2.6. Metode <i>Breadth First Search (BFS)</i> .....	18
2.7. ISO 9126.....	20
2.8. Analisis <i>Function Point</i> .....	21
BAB III PEMBAHASAN .....	23

3.1. Desain Sistem .....	23
3.2. Perancangan Sistem.....	26
3.3. Penelusuran Halaman <i>Website</i> .....	29
3.4. <i>Crawling</i> .....	29
3.5. <i>Scraping</i> .....	31
3.6. <i>Generate File RDF</i> .....	31
3.7. Kerangka Kerja Konseptual .....	31
<b>BAB IV UJI COBA DAN PEMBAHASAN .....</b>	<b>33</b>
4.1. Langkah-langkah Uji Coba .....	33
4.1.1. <i>Crawling</i> dengan Metode <i>Breadth First Search</i> .....	33
4.1.2. Pembuatan <i>File RDF</i> .....	41
4.1.3. Validasi Sintaks RDF .....	44
4.1.4. <i>Linked Data</i> RDF.....	45
4.1.5. Evaluasi Sistem .....	46
4.2. Hasil Uji Coba .....	51
4.3. Pembahasan .....	53
<b>BAB V KESIMPULAN DAN SARAN.....</b>	<b>55</b>
5.1. Kesimpulan.....	55
5.2. Saran .....	55
<b>DAFTAR PUSTAKA .....</b>	<b>56</b>

## DAFTAR GAMBAR

<b>Gambar 2.1</b> Arsitektur <i>Web Semantic</i> .....	11
<b>Gambar 2.2</b> Contoh struktur XML .....	13
<b>Gambar 2.3</b> Contoh <i>unordered graph</i> XML .....	13
<b>Gambar 2.4</b> Contoh <i>ordered graph</i> XML .....	14
<b>Gambar 2.5</b> Contoh struktur RDF .....	15
<b>Gambar 2.6</b> Ilustrasi cara kerja <i>Web Scraping</i> .....	18
<b>Gambar 2.7</b> V8 JavaScript <i>engine</i> .....	18
<b>Gambar 2.8</b> <i>Graph Breadth First Search</i> .....	19
<b>Gambar 2.9</b> <i>Pseudocode</i> algoritma <i>Breadth First Search</i> .....	20
<b>Gambar 3.1</b> Desain sistem .....	23
<b>Gambar 3.2</b> Model kualitas perangkat lunak berdasarkan ISO 9126 .....	25
<b>Gambar 3.3</b> Blok diagram perancangan sistem .....	27
<b>Gambar 3.4</b> <i>Flowchart</i> manajemen hasil ekstraksi data .....	28
<b>Gambar 3.5</b> <i>Flowchart</i> manajemen <i>user</i> .....	28
<b>Gambar 3.6</b> <i>Flowchart</i> penelusuran halaman .....	29
<b>Gambar 3.7</b> Representasi <i>Graph nodes</i> tag HTML .....	30
<b>Gambar 3.8</b> <i>Flowchart</i> penelusuran atribut .....	30
<b>Gambar 3.9</b> Kerangka kerja konseptual .....	32
<b>Gambar 4.1</b> Halaman artikel berita Detik.com .....	33
<b>Gambar 4.2</b> <i>Graph</i> tag HTML beserta urutan traversal .....	40
<b>Gambar 4.3</b> Hasil <i>generate file</i> RDF dari halaman Detik.com .....	43
<b>Gambar 4.4</b> Hasil <i>generate file</i> RDF dari halaman Kompas.com .....	44
<b>Gambar 4.5</b> Tabel <i>triples website</i> W3C <i>RDF Validation Service</i> .....	45
<b>Gambar 4.6</b> Eksekusi <i>query</i> SPARQL untuk menampilkan data .....	46
<b>Gambar 4.7</b> Tabel <i>triples</i> hasil <i>query</i> SPARQL .....	46

## DAFTAR TABEL

<b>Tabel 3.1</b> Tabel klasifikasi parameter <i>Function Point</i> .....	25
<b>Tabel 4.1</b> Tabel perhitungan <i>Function Point</i> .....	47
<b>Tabel 4.2</b> Tabel <i>Complexity Factor</i> .....	47
<b>Tabel 4.3</b> Hasil <i>time behaviour</i> .....	49
<b>Tabel 4.4</b> Hasil pengujian <i>Portability</i> .....	51
<b>Tabel 4.5</b> Pencapaian kualitas OWL <i>Knowledge Base Generator</i> .....	52

## ABSTRAK

Aunillah, Calvin Niam. 2021. **OWL Knowledge Base Generator Dinamis dari Konten Halaman Website dengan Metode Graph Breadth First Search**. Skripsi. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi. Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Pembimbing: (I) A'la Syauqi, M. Kom, (II) Yunifa Miftachul Arif, M. T

---

Kata Kunci: *OWL Knowledge Base, Web Scraping, BFS, crawling*

Berita adalah informasi yang banyak dicari masyarakat, di antaranya ialah berita-berita tentang kesehatan hingga berita-berita tentang pemerintahan. Informasi mengenai kebijakan-kebijakan pemerintah merupakan salah satu informasi yang dibutuhkan oleh masyarakat. Berdasarkan data survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) 2019-2020 (Q2) terhitung 196,71 juta pengguna internet di Indonesia yang mana di antaranya 13,60% pengguna mengakses layanan informasi berita. Terkadang pembaca kesulitan menangkap informasi yang mereka inginkan dikarenakan banyak sekali informasi promosi yang tersebar dari beberapa *website* dan dianggap memakan waktu serta membuat tidak nyaman pembaca, sehingga keakuratan pencarian dan pemilahan informasi sangat dibutuhkan. Pada penelitian ini digunakan metode *Graph Breadth First Search* (BFS) sebagai metode yang digunakan dalam teknik *Web Scraping*. BFS merupakan *Breadth First Search* (BFS) adalah algoritma dari struktur data *graph* yang tepat untuk proses *crawling*. Tujuan dari penelitian ini adalah untuk menghasilkan *OWL Knowledge Base Generator* sesuai dengan dokumen yang diekstrak dari halaman *website* berita. Hasil pengujian kualitas perangkat lunak *OWL Knowledge Base Generator* menghasilkan presentase *Functionality* 70%, *Reliability* 96%, *Usability* 81%, *Efficiency* 81%, *Maintainability* 50%, *Portability* 100% dan total skualitas sebesar 79.66%. Berdasarkan hasil ini, penggunaan metode *Breadth First Search* dianggap sangat baik dan efektif dalam pembuatan *OWL Knowledge Base* dinamis dari konten halaman *website*.

## ABSTRACT

Aunillah, Calvin Niam. 2021. **OWL Knowledge Base Dynamic Generator of Website Page Content with Graph Breadth First Search Method.** Undergraduate Thesis. Informatics Engineering Department, Faculty of Science and Technology. Islamic State of Maulana Malik Ibrahim Malang.

Supervisor: (I) A'la Syauqi, M. Kom, (II) Yunifa Miftachul Arif, M. T

---

Keywords: OWL Knowledge Base, Web Scraping, BFS, crawling

News is information that many people are looking for, including news about health to news about government. Information about government policies is one of the information needed by the community. Based on survey data from the Indonesian Internet Service Providers Association (APJII) 2019-2020 (Q2), there are 196.71 million internet users in Indonesia, of which 13.60% of users access news information services. Sometimes readers have difficulty capturing the information they want because a lot of promotional information is scattered from several websites and is considered time consuming and makes the reader uncomfortable, so that the accuracy of searching and sorting information is needed. In this study, the Graph Breadth First Search (BFS) method is used as the method used in the Web Scraping technique. BFS is Breadth First Search (BFS) is an algorithm of graph data structure that is appropriate for the crawling process. The purpose of this research is to produce an OWL Knowledge Base Generator according to the documents extracted from the news website pages. The results of the OWL Knowledge Base Generator software quality test resulted in the percentage of Functionality 70%, Reliability 96%, Usability 81%, Efficiency 81%, Maintainability 50%, Portability 100%, and total quality of 79.66%. Based on these results, the use of the Breadth First Search method is considered very good and effective in creating a dynamic OWL Knowledge Base from web page content.

## الملخص

عون الله، كالفن نعم. ٢٠٢١. OWL Knowledge Base المولد الديناميكي لمحتوى صفحة موقع الويب باستخدام طريقة البحث الأولى لاتساع الرسم البياني. البحث الجامعي. قسم الهندسة المعلوماتية. كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشريف الأول: (1) علاء سياوقي ماجستير الكمبيوتر (2) يونيفة مفتاح عارف تقنيات إتقان.

---

الكلمة الرئيسية: OWL Knowledge Base ،Web Scraping ،BFS، الزحف

الأخبار هي المعلومات التي يبحث عنها كثير من الناس ، بما في ذلك الأخبار المتعلقة بالصحة والأخبار المتعلقة بالحكومة. المعلومات حول سياسات الحكومة هي إحدى المعلومات التي يحتاجها المجتمع. استنادًا إلى بيانات استقصائية من جمعية مزودي خدمة الإنترنت الإندونيسية (Q2) 2019-2020 (APJII) ، هناك 196.71 مليون مستخدم للإنترنت في إندونيسيا ، منهم 13.60٪ من المستخدمين يصلون إلى خدمات المعلومات الإخبارية. في بعض الأحيان ، يواجه القراء صعوبة في النقاط المعلومات التي يريدونها لأن الكثير من المعلومات الترويجية مبعثرة من عدة مواقع ويب وتعتبر مضيعة للوقت وتجعل القارئ غير مرتاح ، لذا فإن دقة البحث وفرز المعلومات مطلوبة. في هذه الدراسة ، يتم استخدام طريقة (Graph Breadth First Search (BFS) كطريقة مستخدمة في تقنية Web Scraping. BFS هو Breadth First Search (BFS) عبارة عن خوارزمية لهيكل بيانات الرسم البياني المناسب لعملية الزحف. الغرض من هذا البحث هو إنتاج منشئ قاعدة معارف OWL وفقًا للوثائق المستخرجة من صفحات الموقع الإخباري. نتج عن نتائج اختبار جودة برنامج OWL Knowledge Base Generator النسبة المئوية للوظائف 70٪ والموثوقية 96٪ وقابلية الاستخدام 81٪ والكفاءة 81٪ وقابلية الصيانة 50٪ وقابلية 100٪ والجودة الإجمالية 79.66٪. بناءً على هذه النتائج ، يعتبر استخدام طريقة Breadth First Search جيدًا جدًا وفعالاً في إنشاء قاعدة معارف OWL الديناميكية من محتوى صفحة موقع الويب.

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Berita adalah informasi yang banyak dicari masyarakat, di antaranya ialah berita-berita tentang kesehatan hingga berita-berita tentang pemerintahan. Informasi mengenai kebijakan-kebijakan pemerintah merupakan salah satu informasi yang dibutuhkan oleh masyarakat apalagi dalam situasi pandemi. Pemanfaatan teknologi informasi memudahkan masyarakat untuk mengakses informasi terkait kebijakan-kebijakan pemerintah terkait pandemi. Selain itu dengan berbagai keunggulan lainnya: mempermudah penyebaran informasi menjadi lebih cepat, mudah dan murah tanpa ada batasan ruang, jarak, dan waktu.

Berdasarkan data survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) 2019-2020 (Q2) terhitung 196,71 juta pengguna internet di Indonesia yang mana di antaranya 13,60% pengguna mengakses layanan informasi berita (APJII, 2020). Sedangkan *Alexa Traffic Rank* (ATR) melaporkan ada tujuh berita *website* portal berita dalam daftar 10 *website* teratas dengan akses tertinggi di Indonesia, yaitu Okezone, Tribunnews, Kompas, Detik, Kumparan, Suara, dan PikiranRakyat dengan rata-rata halaman yang diakses pengunjung harian per tiga bulan di situs-situs tersebut adalah berkisar 1-4 halaman (Alexa.com, 2021).

Informasi yang terdapat dalam situs berita dibangun menggunakan *database*, sehingga dapat menampilkan berita secara paginasi. Informasi tersebut bisa didapatkan melalui *search engine*. Akan tetapi, untuk mengumpulkan informasi dengan cara mengolah data dalam basis data atau *database* sebagai media penyimpanan data akan lebih mudah dan cepat jika data yang telah dikumpulkan dan yang akan disajikan sebagai informasi diolah menggunakan *database* berbasis *Knowledge Base*. Meskipun begitu, pengunjung *website* yang ingin mendapatkan detail secara lengkap, harus mengakses halaman berita ataupun artikel satu persatu untuk mendapatkan informasi yang diinginkan.

Internet kini menjadi sumber data terbesar dan berharga bagi setiap pengguna karena di *website*, dokumen saling berhubungan dan dapat diakses melalui koneksi internet. Setiap individu dapat menambahkan konten ke halaman

*website*. Semakin banyak informasi yang ditambahkan, semakin besar ukuran *database* dan semakin sulit pula untuk menemukan informasi yang benar-benar diinginkan di halaman *website* dalam waktu yang singkat. Selain itu, menurut Junjoewong et. al (2018) informasi promosi atau iklan merupakan salah satu faktor yang mempengaruhi kesulitan menemukan informasi, di mana vendor mempublikasikan informasi promosi ke banyak saluran *website* untuk mempromosikan produknya. Oleh sebab itu, terkadang pembaca kesulitan menangkap informasi yang mereka inginkan dikarenakan banyak sekali informasi promosi yang tersebar dari beberapa *website* dan dianggap memakan waktu serta membuat tidak nyaman pembaca, sehingga keakuratan pencarian dan pemilahan informasi sangat dibutuhkan. Dengan demikian, diperlukan suatu teknologi untuk mendapatkan informasi yang benar-benar diinginkan, sehingga di sinilah peran *website* berbasis semantik. *Website* berbasis semantik merupakan sistem jaringan yang mampu memahami tidak hanya makna sebuah kata dan konsep, tetapi juga berhubungan antara keduanya, sehingga suatu *website* dapat menghasilkan informasi yang sesuai dan diinginkan oleh pengunjung *website*.

Baskoro et. al, (2018) mengatakan bahwa keunggulan *website* berbasis semantik adalah karena memiliki komponen ontologi relasi, di mana dapat mendeskripsikan kumpulan dari relasi yang mendeskripsikan semantik atau arti dari sebuah entitas. Hal ini berguna untuk menelusuri atribut dalam sebuah dokumen yang memudahkan untuk menelusuri informasi, sehingga tidak perlu lagi membuka halaman *website* satu per satu, tetapi dapat menggunakan *crawler* untuk menjelajahi halaman-halaman *website*. Agen ini secara logis hanya akan memilih hasil yang relevan sesuai dengan kebutuhan pengguna, lalu menyajikannya dalam format yang diinginkan. Saat ini, dokumen mulai disajikan dalam format *eXtensible Markup Language* (XML). Bahasa lain yang digunakan dalam mendukung *website* berbasis semantik selain XML adalah *Resource Description Framework* (RDF)/*Ontology Web Language* (OWL). RDF/OWL dengan kemampuan dan fasilitas yang dimilikinya mampu digunakan untuk merepresentasikan atribut dalam konten *website*, sehingga dapat diproses dalam sistem (Amalia et. al, 2018).

Sistem yang dihasilkan dalam penelitian ini adalah rekayasa perangkat lunak berbasis web yang bertujuan untuk menghasilkan OWL *Knowledge Base Generator* sesuai dengan dokumen yang diekstrak dari halaman *website* berita.

Teknik *Scraping* dapat diimplementasikan pada halaman *website* berita dengan mudah, sehingga mempermudah pembuatan *OWL Knowledge Base Generator*.

Pada zaman digital saat ini, hampir semua pekerjaan menggunakan sebuah sistem informasi. Pada bidang kesehatan, pendidikan, pariwisata, dan lain-lainnya. Sistem merupakan sebuah kesatuan yang di dalamnya terdapat komponen-komponen yang terhubung dan berfungsi untuk mempermudah penyaluran informasi, materi dan *energy* (Eryanto, 1999). Tujuan dibuatnya sistem ini tidak lain adalah untuk membantu meringankan beban pekerjaan. Penelitian ini ditekankan pada pembahasan pengelolaan dalam mencari informasi berita agar dapat membantu individu ataupun instansi dalam mencari informasi agar menjadi lebih efektif dan efisien. Karena Allah menghendaki kemudahan bagi hamba-Nya. Hal ini sesuai dengan firman Allah dalam Al-Qur'an pada surat Al-Baqarah ayat 185, yang berbunyi:

يُرِيدُ اللَّهُ بِكُمْ الْيُسْرَ وَلَا يُرِيدُ بِكُمْ الْعُسْرَ وَتُنَكِّلُوا الْعِدَّةَ وَلِتُكَبِّرُوا اللَّهَ عَلَىٰ مَا هَدَيْتُمْ وَلَعَلَّكُمْ تَشْكُرُونَ

“Allah menghendaki kemudahan bagimu dan tidak menghendaki kesukaran. Hendaklah kamu mencukupkan bilangannya dan mengagungkan Allah atas petunjuk-Nya yang diberikan kepadamu agar kamu bersyukur.” (Qs. Al-Baqarah/2: 185).

Merujuk pada tafsir Ibnu Katsir, penggalan ayat diatas menjelaskan tentang keringanan bagi umat yang menjalankan puasa. “Sesungguhnya diberikan keringanan ini bagi kalian hanya dalam keadaan kalian dalam perjalanan ataupun yang sedang sakit, namun puasa adalah kewajiban bagi umat yang bertaqwa. Hal ini tidak lain adalah untuk meringankan kalian atas rahmat dari Allah SWT.” (Ad-Dimasyqi: 2000). Namun, apabila ayat ini diuraikan lebih lanjut maka, yang dimaksud akan meringankan tersebut adalah bukan hanya berlaku untuk orang yang menjalankan puasa saja, melainkan untuk hal lainnya juga selama itu dalam kebaikan. Seperti halnya dalam menyebarkan kabar atau informasi yang bisa membantu dan memberikan kemudahan bagi umat manusia.

Dalam penelitian ini peneliti juga menekankan betapa pentingnya pengetahuan bagi kelangsungan hidup manusia. Karena dengan pengetahuan, manusia dapat mengetahui apa yang benar dan yang salah, yang baik dan yang buruk, yang membawa manfaat dan yang membawa mudharat. Adapun ayat Al-

Qur'an yang menjelaskan tentang pentingnya sebuah ilmu pengetahuan terdapat dalam surat Al-A'raf ayat 52, yang berbunyi:

وَلَقَدْ جِئْنَاهُمْ بِكِتَابٍ فَصَّلْنَاهُ عَلَىٰ عِلْمٍ هُدًى وَرَحْمَةً لِّقَوْمٍ يُؤْمِنُونَ

“Sungguh, Kami telah mendatangkan kepada mereka Kitab (Al-Qur'an) yang telah Kami jelaskan secara terperinci atas dasar pengetahuan sebagai petunjuk dan rahmat bagi orang-orang yang beriman.” (Qs. Al-A'raf/7: 52).

Data-data yang terdapat dalam *website* disimpan dalam sebuah basis data, sehingga dalam mengumpulkan data diperoleh informasi perhalaman atau juga bisa didapatkan melalui mesin pencari atau *search engine*. Berdasarkan penjelasan di atas dapat disimpulkan bahwa *OWL Knowledge Base* sangat penting adanya dalam penyediaan informasi, sehingga akan memudahkan bagi para *developer website* untuk mengumpulkan semua data yang berada pada suatu halaman *website* akan dimasukkan ke dalam *Knowledge Base* akan membutuhkan waktu yang cukup lama apabila pengambilan data tersebut dilakukan secara manual dengan mengambil data satu persatu. Oleh sebab itu, peneliti rasa sangat kurang efisien, sehingga dibutuhkan sebuah metode yang dapat mengambil konten halaman *website* secara dinamis dan otomatis.

Salah satu teknik yang bisa diterapkan adalah dengan mengekstraksi konten pada dokumen HTML yang ada di dalam halaman *website*. Teknik tersebut dilakukan dengan tujuan untuk mengambil data-data yang dibutuhkan dari halaman HTML dengan merepresentasikan dokumen HTML sebagai struktur seperti pohon di mana simpul daun merepresentasikan sebuah data (Li et. al, 2017). Proses tersebut juga dikenal dengan sebutan *Web Scraping*. Proses pengambilan data menggunakan teknik tersebut dapat dilakukan dengan dinamis, tanpa harus mengambil data dengan cara manual, yaitu mengakses halaman *website* melalui browser. Oleh karena itu, berdasarkan penjelasan mengenai latar belakang di atas peneliti mengangkat judul “**OWL Knowledge Base Generator Dinamis dari Konten Halaman Website dengan Metode Graph Breadth First Search**”.

## 1.2. Pernyataan Masalah

Seberapa tinggi nilai *functionality*, *reliability*, *usability*, *efficiency*, *maintainability*, dan *portability* dari sistem *OWL Knowledge Base Generator* untuk membangun *Knowledge Base* dengan format RDF?

### **1.3. Tujuan Penelitian**

Mengukur *functionality, reliability, usability, efficiency, maintainability, dan portability* dari sistem *OWL Knowledge Base Generator* untuk membangun *Knowledge Base* dengan *output* yang berformat RDF.

### **1.4. Batasan Penelitian**

Agar pembahasan penelitian ini tidak menyimpang, maka diperlukan batasan-batasan. Batasan-batasan dalam penelitian ini adalah sebagai berikut.

1. Sistem yang dibangun menggunakan *Express Framework*.
2. Proses dalam penelitian ini terbagi menjadi dua yaitu pengambilan data pada *website* berita dengan teknik ekstraksi data dari halaman *website* dan *generate Knowledge Base* sebagai *output* sistem.
3. Penelitian berfokus pada proses ekstraksi data berita dalam *Knowledge Base* dengan *output* berformat RDF.
4. Data yang diekstrak berasal dari *website* berita Detik.com dan Kompas.com.
5. Pengukuran kualitas perangkat lunak *OWL Knowledge Base Generator* berbasis ISO 9126.

### **1.5. Manfaat Penelitian**

Diharapkan penelitian ini dapat memberikan manfaat bagi ilmu pengetahuan dalam bidang pengembangan *website* manajemen *Knowledge Base* dimana nantinya hasil dari sistem dapat digunakan dalam sebuah aplikasi multi platform.

### **1.6. Sistematika Penulisan**

#### **BAB I PENDAHULUAN**

Pada Bab pendahuluan menjelaskan kepada pembaca dalam menjawab pertanyaan tentang apa yang diteliti dan untuk apa serta mengapa penelitian ini dilakukan yang termuat dalam sub bab berikut.

### 1.1. Latar Belakang

Latar belakang berisi tentang argumen yang melatari dilakukannya penelitian ini yaitu tentang membuat *Knowledge Base* dengan mengumpulkan konten dari halaman *website* menggunakan teknik ekstraksi data.

### 1.2. Pernyataan Masalah

Sub bab ini berfokus pada permasalahan yang diteliti dalam penelitian ini tentang bagaimana melakukan ekstraksi konten dari halaman *website* dan *generate* file berformat RDF.

### 1.3. Tujuan Penelitian

Sub bab ini menjelaskan tujuan penelitian atau hasil akhir penelitian ini yaitu tentang bagaimana melakukan ekstraksi konten dari halaman *website* dan *generate* file berformat RDF.

### 1.4. Batasan Penelitian

Batasan masalah pada penelitian ini dibatasi dalam ruang lingkup tertentu yang dijelaskan dalam sub bab ini.

### 1.5. Manfaat Penelitian

Manfaat dari hasil penelitian ini adalah memudahkan pengembangan *website* manajemen *Knowlede Base* lebih lanjut yang bisa diimplementasikan dalam berbagai platform.

### 1.6. Sistematika Penulisan

Sub bab ini menjelaskan orientasi atau maksud dari masing-masing bab atau sub bab.

## **BAB II TINJAUAN PUSTAKA**

Bab ini menjelaskan tentang teori dasar yang digunakan sebagai landasan dalam penulisan tugas akhir ini dan sebagai referensi serta acuan dalam pembuatan tugas akhir.

### 2.1. *Web Semantic*

Sub bab yang meliputi pengertian *Web Semantic* serta fungsi- fungsinya secara menyeluruh.

### 2.2. *Knowledge Base*

Pada sub bab ini dijelaskan mengenai segala hal yang terkait dengan *Knowledge Base* serta apa fungsinya.

### 2.3. *Resource Description Framework* (RDF)

Menjelaskan tentang pengertian ataupun manfaat dari *Resource Description Framework* (RDF) untuk membangun sebuah *Knowledge Base* dan juga bagaimana bentuk serta skemanya.

### 2.4. *Web Scraping*

Pembahasan tentang teknik yang digunakan dalam penelitian untuk ekstraksi data dari halaman *website*.

### 2.5. Node Js

Pengertian tentang platform perangkat lunak untuk mengembangkan aplikasi berbasis web yang ditulis dalam sintaks bahasa pemrograman JavaScript.

### 2.6. Metode *Breadth First Search* (BFS)

Pembahasan tentang metode yang digunakan dalam penelitian untuk pengambilan data pada *website*.

### 2.7. ISO 9126

Pada sub bab berikut dijelaskan tentang pengertian ISO 9126 serta apa saja indikator karakteristiknya.

### 2.8. Analisis *Function Point*

Sub bab yang meliputi pengertian *Function Point* serta fungsinya sebagai pengukur sebuah perangkat lunak.

## **BAB III METODE PENELITIAN**

Dalam bab Metode Penelitian ini menjelaskan sub bab tentang metode yang diimplementasikan dalam menyelesaikan penelitian dan langkah-langkah yang dilakukan dalam penelitian pada sistem yang dibangun.

### 3.1. Desain Sistem

Pada sub bab berikut ini diperlihatkan desain sistem penelitian untuk mengetahui gambaran dan rancangan dari sistem.

### 3.2. Perancangan Sistem

Pada sub bab ini dijelaskan secara detil tentang desain, alur dan proses pengambilan data, serta pembuatan *Knowledge Base*.

### 3.3. Penelusuran Halaman *Website*

Sub bab yang menggambarkan bagaimana penelusuran halaman *website* dilakukan.

### 3.4. *Crawling*

Pada sub bab ini menjelaskan proses penelusuran dengan menggunakan metode *Breadth First Search* untuk memindai konten halaman *website*.

### 3.5. *Scraping*

Pada sub bab ini menjelaskan proses pengambilan konten pada halaman *website* yang mengacu pada teknik ekstraksi data.

### 3.6. *Generate File RDF*

Sub bab ini menjelaskan proses pembuatan *file* RDF dengan hasil ekstraksi data.

### 3.7. Kerangka Kerja Konseptual

Pada sub bab ini menampilkan dan menjelaskan kerangka kerja konseptual penelitian.

## **BAB IV UJI COBA DAN PEMBAHASAN**

Bab uji coba dan pembahasan ini berisi uji coba penelitian yang dilakukan sesuai desain yang telah dengan dirancagn.

### 4.1. Implementasi

Dalam sub bab berikut ini dijabarkan mengenai implementasi dari uji coba yang telah dibuat dan permasalahan selama penelitian dalam pembangunan dan penerapan sistem.

### 4.2. Pengujian Sistem

Sub bab yang berisikan proses pengujian penelitian setelah semua proses dalam sistem selesai dilakukan.

## **BAB V PENUTUP**

Dalam bab ini berisi kesimpulan dan saran berdasarkan penelitian yang dilakukan.

### 5.1. Kesimpulan

Sub bab ini menjawab dari pernyataan masalah yang telah ditetapkan sebelumnya.

## 5.2. Saran

Sub bab ini menyampaikan saran berdasarkan hasil penelitian yang dilakukan untuk pengembangan sistem lebih lanjut.

### **DAFTAR PUSTAKA**

Sub bab ini menyantumkan seluruh referensi atau bahan rujukan yang digunakan dalam penulisan tugas akhir ini.

## **BAB II KAJIAN PUSTAKA**

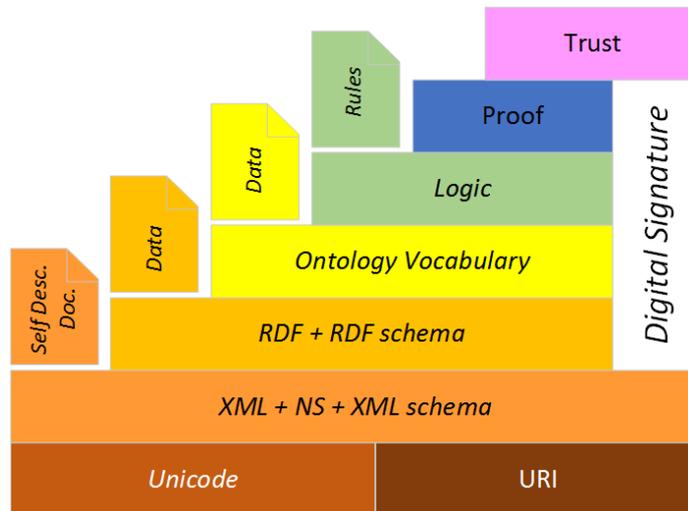
Pada bab ini membahas mengenai teori-teori yang menjadi dasar penelitian yang dilakukan. Teori-teori tersebut antara lain *Web Semantic*, *Knowledge Base*,

*Resource Description Framework (RDF), Simple Protocol and RDF Query Language (SPARQL), Web Scraping, dan Node Js.*

### **2.1. Web Semantic**

*Web Semantic* atau web semantik merupakan salah satu perkembangan pada aplikasi berbasis web. Menurut bahasa, web semantik mempunyai arti yaitu web yang memiliki makna atau dengan kata lain, web semantik merupakan suatu aplikasi berbasis web yang mempunyai *Knowledge Base* tertentu sehingga bisa dikatakan web semantik mempunyai sifat lebih pintar dari web sebelumnya. Web semantik adalah sebuah evolusi dan perluasan dari web yang telah ada, yang memungkinkan komputer untuk memanipulasi data dan informasi (Nandini, 2014). Web semantik sering disebut sebagai web versi 3.0.” Web semantik memungkinkan data tidak hanya ditujukan kepada manusia sebagai pembaca. Akan tetapi, juga ditujukan kepada mesin atau komputer agar bisa diproses.

Web semantik terdiri atas 2 buah kata yang masing-masing memiliki pengertian yang cukup berbeda. Web di sini dimaksudkan sebagai jaringan komputer yang luas yaitu *World Wide Web (WWW)*. Sedangkan semantik dapat diartikan sebagai ilmu tentang makna atau tentang arti, yaitu salah satu dari tiga tataran analisis bahasa yaitu fonologi, gramatika dan semantik. Oleh sebab itu, web semantik adalah kemampuan aplikasi komputer yang berfungsi untuk lebih memahami bahasa manusia dan bukan hanya bahasa yang baku dari para penggunanya tetapi juga bahasa yang lebih kompleks seperti dalam bahasa percakapan, sehingga memudahkan penggunanya untuk berkomunikasi dengan mesin (Satria, 2016). Arsitektur dari *Web Semantic* dapat dilihat pada Gambar 2.1.



**Gambar 2.1** Arsitektur *Web Semantic*

Menurut Satria (2016), arsitektur *Web Semantic* di atas dijelaskan sebagai berikut.

1. *URI dan Unicode*

Digunakan untuk mengidentifikasi dan menempatkan *resource*. URI dianggap sebagai dasar dari web yang digunakan untuk memberikan nama unik untuk setiap *resources*. *Unicode* merupakan standar representasi karakter komputer.

2. *Extensible Markup Language (XML)*

Merupakan bahasa *markup* yang dapat dimengerti oleh mesin dan mempunyai format penulisan tersendiri. XML memiliki format teks yang *flexible* dan didesain untuk menggambarkan data serta berperan penting dalam pertukaran jenis data yang berbeda pada web. Setiap dokumen XML diawali dengan pendeklarasian *namespace* menggunakan XML *namespace*.

3. *Resource Description Framework (RDF)*

Merupakan lapisan pertama dari web semantik. RDF merupakan kerangka dalam menggunakan dan merepresentasikan metadata dan menggambarkan informasi semantik mengenai *resource* web yang mudah diakses mesin atau komputer. RDF menggunakan URI untuk mengidentifikasi *resource* dengan menggunakan model grafik.

4. *Ontology Vocabulary*

Merupakan bahasa yang menyediakan kosa kata umum dan tata bahasa untuk publikasi data serta deskripsi semantik dari data yang digunakan

untuk mengamankan dan menjaga *ontology* tersebut agar siap untuk *inference*. *Ontology* menggambarkan data semantik dan menyediakan data yang seragam yang memungkinkan komunikasi dapat dimengerti oleh pihak manapun.

5. *Logic dan proof*

Pembangunan sistem pada web semantik mengikuti logika pada struktur *ontology*. Sebuah *reasoned* dapat digunakan untuk memeriksa dan menyelesaikan permasalahan konsistensi dan redundansi dari terjemahan kelas.

6. *Trust*

Merupakan lapisan terakhir pada web semantik. Komponen ini berkaitan dengan kepercayaan informasi pada web untuk memberikan jaminan kepercayaan terhadap informasi tersebut.

## **2.2. Knowledge Base**

*Knowledge Base* adalah tempat basis data informasi terkait tentang subjek tertentu. Pada kaitannya dengan Teknologi Informasi (TI), *Knowledge Base* adalah sumber data yang dapat dibaca mesin untuk penyebaran informasi, umumnya secara *online*. Cara untuk mendukung pengambilan keputusan dan mengatasi masalah manajemen data adalah dengan *Knowledge Base*, di mana *knowledge* dikodekan dalam bahasa formal yang semantiknya dapat diproses oleh sistem komputer (Sanfillippo, 2019). *Knowledge Base* telah dikembangkan untuk menyimpan, menalar, dan mendapatkan informasi *web semantik*. Beberapa *Knowledge Base* mendukung RDF/RDFS dan OWL. Penelitian ini mempertimbangkan masalah bagaimana memilih *Knowledge Base* yang sesuai untuk OWL. Menurut Costa dan Monteiro (2016), perusahaan dapat mengalami keunggulan kompetitif berkelanjutan ketika mereka menerapkan *knowledge* dalam produk dan layanan baru dan signifikan yang ditingkatkan, praktik organisasi, proses produksi, strategi pemasaran dan inovasi.

## **2.3. Resource Description Framework (RDF)**

RDF adalah bahasa representasi pengetahuan yang didedikasikan untuk penjelasan dokumen dan lebih umum sumber daya dalam web semantik. RDF

Merupakan infrastruktur yang memungkinkan pertukaran *encoding* dan penggunaan kembali metadata terstruktur. Struktural RDF menyediakan metode yang tidak ambigu dalam mengekspresikan semantik. RDF menyediakan model untuk mendeskripsikan dokumen. Dokumen memiliki properti (atribut atau karakteristik). RDF mendefinisikan dokumen sebagai objek yang diidentifikasi secara unik menurut tipe properti, dan tipe properti memiliki nilai yang sesuai.

RDF adalah aplikasi XML. Salah satu penggunaan XML adalah pertukaran data elektronik antara dua atau lebih sumber data di web. Data elektronik terutama ditujukan untuk konsumsi komputer. Misalnya, robot penelusuran dapat mengintegrasikan informasi secara otomatis dari sumber terkait yang mempublikasikan datanya dalam format XML. Peluang baru akan muncul bagi pihak ketiga dengan mengintegrasikan, mengubah, membersihkan, dan menggabungkan data XML (Deutsch et. al, 1999). Contoh dari struktur XML dapat dilihat pada Gambar 2.2.

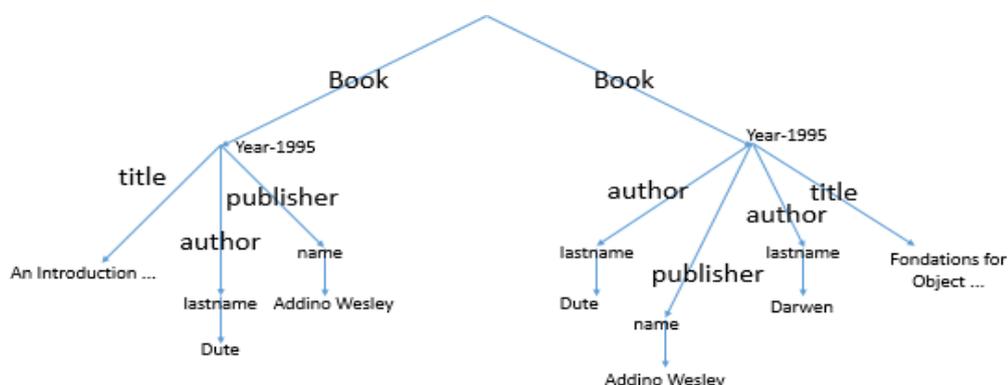
```

<?xml version="1.0" encoding="UTF-8"?>
- <note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this weekend!</body>
</note>

```

Gambar 2.2 Contoh struktur XML

Menurut Deutsch et. al (1999) model data XML dapat dibedakan menjadi *ordered data model* dan *unordered model*. Contoh *Unordered Graph XML* dapat dilihat pada Gambar 2.3.

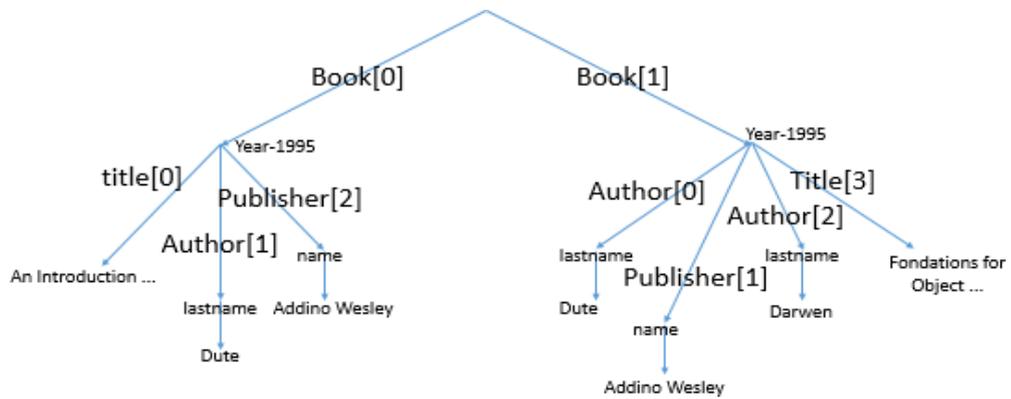


Gambar 2.3 Contoh *unordered graph XML*

- Setiap *node* diwakili oleh string unik yang disebut pengidentifikasi objek.
- Tepi (*edge*) *graph* diberi label dengan tag elemen.

- *Node graph* diberi label dengan set pasangan atribut-nilai.
- Daun (*leaves*) graph diberi label dengan satu nilai *string*.
- Graph memiliki *root node*.

Sedangkan *Ordered Graph XML* digambarkan dan didefinisikan sebagai berikut.



**Gambar 2.4** Contoh *ordered graph XML*

Pada Gambar 2.4 merupakan contoh *Ordered Graph XML* di mana *graph XML* ini memiliki urutan untuk atributnya. Pada model terurut, boleh jadi memiliki banyak tepi atau *edge* dengan sumber yang sama, label tepi yang sama, dan nilai tujuan yang sama.

```

<?xml version="1.0"?>

<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:cd="http://www.recshop.fake/cd#">

<rdf:Description
rdf:about="http://www.recshop.fake/cd/Empire Burlesque">
<cd:artist>Bob Dylan</cd:artist>
<cd:country>USA</cd:country>
<cd:company>Columbia</cd:company>
<cd:price>10.90</cd:price>
<cd:year>1985</cd:year>
</rdf:Description>

<rdf:Description
rdf:about="http://www.recshop.fake/cd/Hide your heart">
<cd:artist>Bonnie Tyler</cd:artist>
<cd:country>UK</cd:country>
<cd:company>CBS Records</cd:company>
<cd:price>9.90</cd:price>
<cd:year>1988</cd:year>
</rdf:Description>
.
.
.
</rdf:RDF>

```

**Gambar 2.5** Contoh struktur RDF

RDF memiliki infrastruktur dalam membangun sebuah ontologi. Infrastruktur tersebut terdiri atas subjek, predikat, dan objek yang menyatakan hubungan antara data atau *Linked Data*. RDF menggunakan format XML. Format XML mempresentasikan model data pada RDF. RDF menjelaskan tentang deskripsi yang akan dijelaskan. Gambar 2.5 menunjukkan Struktur data RDF yang diambil dari konten halaman *website*.

#### **2.4. Web Scraping**

*Web scraping* atau pengikis web adalah teknik yang digunakan untuk mengekstrak informasi dari situs web. Teknik ini mengubah data tidak terstruktur menjadi bentuk terstruktur. Data yang telah diekstrak ini dapat disimpan di mana saja di komputer lokal atau server jarak jauh. Proses ini seringkali efisien, lebih cepat, dan tidak terlalu rentan terhadap kesalahan saat otomatis. Web scraping dapat dibangun menggunakan teknologi apa pun seperti PHP, Node JS, dan Python dll.

Pengetahuan tentang struktur skrip HTML dan format data seperti JavaScript *Object Notation* (JSON) sangat penting untuk membangun perangkat lunak *Web Scraping* yang andal.

Data dari internet biasanya berada di halaman *website* menggunakan format HTML, dengan sebagian isi halaman tersebut merupakan informasi untuk pengguna manusia seperti tombol navigasi, pencarian gambar dan *layout* untuk memperindah dan mempermudah halaman *website* untuk dibaca. Beberapa kasus teknologi *Web Scraping* mungkin satu-satunya solusi yang bisa diterapkan ketika ingin mengumpulkan data secara besar (Mahajan & Kumar, 2015).

Menurut Vishwakarma et. al (2019), membangun *Web Scraping* tidaklah mudah. Beberapa masalah yang sering dihadapi selama ekstraksi data dari gambar. Misalnya, tidak dapat mengekstrak teks dengan benar karena berbagai karakteristik gambar seperti teks dengan efek bayangan. Proses ekstraksi data dilakukan untuk mendapatkan informasi dengan menggunakan bermacam-macam teknik. Teknik *scraping* dapat dilakukan diantaranya dengan menggunakan analisa *Document Object Model* (DOM) dan dengan menggunakan teknik pemrograman regular ekspresi.

*Document Object Model* (DOM) adalah platform dokumen, yang diatur oleh beberapa hierarki *node* atau terdiri dari beberapa segmen informasi yang ditetapkan, sehingga dianggap sebagai pohon, yang disebut sebagai Dom-Tree. *Browser* dapat mengenali setiap halaman *website* sebagai Dom-Tree, yang mencerminkan informasi internal halaman *website* dan sangat sering digunakan dalam pencarian informasi dalam suatu *website*. Misalnya, topik paragraf yang memiliki lebih banyak *node* tetangga bisa lebih spesifik daripada yang lain yang memiliki lebih sedikit *node* tetangga. Sementara jika sebuah *node* ataupun *edge* memiliki banyak paragraf anak, topiknya akan abstrak. Dom-Tree digunakan dalam penyaringan konten halaman *website* dan dapat digunakan untuk menyaring topik halaman *website* menggunakan jumlah *node* tetangga dan hubungan *node* terdekat serta paragraf teks terdekat di Dom-Tree.

Le dan Pishva (2015), menggunakan teknik *Web Scraping* dan layanan Google API untuk mengumpulkan data terkait pendistribusian barang. Penelitian tersebut menunjukkan teknik *Web Scraping* yang digunakan dapat mengukur

akurasi dan performa pendistribusian barang, sehingga dapat memecahkan masalah manajemen pendistribusian yang dilakukan secara maksimal.

Julian et. al (2015), melakukan penelitian tentang *Web Scraping* yang digunakan untuk mendapatkan rekomendasi perakitan komputer yang memungkinkan pencarian harga terendah dan pembelian komponen komputer pada lima *website* toko komputer *online*. Peneliti menyimpulkan bahwa teknik *Web Scraping*, membantu pengguna untuk mendapatkan rekomendasi produk untuk perakitan sebuah PC.

Sundaramoorthy et. al (2017), membuat sistem yang dapat mengambil url dari *website* portal berita. Penelitian tersebut bertujuan untuk mengelompokkan artikel berita berdasarkan kategori. Peneliti menyimpulkan bahwa teknik *Web Scraping* dapat memudahkan pembaca untuk mengakses berita yang sesuai dengan minat ataupun bidang pembaca.

Menurut Rosario (2017), *Web Scraping* mempunyai beberapa langkah-langkah sebagai berikut.

1. Membuat *template scraping*

Proses ini melakukan observasi terhadap dokumen HTML *website* yang akan diambil informasinya atau pada dikenai proses *scraping*. Caranya adalah dengan melakukan tag HTML untuk mengapit informasi yang akan diambil.

2. Eksplorasi Navigasi Situs

Proses ini melakukan menelusuri navigasi pada *website* yang akan diambil informasinya atau dikenai *scraping* untuk ditirukan pada aplikasi *web scraper* yang dibuat.

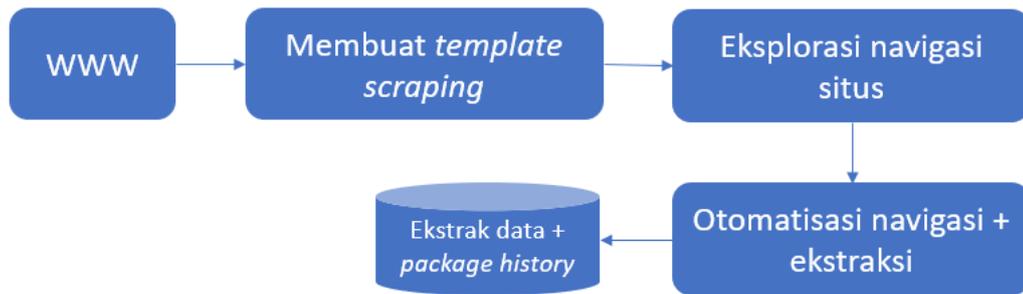
3. Mengotomatisasi navigasi dan mengekstraksi informasi

Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, aplikasi *web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang ditentukan.

4. Ekstraksi data dan menyimpan histori

Informasi yang didapat dari langkah 3 disimpan dalam tabel di dalam database.

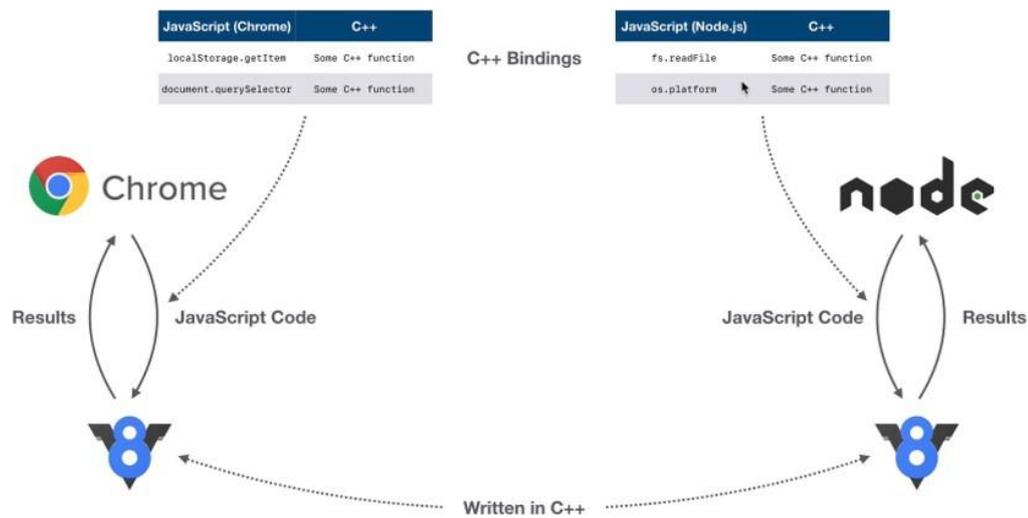
Adapun cara kerjanya dapat dilihat pada Gambar 2.6.



Gambar 2.6 Ilustrasi cara kerja Web Scraping

## 2.5. Node Js

Node js adalah platform perangkat lunak JavaScript yang didasarkan pada implementasi runtime yang diberi nama "V8" (Kyriakou et. al, 2019). V8 adalah implementasi C ++ dari sistem runtime untuk JavaScript dan bahasa pemrograman WebAssembly (). V8 dan Node Js sebagian besar diimplementasikan dalam C dan C ++, dengan fokus pada kinerja dan konsumsi memori yang rendah. V8 mendukung sebagian besar JavaScript di browser (terutama, Google Chrome). Keterkaitan antara penggunaan V8 engine pada Node Js dapat dilihat pada Gambar 2.7.



Gambar 2.7 V8 JavaScript engine

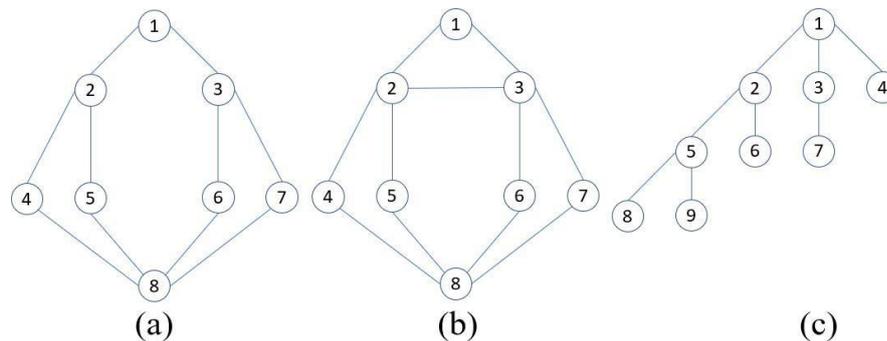
## 2.6. Metode Breadth First Search (BFS)

*Breadth First Search* (BFS) adalah algoritma dari struktur data *graph* yang melakukan pencarian secara melebar dengan mengunjungi *node* secara *pre-order*, yaitu mengunjungi suatu *node* kemudian mengunjungi semua *node* yang bertetangga dengan *node* tersebut terlebih dahulu. Kemudian, *node* yang belum

dikunjungi dan bertetangga dengan setiap *node* yang tadi dikunjungi, dst. Langkah-langkah penyelesaian menggunakan metode *Breadth First Search* adalah sebagai berikut.

1. Memasukkan *node* ke dalam antrian.
2. Mengunjungi *node* dari awal antrian, lalu memeriksa apakah *node* tersebut merupakan target.
3. Jika *node* merupakan target, maka pencarian selesai dan hasil dikembalikan.
4. Jika *node* bukan target, maka memasukkan seluruh *node* yang bertetangga dengan *node* ke dalam antrian.
5. Jika antrian kosong dan setiap *node* sudah diperiksa, maka pencarian selesai dan mengembalikan hasil target tidak ditemukan.
6. Mengulangi pencarian dari langkah kedua.

Contoh dari implementasi metode *Breadth First Search* dalam *graph* dapat dilihat pada Gambar 2.8.



**Gambar 2.8** *Graph Breadth First Search*

Di mana urutan *node* dikunjungi adalah sebagai berikut.

- *Graph* (a): 1, 2 3, 4, 5, 6, 7, 8.
- *Graph* (b): 1, 2 3, 4, 5, 6, 7, 8.
- *Graph* (c): 1, 2 3, 4, 5, 6, 7, 8, 9.

Pada algoritma *Breadth First Search*, semua *node* pada *level*  $n$  akan dikunjungi terlebih dahulu sebelum mengunjungi *node* pada *level*  $n+1$ . Pencarian dimulai dari *node* akar terus ke *level* 1 dari kiri ke kanan, kemudian berpindah ke *level* berikutnya demikian pula dari kiri ke kanan hingga ditemukan targetnya. Adapun *pseudocode* algoritma *Breadth First Search* dapat dilihat pada Gambar 2.9.

```

function BFS(node) is
    Deklarasi Q sebagai a      ntrian
    Tandai root sudah dikunjungi
    while Q tidak kosong do
        v := Q.dequeue()
        if v adalah target then
            return v
        foreach tetangga dari v sampai w in
nodes.adjacentEdges(v) do
            if w belum dikunjungi then
                Tandai w sudah dikunjungi
                Q.push(w)

```

**Gambar 2.9** Pseudocode algoritma *Breadth First Search*

Josi dan Abdillah (2014), juga menerapkan salah satu algoritma dari struktur data *graph* yaitu algoritma *Depth First Search* (DFS) untuk mengumpulkan informasi mengenai artikel/jurnal ilmiah pada portal Google Cendekia (Google *Scholar*). Cara kerja algoritma sedikit berbeda dengan algoritma BFS. Yaitu DFS berprioritas pada kedalaman *node*, sehingga dimulai dari *child node* kemudian mengunjungi *node* semakin dalam sampai target *node* ditemukan. Sedangkan BFS, dimulai dari akar mengunjungi semua *node* tetangga terdekat. Pada penelitian, peneliti menyimpulkan bahwa teknik *Web Scraping* sangat efektif untuk mendapatkan artikel/jurnal yang kemudian disimpan secara otomatis hasil scraping ke dalam *database*.

## 2.7. ISO 9126

ISO-9126 merupakan standard pengukuran kualitas perangkat lunak. Ide dasarnya adalah menentukan dan mengevaluasi produk perangkat lunak untuk mengetahui kualitas internal dan kualitas eksternal serta hubungannya dengan atribut kualitas. Konten utamanya adalah representasi kualitas perangkat lunak baik yang terlihat oleh pengguna perangkat lunak yang terdiri dari enam karakteristik yang didefinisikan dalam standar sebagai blok bangunan kualitas produk perangkat lunak (Supriyono et. al, 2019). Keenam karakteristik tersebut di antaranya adalah sebagai berikut.

- a. *Functionality*, “kemampuan dalam perangkat lunak digunakan untuk penyediaan fungsi-fungsi yang memenuhi kebutuhan user dengan dinyatakan atau tersirat”;
- b. *Reliability*, “kemampuan suatu produk perangkat lunak (software) dalam mempertahankan tingkat kinerjanya pada kondisi tertentu yang telah ditetapkan pada periode waktu yang ditentukan”;
- c. *Usability*, “seperangkat atribut yang mengukur persepsi utilitas sistem dan kepuasan untuk seperangkat pengguna yang dinyatakan atau yang tersirat”;
- d. *Efficiency*, “kemampuan software dalam pemberian kinerja yang tepat, relatif pada jumlah sumber daya yang digunakan”;
- e. *Maintainability*, “seperangkat atribut yang mendukung upaya yang diperlukan untuk membuat modifikasi yang ditentukan”;
- f. *Portability*, “kemampuan produk software yang dapat dikirim dari satu lingkup ke lingkup lain”.

## 2.8. Analisis *Function Point*

Konsep *Function Point* diperkenalkan oleh Alan J. Albrecht dari IBM pada tahun 1979. Pada tahun 1984, *International Function Point Users Group* (IFPUG) dibentuk untuk memperjelas aturan, menetapkan standar, dan mempromosikan penggunaan dan evolusinya (Behrens, 1983). Analisis *Function Point* digunakan untuk mengukur perangkat lunak berdasarkan kegunaan produk. Menurut Kumawat & Sharma (2018), ada dua keuntungan analisis *Function Point* antara lain:

1. Dapat dihitung pada fase awal pengembangan perangkat lunak.
2. Teknik *Function Point* membangun komunikasi yang efektif antara pelanggan dan pengembang perangkat lunak.

Dalam Analisis *Function Point*, terdapat 5 fungsi sebagai parameter pengukuran sebuah perangkat lunak, yaitu *inputs*, *outputs*, *inquires*, *files*, dan *interfaces* (Sangeetha & Dalal, 2015).

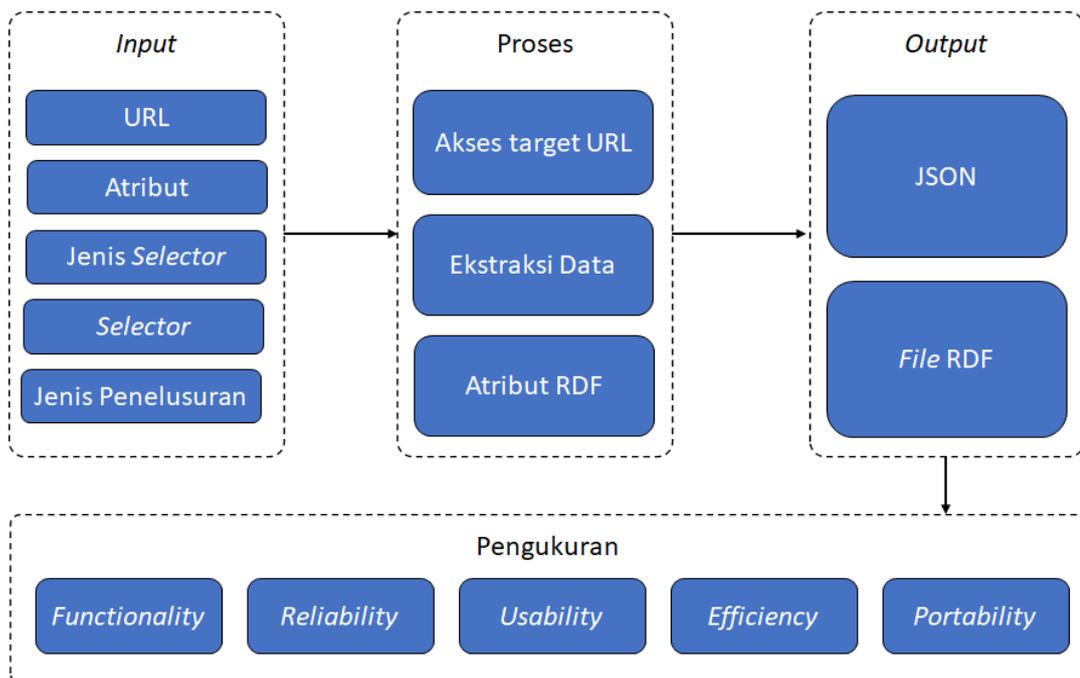
- *Inputs* adalah proses dasar yang memproses data dan informasi kontrol yang datang dari luar batasan aplikasi.

- *Outputs* adalah sebuah proses dasar dimana hasil data dilewatkan dari dalam ke keluar dari batasan aplikasi.
- *Inquiries* fungsi utamanya adalah menyediakan informasi ke user melalui pengambilan/pemrosesan data atau informasi kontrol.
- *Files* adalah kelompok data atau kelompok informasi kontrol yang digunakan dalam aplikasi.
- *Interfaces* adalah kelompok data berelasi atau informasi kontrol yang dirujuk oleh aplikasi, tapi dipelihara oleh aplikasi lain.

## BAB III PEMBAHASAN

### 3.1. Desain Sistem

Desain sistem berperan untuk mengetahui gambaran dari rangkaian proses dalam sistem yang dibuat. Gambaran dari desain sistem ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Desain sistem

Secara berurutan Gambar 3.1 menjelaskan rangkaian proses pada sistem yang dibuat. Tahap *Input* merupakan tahap penginputan parameter-parameter yang dibutuhkan untuk proses ekstraksi data, mencakup URL halaman *website*, atribut hingga *selector* yang akan ditentukan. Kemudian setiap parameter digunakan untuk proses pengambilan konten halaman *website*. Parameter-parameter tersebut antara lain adalah sebagai berikut.

1. URL

Parameter URL merupakan masukan dari pengguna untuk alamat halaman *website* yang akan diekstrak kontennya.

2. Atribut

Atribut adalah parameter yang digunakan untuk menentukan data apa saja yang ingin diambil dari halaman *website*. Misalnya, judul, *author* dan detail artikel berita.

### 3. Jenis *Selector*

Jenis *selector* adalah parameter yang digunakan sebagai kata kunci dan simbol untuk menyeleksi atau memilih tag HTML yang sedang ditelusuri. Jenis *selector* ini berdasarkan *selector* di *Cascading Style Sheet* (CSS). Misalnya, selektor tag, selektor *class*, dll.

### 4. *Selector*

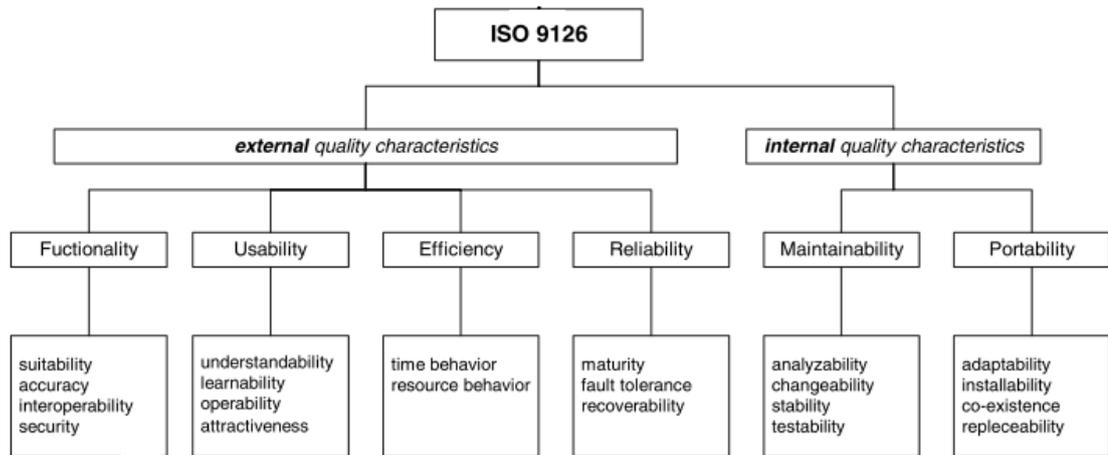
Parameter ini merupakan masukan untuk memilih elemen HTML berdasarkan nama tag atau *class* yang ingin diekstrak datanya berdasarkan jenis *selector* yang telah dipilih pada parameter sebelumnya yaitu jenis *selector*.

### 5. Jenis Penelusuran

Parameter ini digunakan untuk menentukan apakah elemen pertama atau semua elemen HTML yang akan diekstrak datanya setelah tag HTML yang diinginkan berhasil ditemukan.

Tahap selanjutnya adalah proses ekstraksi data yang diawali dengan mengakses halaman *website* kemudian mulai mengekstrak data dari halaman *website*. Setelah proses berhasil dijalankan *user* akan dapat mengunduh *file* RDF dari hasil ekstraksi konten halaman *website* sebagai *output* atau keluaran dari sistem.

Dalam penelitian ini, peneliti menggunakan enam karakteristik yang didasarkan pada ISO 9126 untuk mengukur kualitas perangkat lunak dari sistem yang dibuat. Kelima karakteristik tersebut yaitu *functionality*, *reliability*, *usability*, *efficiency*, *maintainability*, dan *portability*. Keenam komponen tersebut digunakan untuk merepresentasikan kualitas dari perangkat lunak seperti dimodelkan pada Gambar 3.2 berikut.



**Gambar 3.2** Model kualitas perangkat lunak berdasarkan ISO 9126

Sistem yang dibangun peneliti tidak dapat langsung diukur. Jadi, sebelum melakukan pengukuran, indikator karakteristik ISO 9126 dihitung terlebih dahulu dengan perhitungan *function point*. Perhitungan *function point* membutuhkan parameter berupa *inputs*, *outputs*, *inquires*, *files*, dan *interfaces*. Parameter-parameter tersebut didapatkan dari klasifikasi seperti yang ditunjukkan pada Tabel 3.1 berikut.

**Tabel 3.1** Tabel klasifikasi parameter *Function Point*

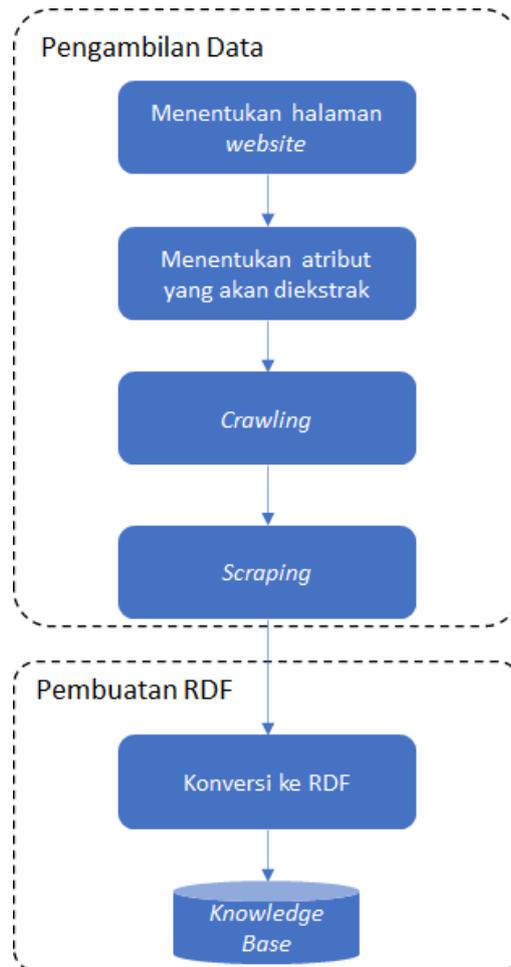
<b>Nama Parameter</b>	<b>Item</b>	<b>Jumlah</b>
<i>Inputs</i>	URL	5
	Atribut	
	Jenis <i>selector</i>	
	<i>Selector</i>	
	Jenis penelusuran	
<i>Outputs</i>	JSON	2
	<i>File</i> RDF	
<i>Inquires</i>	<i>Login</i>	13
	<i>Logout</i>	
	Menu dasbor	
	Menu <i>entry</i> ekstraksi data ( <i>scraping</i> )	
	Unduh <i>file</i> RDF	
	Riwayat hasil ekstraksi data	
	Edit hasil ekstraksi data	

	Hapus Hasil Ekstraksi Data	
	Tampil data <i>user</i>	
	Tambah <i>user</i>	
	Edit <i>user</i>	
	Hapus <i>user</i>	
	Import data <i>user</i>	
<i>Files</i>	Tabel <i>scrapes</i>	2
	Tabel <i>users</i>	
<i>Interfaces</i>	<i>File excel</i> data <i>user</i>	1
<i>Failures</i>	<i>Request Timeout</i> ketika mengakses halaman <i>website</i>	3
	TomboI unduh <i>file</i> RDF tidak berfungsi	
	Format sintaks RDF	
Modul Program Saat Ini	Manajemen hasil ekstraksi data	2
	Manajemen akun	
Modul Terganti	Manajemen hasil ekstraksi data	1
Modul Terhapus		0
Bahasa Pemrograman	JavaScript	1

### 3.2. Perancangan Sistem

Sistem yang dibuat adalah berbasis web yang menggunakan bahasa pemrograman JavaScript dengan *Framework* Express versi 4.15.2 dan *library* Puppeteer. Dengan menggunakan Puppeteer proses ekstraksi data akan mudah dilakukan. Tema yang digunakan untuk *interface* sistem adalah Now Ui Dashboard versi pro dengan sedikit modifikasi. Diperlukan aplikasi *web browser* seperti Google Chrome, Mozilla Firefox, Safari, dsb. untuk dapat menjalankan sistem. Dikarenakan sistem akan dijalankan untuk mengakses halaman *website*, maka dibutuhkan jaringan internet untuk membangun sistem.

Penelitian ini dibagi menjadi dua tahapan inti yaitu pengambilan data dan pembuatan dokumen *Knowledge Base*. Adapun diagram perancangan sistem pada penelitian ini digambarkan seperti pada Gambar 3.3.

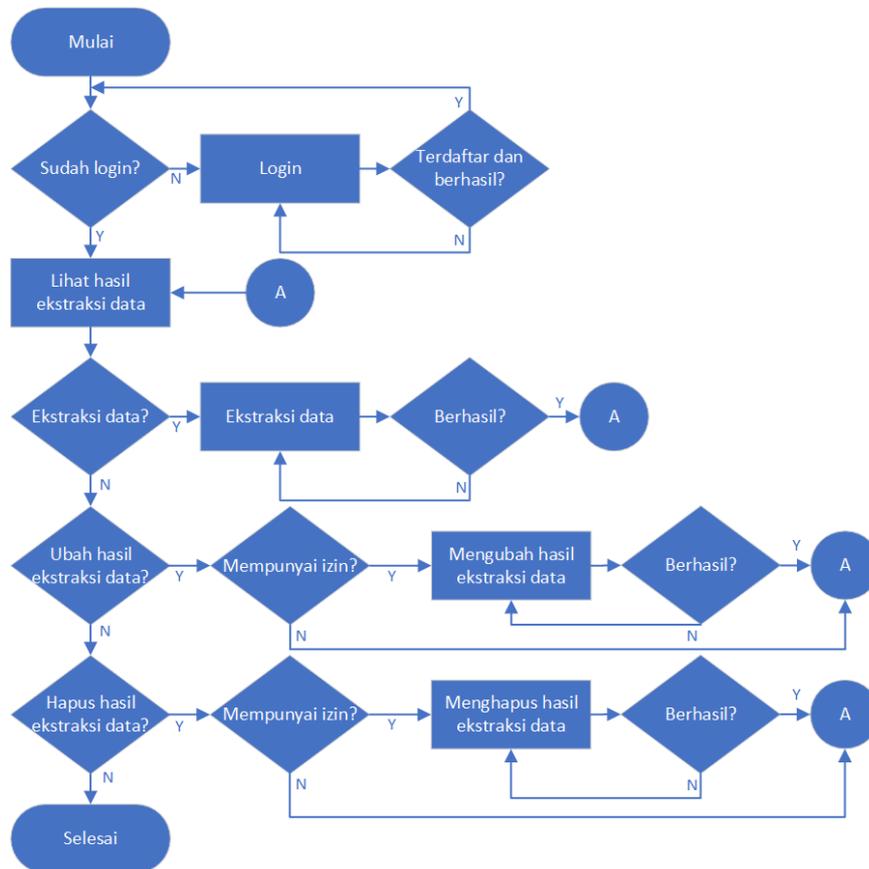


**Gambar 3.3** Blok diagram perancangan sistem

Berdasarkan gambar di atas, atribut RDF didapatkan dari data-data atribut yang berhasil diekstrak dari konten halaman *website* kemudian dijadikan *file* RDF, yang nantinya *file* tersebut akan dijalankan dalam *Knowledge Base*. Sedangkan perancangan *flowchart* sistem adalah sebagai berikut.

- Manajemen Hasil Ekstraksi Data

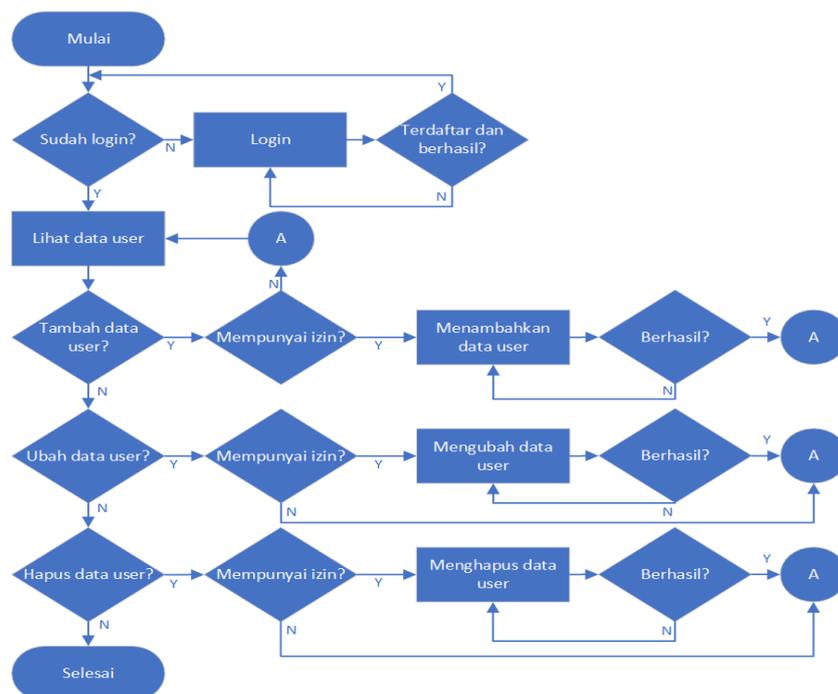
*Flowchart* manajemen hasil ekstraksi data dapat dilihat pada Gambar 3.4.



**Gambar 3.4** Flowchart manajemen hasil ekstraksi data

- Manajemen User

Flowchart manajemen user dapat dilihat pada Gambar 3.5.



**Gambar 3.5** Flowchart manajemen user

### 3.3. Penelusuran Halaman *Website*

Halaman *website* yang akan digunakan adalah halaman *website* portal berita yang valid dan dapat dipercaya informasi maupun datanya, sehingga dapat digunakan dan diolah untuk kebutuhan penelitian. Halaman *website* tersebut adalah halaman *website* portal berita Detik.com dan Kompas.com. Pada kedua *website* tersebut diambil beberapa atribut dari konten halaman artikel berita yaitu judul, *author* dan detail artikel berita. Penelusuran halaman dalam penelitian ini dibutuhkan dalam menentukan halaman *website* yang memiliki konten yang diinginkan. Setelah alamat URL diketahui, yaitu pada halaman indeks *website* berita. Kemudian dilanjutkan dengan penentuan tanggal kapan artikel berita dipublikasikan, sehingga sistem yang dibuat akan langsung dapat mencari halaman yang ditentukan.

Pada sistem yang dibuat, penentuan halaman awal adalah pada halaman “<https://news.detik.com/indeks>”, dengan alur penentuan halaman sesuai dengan seperti pada Gambar 3.6.

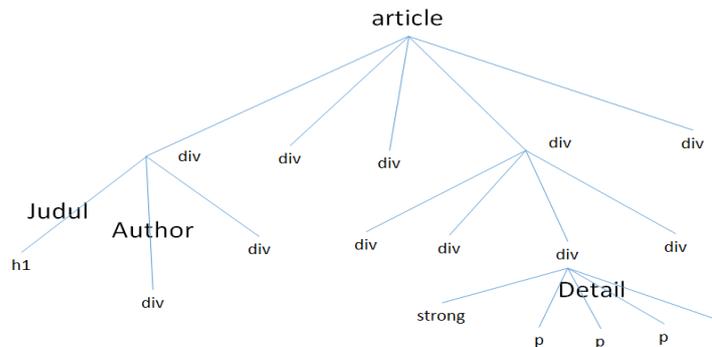


**Gambar 3.6** Flowchart penelusuran halaman

### 3.4. *Crawling*

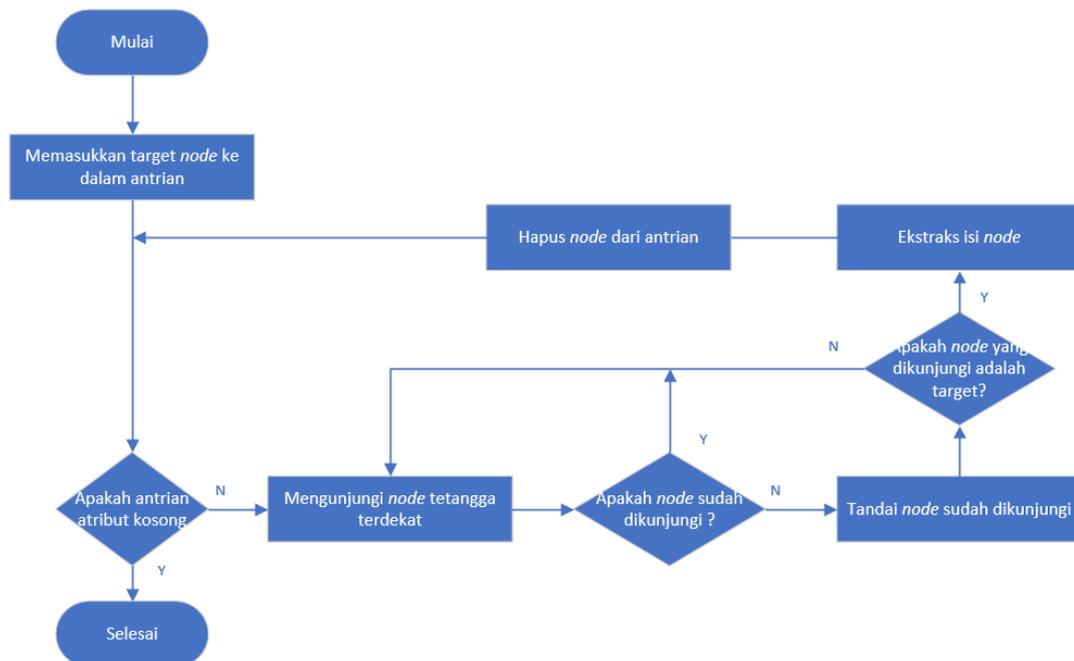
*Crawling* dilakukan untuk menemukan target tag HTML yang diinginkan dengan cara menelusuri tag-tag HTML pada halaman *website* dengan menggunakan algoritma atau metode tertentu. Proses ini dilakukan sebelum mengekstraksi informasi atau pun konten dari halaman *website*. Dalam proses ini, *crawling* dimulai dari menerima *input* berupa URL halaman *website* dan atribut-atribut yang

akan diekstrak dari pengguna, kemudian mengunjungi halaman *website* dari URL tersebut. Setelah proses mengunjungi halaman *website* berhasil, maka akan didapatkan *Document Object Model* (DOM) dari halaman *website* yang direpresentasikan oleh sistem sebagai *graph* seperti ditunjukkan pada Gambar 3.7 berikut.



**Gambar 3.7** Representasi *Graph nodes* tag HTML

Kemudian sistem akan mengunjungi setiap *node* dengan menggunakan metode *Breadth First Search*. Proses *crawling* dengan metode *Breadth First Search*, dilakukan dengan mengunjungi semua *node* tag HTML pada tingkatan yang sama atau *node* tetangga (*neighbor node*) terlebih dahulu, baru setelah itu menelusuri *node* anak (*child node*) untuk menemukan target tag yang diinginkan. Adapun *flowchart crawling* dengan menggunakan metode *Breadth First Search* dapat dilihat pada Gambar 3.8.



**Gambar 3.8** *Flowchart* penelusuran atribut

### **3.5. Scraping**

Proses *scraping* atau ekstraksi data dilakukan setelah mendapatkan target tag HTML di mana teks dari tag HTML diambil dan dibersihkan dari sisa teks yang tidak diinginkan seperti sintaks *script JavaScript* hingga tag HTML itu sendiri. Kemudian, dilakukan pembuatan data berseri yang mengandung meta ataupun detail dari konfigurasi ekstraksi data yang digunakan. Meta tersebut meliputi atribut, *selector*, dan jenis traversal *crawling*. Proses ini melakukan konversi data ke dalam format standar JSON yang dapat disimpan di *database*. Format JSON dipilih karena serialisasi JSON dapat menggunakan *array* dan tidak perlu menggunakan *parser* khusus untuk diubah menjadi sebuah objek. Akan tetapi, hanya perlu menggunakan fungsi *default* dari JavaScript. Dengan demikian, proses ekstraksi teks dari tag HTML untuk pembuatan *Knowledge Base* akan dapat dengan mudah dilakukan.

### **3.6. Generate File RDF**

Pada pembuatan *Knowledge Base*, terdapat beberapa hal yang harus diperhatikan. Salah satunya yaitu *framework* yang dibutuhkan untuk membangun *Knowledge Base*. Dalam penelitian ini, digunakan *framework* RDF. Selain itu, hal yang perlu diperhatikan adalah menentukan ruang lingkungannya *Knowledge Base*. Penentuan ruang lingkup bertujuan agar membantu membatasi ruang lingkup model dan juga mempermudah dalam merancang struktur dari model ontologi.

Data hasil ekstraksi disimpan ke dalam *database* untuk dapat digunakan dalam pembuatan *file* RDF dengan format owl. Pembuatan *file* RDF dilakukan setelah data hasil ekstraksi didapatkan dan disimpan ke dalam *database*, sehingga pembuatan *file* RDF dapat dilakukan tanpa harus melalui proses *crawling* dan *scraping* lagi karena data atribut yang telah tersimpan dalam *database* dan dapat digunakan lagi untuk *generate file* RDF. Kemudian, proses selanjutnya adalah menuliskan dokumen dengan format yang sesuai dan berisi atribut-atribut ke dalam *file* RDF.

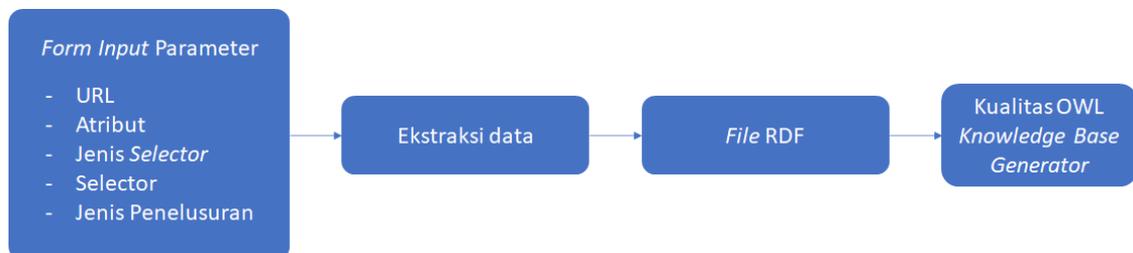
### **3.7. Kerangka Kerja Konseptual**

Penelitian ini mengembangkan teknik *Web Scraping* untuk pengambilan data dari konten halaman *website*. Proses pengambilan data tersebut dapat

dilakukan dengan metode penelusuran. Metode penelusuran yang digunakan dalam penelitian adalah metode *Breadth First Search*. Proses pengambilan data membutuhkan beberapa parameter pendukung. Parameter pendukung tersebut dibutuhkan sebagai *input* dalam pengoperasian sistem. Sistem yang dibangun memiliki tujuan akhir untuk membuat dokumen *Knowledge Base* secara dinamis. Adapun dokumen yang akan dibuat adalah hasil pengolahan dari ekstraksi data yang menghasilkan *output* berupa *file* RDF.

Pengukuran penelitian ini dilakukan berdasarkan atribut kualitas perangkat lunak ISO 9126 sebagai variabel bebas yang secara teknis diklasifikasikan menjadi beberapa atribut yang dapat dilihat pada Tabel 1. Sedangkan variabel terikat adalah variabel yang dipengaruhi atau disebabkan oleh variabel bebas. Dalam penelitian ini, variabel terikatnya adalah kualitas perangkat lunak dari sistem yang telah dibuat. Dengan demikian, peneliti dapat mengetahui bahwa penelitian telah selesai atau sebaliknya. Maka dengan adanya komponen ISO 9126 pengukuran kualitas perangkat yang telah dibuat dapat diukur, sehingga dapat diambil kesimpulan.

Berdasarkan penjelasan tersebut maka kerangka kerja konseptual pada penelitian ini digambarkan seperti pada Gambar 3.9.



**Gambar 3.9** Kerangka kerja konseptual

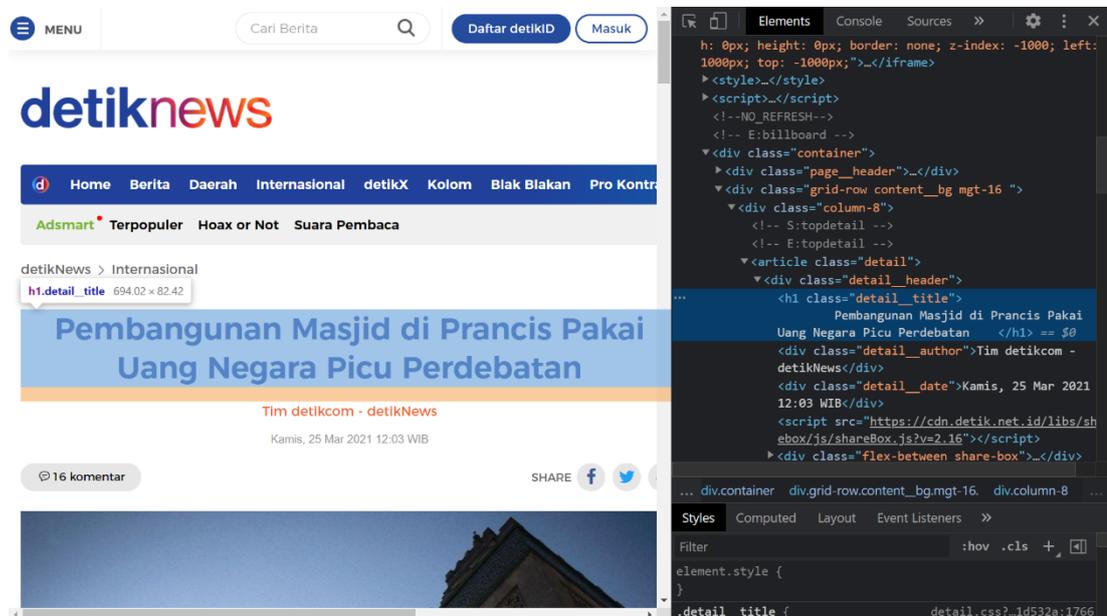
## BAB IV UJI COBA DAN PEMBAHASAN

### 4.1. Langkah-langkah Uji Coba

Implementasi sistem adalah pelaksanaan dari sebuah sistem yang sudah dirancang dan terencana agar sistem berjalan dengan kebutuhan pengguna.

#### 4.1.1. *Crawling* dengan Metode *Breadth First Search*

Penerapan ekstraksi diawali dengan penentuan halaman *website* yang akan diekstrak datanya. Peneliti kemudian menentukan data atau atribut apa saja yang akan diekstrak. Pada kasus ini halaman yang akan diekstrak adalah halaman pada portal berita Detik.com.



Gambar 4.1 Halaman artikel berita Detik.com

Pada Gambar 4.1 menunjukkan beberapa atribut dapat diekstrak pada halaman tersebut. Atribut yang akan diekstrak diidentifikasi berdasarkan tag maupun kelas yang ada pada tag HTML. Salah satunya yaitu judul dari artikel berita yang berada pada tag HTML dengan kelas *detail\_\_title*. Selain judul, atribut yang akan peneliti ekstrak adalah penulis dan detail artikel berita itu sendiri. Penerapan *crawling* dengan metode *Breadth First Search* dilakukan dengan pengkodean pemrograman JavaScript seperti ditunjukkan pada *source code* berikut.

```

let BFS = function(nodes, selector, selector_type,
selector_traversal) {
  let queue = [];
  let elements = [];
  let nodesVisited = [];
  if (nodes != null) {
    queue.unshift(nodes);
    while (queue.length != 0) {
      let element = queue.shift();
      if(selector_type == 'tag')
nodesVisited.push(element.tagName.toLowerCase());
      else if(selector_type == 'class')
nodesVisited.push(`${element.tagName.toLowerCase()}.${element.c
lassName}`);

      if(selector_type == 'tag') {
        if(element.tagName.toLowerCase() == selector &&
selector_traversal == 'first') return {elements: [element],
nodesVisited};
        else if(element.tagName.toLowerCase() == selector
&& selector_traversal == 'all') elements.push(element);
      } else if(selector_type == 'class') {
        if(element.className == selector &&
selector_traversal == 'first') return {elements: [element],
nodesVisited};
        else if(element.className == selector &&
selector_traversal == 'all') elements.push(element);
      }
      let childrens = element.children;
      for (let i = 0; i < childrens.length; i++) {
        queue.push(childrens[i]);
      }
    }
  }
  return {elements, nodesVisited};
}

```

Metode ini mengunjungi setiap *node* tag pada DOM halaman *website*. Kemudian menambahkan ke sebuah *array* atau larik. Setelah itu, *crawler* menelusuri setiap tag sampai setiap tag telah dikunjungi. Pada saat *crawling*, ketika mengunjungi tag pada halaman *website*, setiap tag diperiksa apakah sudah ada dalam larik atau belum, untuk tag yang belum dikunjungi akan ditambahkan ke larik. Adapun hasil *running* program ditunjukkan pada cuplikan *log* sebagai berikut.

```

{
  url:      'https://news.detik.com/internasional/d-
5507119/pembangunan-masjid-di-prancis-pakai-uang-
negara-picu-perdebatan?tag_from=wp_nhl_6',
  objectId: '605c32a0a62a9d8f71d3dc7f',
  attribute: [ 'Judul', 'Author', 'Detail' ],
  selector: [ 'h1', 'detail__author', 'p' ],
  selector_type: [ 'tag', 'class', 'tag' ],
  selector_traversal_type: [ 'first', 'first', 'all' ]
},
{
  attributes: {
    Judul: 'Pembangunan Masjid di Prancis Pakai Uang
Negara Picu Perdebatan',
    Author: 'Tim detikcom - detikNews',
    Detail: 'Perdebatan tengah mencuat di Prancis
terkait rencana membangun sebuah masjid di wilayah
Strasbourg. Otoritas Strasbourg menyetujui penggunaan
dana publik untuk pembangunan masjid itu, tapi
Kementerian Dalam Negeri Prancis menyampaikan
keberatan.\n' +
"Seperti dilansir AFP, Kamis (25/3/2021), Kementerian
Dalam Negeri Prancis menyatakan bahwa otoritas
Strasbourg menggunakan dana negara untuk mendanai
'campur tangan asing' di wilayah Prancis. Diketahui
bahwa rencana pembangunan masjid di Strasbourg itu
didukung oleh kelompok Muslim Turki
terkemuka.
Skip\n" +
'Pada Senin (22/3) waktu setempat, otoritas kota
Strasbourg, yang dipimpin Wali Kota Jeanne Barseghian
dari Partai Hijau mengabdulkan hibah sebesar 2,5 juta
Euro

```

'Namun CMIG merupakan salah satu dari tiga konfederasi Muslim di Prancis yang menolak untuk menandatangani piagam antiekstremisme yang diperjuangkan Presiden Emmanuel Macron. Diketahui bahwa Macron ingin menindak tegas ekstremisme Islam di Prancis, yang diyakini mendalangi serentetan serangan teror di negara itu sejak tahun 2015.\n' +

"Macron ingin kelompok-kelompok tersebut berkomitmen secara tertulis menolak 'Islam politis' dan menghormati aturan hukum di Prancis, saat dia berupaya memerangi kaum radikal yang dipandang sebagai ancaman terhadap sistem sekuler yang berlaku di Prancis sejak lama.\n" +

'Pemerintah Prancis juga telah menyusun rancangan undang-undang yang akan memaksa kelompok-kelompok Muslim untuk mendeklarasikan pendanaan asing yang besar dan akan memberikan wewenang lebih besar kepada negara untuk menghentikan pidato yang dinilai menyebarkan kebencian atau kekerasan.\n' +

"Kami yakin bahwa asosiasi ini tidak lagi bisa menjadi perwakilan Islam di Prancis," tegas Menteri Dalam Negeri Prancis, Gerald Darmanin, merujuk pada CMIG.\n' +

"Kami meyakini bahwa otoritas kota ini tidak seharusnya mendanai campur tangan asing di tanah kita," imbuh Darmanin kepada televisi BFM.\n' +

'\n' +

"Lihat juga video 'Prancis Cabut Travel Ban untuk 7 Negara Non Uni Eropa':"

},

nodesVisited: {

Judul: [

'article', 'div',

```

        'div',      'script',
        'h1'
    ],
    Author: [
        'article.detail',
        'div.detail__header',
        'div.detail__media',
        'style.',
        'div.',
        'div.detail__body itp_bodycontent_wrapper',
        'div.',
        'script.',
        'div.',
        'script.',
        'div.flex-between share-box',
        'script.',
        'h1.detail__title',
        'div.detail__author'
    ],
    Detail: [
        'article',      'div',      'div',      'style',
'div',
        'div',          'div',      'script', 'div',
'script',
        'div',          'script', 'h1',      'div',      'div',
        'script',      'div',      'script', 'figure',
'script',
        'script',      'div',      'div',      'div',      'div',
        'div',          'div',      'div',      'div',      'img',
        'figcaption', 'div',      'strong', 'p',      'p',
        'p',            'div',      'p',      'table', 'p',

```

```

        'p',          'p',          'p',          'p',          'p',
        'p',          'p',          'p',          'p',          'div',
        'div',        'div',        'iframe',     'a',          'a',
        'a',          'a',          'a',          'a',          'a',
        'a',          'a',          'a',          'em',         'a',
        'div',        'a',          'tbody',     'strong',
'strong',
        'a',          'strong',    'a',          'div',        'div',
        'div',        'div',        'i',          'span',       'i',
        'i',          'i',          'span',       'i',          'span',
        'i',          'i',          'i',          'span',       'div',
        'tr',         'br',        'div',        'br',         'br',
        'i',          'div',        'a',          'a',          'a',
        ... 99 more items
    ]
},
    scraped_page_title: 'Pembangunan Masjid di Prancis
Pakai Uang Negara Picu Perdebatan'
}

```

Berdasarkan cuplikan *log* di atas dapat dilihat bahwa sistem melakukan *crawling* pada tag-tag HTML yang dikunjungi. Sistem mengimplementasikan metode *Breadth first Search* untuk mengunjungi semua *node* tag HTML pada tingkatan yang sama atau *node* tetangga (*neighbor node*) terlebih dahulu, lalu menelusuri *node* anak (*child node*) untuk menemukan target tag HTML yang diinginkan. Susunan tag HTML yang ada pada website Detik.com diperlihatkan pada cuplikan code berikut.

```

<article class="detail">
  <div class="detail__header">
    <h1 class="detail__title">Pembangunan Masjid
di Prancis Pakai Uang Negara Picu Perdebatan</h1>
    <div class="detail__author">Tim detikcom -
detikNews</div>
    <div class="detail__date">Kamis, 25 Mar 2021
12:03 WIB</div>
    <script
src="https://cdn.detik.net.id/libs/sharebox/js/shareB
ox.js?v=2.16"></script>
    <div class="flex-between share-box">
      <div>
        <a dtr-evt="komentar top" dtr-sec="" dtr-
act="tombol komentar" onclick="_pt(this)">
          <i class="icon icon-comment icon--
xs"></i>
          <span>16 komentar</span>
        </a>
      </div>
      <div class="detail__share">
        <a dtr-evt="share top" dtr-sec="" dtr-
act="share facebook" onclick="_pt(this)"><i
class="icon icon-facebook"></i></a>
        <a dtr-evt="share top" dtr-sec="" dtr-
act="share twitter" onclick="_pt(this)"><i
class="icon icon-twitter"></i></a>
        <a dtr-evt="share top" dtr-sec="" dtr-
act="copy link" onclick="_pt(this)"><i class="icon
icon-link"></i>
          <span class="notif">URL telah
disalin</span>
        </a>
      </div>
    </div>
  </div>

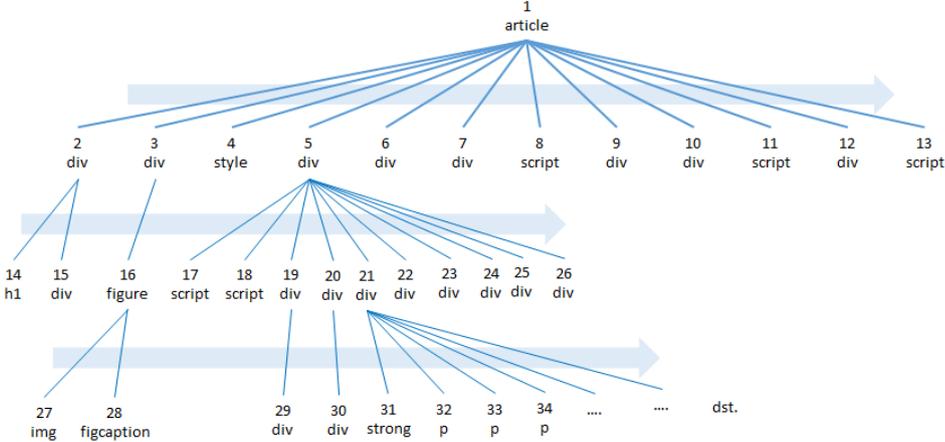
```

```

</div>
<div class="detail__media">
  <figure dtr-evt="cover image" dtr-sec="image"
dtr-act="cover image" onclick="_pt(this)"
class="detail__media-image">
    
    <figcaption class="detail__media-
caption">Ilustrasi masjid di Eropa (dok. AFP
Photo/Joel Saget)</figcaption>

```

Pada proses pencarian tag HTML yang dilakukan di dalam sistem adalah dengan merepresentasikan DOM dari halaman *website* menjadi bentuk *graph*. Kemudian sistem menyaring terlebih dahulu tag-tag dari tag yang tidak diinginkan seperti tag `<style>` dan tag `<script>` serta memulai pencarian dari tag yang diinginkan, yang dalam kasus ini pencarian dimulai dari tag `<article>`. Adapun bentuk representasi *graph nodes* tag HTML pada halaman *website* pada proses *crawling* dengan metode *Breadth First Search* dari sistem ditunjukkan pada Gambar 4.2 berikut.



**Gambar 4.2** *Graph* tag HTML beserta urutan traversal

Pada Gambar 4.2 dapat dilihat urutan traversal *nodes* tag HTML dari proses *crawling* yang dilakukan oleh sistem untuk menemukan target tag HTML yang diinginkan. Dengan demikian, urutan traversal *node* pada proses *crawling* dengan menggunakan metode *Breadth First Search* untuk atribut judul, *author*, dan detail adalah sebagai berikut.

1. Judul: article, div, div, style, div, div, div, script, div, script, div, script, h1.
2. *Author*: article, div, div, style, div, div, div, script, div, script, div, script, h1, div.
3. Detail: article, div, div, style, div, div, div, script, div, script, div, script, h1, div, div, script, div, script, figure, script, script, div, div, div, div, div, div, div, div, img, figcaption, div, strong, p, dst.

#### 4.1.2. Pembuatan *File* RDF

Pembuatan *file* RDF bertujuan untuk digunakan dalam pembuatan *Knowledge Base*, di mana membutuhkan *file* berekstensi owl. Struktur penulisan *file* RDF pada proses konversi disesuaikan dengan struktur data dari web aslinya. Proses konversi data memperhatikan bagian yang telah ditentukan yang terdiri dari subjek, predikat, dan objek dalam struktur RDF. Adapun pengkodean dari proses ini ditunjukkan pada *source code* berikut.

```
// Penulisan isi file RDF
let result = await Scraper.scrape(input);
let strXML = `<?xml version="1.0"?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#"
  xmlns:dc="${baseUrl}/#konten">

  <rdf:Description rdf:about="${input.url}">`;

  for (let i = 0; i < input.attribute.length; i++) {
    strXML += `\\n<dc:${input.attribute[i].replace(/\\s/g, "-
")}>${
      result.attributes[input.attribute[i]]
    }</dc:${input.attribute[i].replace(/\\s/g, "-")}>`;
  }
  strXML += `
  </rdf:Description>
</rdf:RDF>`;

.....
```

Kemudian dari isi *file* RDF ini disimpan ke dalam *database* untuk dapat dilakukan proses *generate file* RDF pada saat *user* ingin mengunduh hasil ekstraksi. Pengkodean untuk *generate file* RDF dapat dilihat pada *source code* berikut.

Hasil dari cuplikan *source code* tersebut adalah *file* RDF berekstensi owl yang

```
// Pembuatan file RDF
let objectId = req.params.objectId;
const scrape = await Scrape.findOne({
  where: {
    code: objectId
  }
});
if (scrape === null) {
  return res.status(404).send("Cannot find data.");
} else {
  let content = JSON.parse(scrape.content);
  const fileData = content.xml;
  const fileName = `${scrape.page_title}.owl`;
  const fileType = 'text/plain';

  res.writeHead(200, {
    'Content-Disposition': 'attachment;
filename="${fileName}"',
    'Content-Type': fileType,
  })

  const download = Buffer.from(fileData);
  return res.end(download);

})

const download = Buffer.from(fileData);
return res.end(download);
```

dapat diunduh oleh *user*. Adapun hasil dari proses *generate file* RDF ditunjukkan pada Gambar 4.3.

```

<?xml version="1.0"?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="https://news.detik.com/konten#">
    <rdf:Description rdf:about="https://news.detik.com/internasional/d-5507119/
      pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6">
    <dc:Judul>Pembangunan Masjid di Prancis Pakai Uang Negara Picu Perdebatan</dc:Judul>
    <dc:Author>Tim detikcom - detikNews</dc:Author>
    <dc:Detail>Perdebatan tengah mencuat di Prancis terkait rencana membangun sebuah masjid di
      wilayah Strasbourg. Otoritas Strasbourg menyetujui penggunaan dana publik untuk pembangunan
      masjid itu, tapi Kementerian Dalam Negeri Prancis menyampaikan keberatan.
      Seperti dilansir AFP, Kamis (25/3/2021), Kementerian Dalam Negeri Prancis menyatakan bahwa
      otoritas Strasbourg menggunakan dana negara untuk mendanai 'campur tangan asing' di wilayah
      Prancis. Diketahui bahwa rencana pembangunan masjid di Strasbourg itu didukung oleh
      kelompok Muslim Turki terkemuka.
      Pada Senin (22/3) waktu setempat, otoritas kota Strasbourg, yang dipimpin Wali Kota Jeanne
      Barseghian dari Partai Hijau mengabdikan hibah sebesar 2,5 juta Euro untuk Milli Gorus
      Islamic Confederation (CMIG), sebuah gerakan pan-Eropa untuk diaspora Turki.
      Namun CMIG merupakan salah satu dari tiga konfederasi Muslim di Prancis yang menolak untuk
      menandatangani piagam antiekstremisme yang diperjuangkan Presiden Emmanuel Macron.
      Diketahui bahwa Macron ingin menindak tegas ekstremisme Islam di Prancis, yang diyakini
      mendalangi serentetan serangan teror di negara itu sejak tahun 2015.
      Macron ingin kelompok-kelompok tersebut berkomitmen secara tertulis menolak 'Islam politis'
      dan menghormati aturan hukum di Prancis, saat dia berupaya memerangi kaum radikal yang
      dipandang sebagai ancaman terhadap sistem sekuler yang berlaku di Prancis sejak lama.
      Pemerintah Prancis juga telah menyusun rancangan undang-undang yang akan memaksa
      kelompok-kelompok Muslim untuk mendeklarasikan pendanaan asing yang besar dan akan
      memberikan wewenang lebih besar kepada negara untuk menghentikan pidato yang dinilai
      menyebarkan kebencian atau kekerasan.
      "Kami yakin bahwa asosiasi ini tidak lagi bisa menjadi perwakilan Islam di Prancis," tegas
      Menteri Dalam Negeri Prancis, Gerald Darmanin, merujuk pada CMIG.
      "Kami meyakini bahwa otoritas kota ini tidak seharusnya mendanai campur tangan asing di
      tanah kita," imbuh Darmanin kepada televisi BFM.

      Lihat juga video 'Prancis Cabut Travel Ban untuk 7 Negara Non Uni Eropa':</dc:Detail>
    </rdf:Description>
  </rdf:RDF>

```

**Gambar 4.3** Hasil *generate file* RDF dari halaman Detik.com

Sedangkan hasil dari *generate file* RDF dari halaman Kompas.com dapat dilihat pada Gambar 4.4.

```

<?xml version="1.0"?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="https://www.kompas.com/konten#">
    <rdf:Description rdf:about="https://www.kompas.com/tren/read/2021/03/25/131100365/
      kartu-prakerja-gelombang-16-dibuka-perhatikan-hal-berikut-agar-berpeluang">
    <dc:Judul>Kartu Prakerja Gelombang 16 Dibuka, Perhatikan Hal Berikut agar Berpeluang Lolos</
    dc:Judul>
    <dc:Author>Penulis Dandy Bayu Bramasta
    Editor Sari Hardiyanto</dc:Author>
    <dc:Detail>Kartu Prakerja Gelombang 16 Dibuka, Perhatikan Hal Berikut agar Berpeluang Lolos
    KOMPAS.com - Pendaftaran Kartu Prakerja gelombang 16 telah resmi dibuka pada Kamis (25/3/
    2021) siang.
    Hal tersebut telah dikonfirmasi oleh Head of Communications Manajemen Kartu Prakerja Louisa
    Tahutu.
    "Hari ini gelombang 16 dibuka pukul 12.00 WIB," kata Louisa kepada Kompas.com, Kamis (25/3/
    2021).
    Mengenai kuotanya, gelombang 16 yang akan menjadi penutup pada semester I 2021 ini hanya
    tersisa 300.000 dari target 2,7 juta orang.
    Baca juga: Cara Daftar Prakerja Gelombang 16, Gelombang Terakhir di Semester I 2021
    Baca juga: Kartu Prakerja Gelombang 16 Dibuka Pukul 12.00 WIB, Kuota 300.000 Orang
    Berikut hal yang perlu diperhatikan agar memiliki peluang lolos pada gelombang 16 ini.
    Hanya ada satu link resmi pendaftaran Kartu Prakerja, yaitu di laman www.prakerja.go.id.
    Selain link itu, dapat dipastikan palsu atau abal-abal, meski mengatasnamakan Kartu
    Prakerja.
    Berdasarkan pantauan Kompas.com, berikut 5 link palsu yang banyak beredar:
    Baca juga: Ramai soal Gambar Lonceng di Halaman Dashboard Kartu Prakerja, Apa Itu?

    Tulis komentar dengan menyertakan tagar #JernihBerkomentar dan #MelihatHarapan di kolom
    komentar artikel Kompas.com. Menangkan E-Voucher senilai Jutaan Rupiah dan 1 unit
    Smartphone.
    A post shared by Kartu Prakerja (@prakerja.go.id)
    Kartu Prakerja Gelombang 15 Resmi Dibuka Hari Ini
    Kartu Prakerja Gelombang 15 Resmi Dibuka Hari Ini
    Pendaftaran Kartu Prakerja Gelombang 13 Dibuka Hari Ini, Simak Cara Daftarnya
    Ingat! Ini Sejumlah Alasan Tak Lolos Program Kartu Prakerja
    Buntut Deportasi Gubernur Papua, Imigrasi Akan Panggil Lukas Enembe untuk Dimintai
    Klarifikasi
    Mobil Tabrak Rumah Warga Hingga Rusak Parah, Diduga Akibat Sopir Mengantuk
    Sempat Hilang Selama Satu Hari, Seorang Pria Ditemukan Tewas Diduga Tersengat Listrik
    Keterbatasan Fisik Bukanlah Penghalang
    Tawuran Warga di Johar Baru Berhasil Dibubarkan Petugas dengan Gas Air Mata</dc:Detail>
  </rdf:Description>
</rdf:RDF>

```

**Gambar 4.4** Hasil *generate file* RDF dari halaman Kompas.com

### 4.1.3. Validasi Sintaks RDF

Sistem menerapkan *Web Scraping* pada halaman *website* berbasis ontologi menggunakan algoritma *Breadth First Search* dengan bahasa pemrograman JavaScript. Penerapan algoritma tersebut digunakan untuk proses *crawling*, sehingga target tag HTML dapat diekstrak datanya. Hasil dari ekstraksi data digunakan untuk membuat *file* RDF. *File* RDF yang berhasil dibuat oleh sistem, kemudian diuji dengan dua pengujian. Pengujian pertama dilakukan pada *website* W3C *RDF Validation Service* untuk mengetahui apakah sintaks RDF yang telah dituliskan pada kode program adalah sesuai dengan format RDF atau tidak.

## Validation Results

Your RDF document validated successfully.

### Triples of the Data Model

Number	Subject	Predicate	Object
1	<a href="https://news.detik.com/internasional/d-5507119/pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6">https://news.detik.com/internasional/d-5507119/pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6</a>	<a href="https://news.detik.com/konten#Judul">https://news.detik.com/konten#Judul</a>	"Pembangunan Masjid di Prancis Pakai Uang Negara Picu Perdebatan"
2	<a href="https://news.detik.com/internasional/d-5507119/pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6">https://news.detik.com/internasional/d-5507119/pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6</a>	<a href="https://news.detik.com/konten#Author">https://news.detik.com/konten#Author</a>	"Tim detikcom - detikNews"
3	<a href="https://news.detik.com/internasional/d-5507119/pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6">https://news.detik.com/internasional/d-5507119/pembangunan-masjid-di-prancis-pakai-uang-negara-picu-perdebatan?tag_from=wp_nhl_6</a>	<a href="https://news.detik.com/konten#Detail">https://news.detik.com/konten#Detail</a>	"Perdebatan tengah mencuat di Prancis terkait rencana membangun sebuah masjid di wilayah Strasbourg. Otoritas Strasbourg menyetujui penggunaan dana publik untuk pembangunan masjid itu, tapi Kementerian Dalam Negeri Prancis menyampaikan keberatan. Seperti dilansir AFP, Ramis (25/3/2021), Kementerian Dalam Negeri Prancis menyatakan bahwa otoritas Strasbourg menggunakan dana negara untuk mendanai "campur tangan asing" di wilayah Prancis. Diketahui bahwa rencana pembangunan masjid di Strasbourg itu didukung oleh kelompok Muslim Turki terkemuka. Pada Senin (22/3) waktu setempat, otoritas kota Strasbourg, yang dipimpin Wali Kota Jeanne Barseghian dari Partai Hijau mengabdikan hibah sebesar 2,5 juta Euro untuk Milli Gorus Islamic Confederation (CMIG), sebuah gerakan pan-Eropa untuk diaspora Turki. Namun CMIG merupakan salah satu dari tiga konfederasi Muslim di Prancis yang menolak untuk menandatangani piagam antiekstremisme yang diperjuangkan Presiden Emmanuel Macron. Diketahui bahwa Macron ingin menindaki tegas ekstremisme Islam di Prancis, yang diyakini mendalangi serentetan serangan teror di negara itu sejak tahun 2015. Macron ingin kelompok-kelompok tersebut berkomitmen secara tertulis menolak "Islam politis" dan menghormati aturan hukum di Prancis, saat dia berupaya memerangi kaum radikal yang dipandangnya sebagai ancaman terhadap sistem sekuler yang berlaku di Prancis sejak lama. Pemerintah Prancis juga telah menyusun rancangan undang-undang yang akan memaksa kelompok-kelompok Muslim untuk mendeklarasikan pendanaan asing yang besar dan akan memberikan wewenang lebih besar kepada negara untuk menghentikan pidato yang dinilai menyebarkan kebencian atau kekerasan. "Kami yakin bahwa asosiasi ini tidak lagi bisa menjadi perwakilan Islam di Prancis," tegas Menteri Dalam Negeri Prancis, Gerald Darmanin, merujuk pada CMIG. "Kami meyakini bahwa otoritas kota ini tidak seharusnya mendanai campur tangan asing di tanah kita," imbuh Darmanin kepada televisi BFM. Lihat juga video 'Prancis Cabut Travel Ban untuk 7 Negara Non Uni Eropa':"

Gambar 4.5 Tabel triples website W3C RDF Validation Service

Pada Gambar 4.5 menunjukkan hasil validasi sintaks pada isi *file* RDF yang dibuat sistem adalah benar dan tidak ada *error*. Dibuktikan dengan pesan validasi yang ditampilkan yaitu "Your RDF document validated successfully." dan tabel *triples* yang tampil sesuai dengan isi *file* RDF yang telah dibuat oleh sistem.

### 4.1.4. Linked Data RDF

Pengujian kedua dilakukan menggunakan kueri SPARQL guna menentukan apakah *Knowledge Base* telah berjalan dengan benar atau tidak dengan *tool* ontologi OpenLink Virtuoso, sehingga menghasilkan sebuah pernyataan RDF (subjek, predikat, objek) yang saling berhubungan satu dengan lainnya. Pengujian kedua dilakukan setelah *file* RDF telah diunggah ke dalam Virtuoso untuk dilakukan pengecekan isi dokumen. Pada Gambar 4.6 menunjukkan *query* SPARQL yang akan dieksekusi.

```

Default Graph IRI
Query
PREFIX dc: <https://news.detik.com/konten#>

SELECT ?Judul ?Author ?Detail
WHERE {
  ?x dc:Judul ?Judul.
  ?x dc:Author ?Author.
  ?x dc:Detail ?Detail.
}

```

Execute Save Load Clear

**Gambar 4.6** Eksekusi *query* SPARQL untuk menampilkan data

Sedangkan hasil *query* SPARQL yang telah dieksekusi ditunjukkan pada Gambar 4.7.

Judul	Author	Detail
"Pembangunan Masjid di Prancis Pakai Uang Negara Picu Perdebatan"	"Tim detikcom - detikNews"	"Perdebatan tengah mencuat di Prancis t Seperti dilansir AFP, Kamis (25/3/2021) Pada Senin (22/3) waktu setempat, otori Namun CMIG merupakan salah satu dari ti Macron ingin kelompok-kelompok tersebut Pemerintah Prancis juga telah menyusun "Kami yakin bahwa asosiasi ini tidak le "Kami meyakini bahwa otoritas kota ini  Lihat juga video 'Prancis Cabut Travel

**Gambar 4.7** Tabel *triples* hasil *query* SPARQL

#### 4.1.5. Evaluasi Sistem

Pada tahap ini akan ditampilkan evaluasi sistem dengan mencari nilai *functionality*, *reliability*, *usability*, *efficiency*, *maintainability*, dan *portability* dengan menghitung *Function Point*. *Function Point* dihitung dengan mencari jumlah dari komponen fungsional sistem dan dilanjutkan dengan mengevaluasi kuantitasi bobot kompleksitas dari tiap komponen tersebut. Kemudian menghitung TCF (*Technical Complexity Factor*). Cara menghitung TCF adalah dengan menggunakan 14 *General System Characteristic* (GSC), dimana masing-masing GSC berskala nol hingga lima. Skala nol menunjukkan tidak adanya pengaruh dan skala lima menunjukkan adanya pengaruh yang banyak terhadap keseluruhan sistem. Setelah setiap karakteristik diberi bobot masing-masing dan dijumlahkan, sehingga nilai *Function Point* didapatkan.

a. *Functionality*

Indikator *functionality* dapat dihitung dari *Function Point* yang telah didapatkan. Perhitungan *Function Point* membutuhkan parameter berupa *inputs*, *outputs*, *inquires*, *files*, dan *interfaces* yang didapatkan dari Tabel 3.1. Setiap parameter tersebut harus dinilai kompleksitasnya secara umum yaitu bernilai sederhana, sedang atau kompleks yang ditunjukkan pada Tabel 4.1 dan dihitung dengan faktor peubah kompleksitas yang didapatkan dari Tabel 4.2.

**Tabel 4.1** Tabel perhitungan *Function Point*

Paramater	Jumlah	Faktor Pembobotan			
		Sederhana	Sedang	Kompleks	Total
<i>Inputs</i>	5 x	3			15
<i>Outputs</i>	2 x			7	14
<i>Inquires</i>	13 x		4		52
<i>Files</i>	2 x	7			14
<i>Interfaces</i>	1 x		7		7
<b>Jumlah Total</b>					102

**Tabel 4.2** Tabel *Complexity Factor*

No.	Characteristic	Degrees of Influence (DI)					
		0	1	2	3	4	5
1	<i>Data communications</i>		✓				
2	<i>Heavely use configuration</i>			✓			
3	<i>Transaction rate</i>			✓			
4	<i>End-user efficiency</i>				✓		
5	<i>Complex processing</i>				✓		
6	<i>Installation ease</i>		✓				
7	<i>Multiple sites</i>				✓		
8	<i>Performance</i>			✓			
9	<i>Distributed functions</i>		✓				
10	<i>On-line data entry</i>			✓			

11	<i>On-line update</i>			✓			
12	<i>Reusability</i>				✓		
13	<i>Operational ease</i>			✓			
14	<i>Extensibility</i>				✓		
<b>Totai Skala</b>		30					

*Technical Complexity Factor* (TCF) ditentukan dengan memperkirakan tingkat pengaruh dari 14 komponen karakteristik yang ditunjukkan pada Tabel 4.2. Tingkat skala pengaruh berkisar dari nol (tidak ada, atau tidak ada pengaruh) hingga lima (pengaruh kuat secara keseluruhan). Jumlah skor dari 14 karakteristik, yaitu total *Degrees of Influence* (DI) diubah menjadi (TCF), sehingga perhitungan *Function Point* adalah sebagai berikut.

$$FP = \text{Jumlah total} * (0.65 + (0.01 * TCF))$$

$$FP_{\max} = 1.35 * \text{Jumlah total}$$

Di mana:

FP = *Function Point*

FP<sub>max</sub> = *Function Point* maksimum

Jumlah total = Jumlah total domain informasi

TCF = Jumlah total faktor kompleksitas

Berdasarkan *Function Point* dan *Function Point* maksimum yang sudah dihitung, maka *functionality* perangkat lunak adalah sebagai berikut.

$$Functionality = \frac{FP}{FP_{\max}}$$

#### b. *Reliability*

Indikator *reliability* (keandalan) perangkat lunak didapatkan dari metrik *Rate of Failure Occurrence* (ROCOF). Untuk menghitung ROCOF dibutuhkan variabel *Function Point* dan jumlah kegagalan (*Failure*).

$$ROCOF = \frac{Failures}{Function\ Point}$$

Dengan demikian, *reliability* OWL *Knowledge Base Generator* adalah sebagai berikut.

$$Reliability = 1 - ROCOF$$

*Reliability* OWL Knowledge Base Generator adalah sebesar 0.9691, di mana hasil tersebut mengindikasikan bahwa dalam 1000 kali pengoperasian, diperkirakan program mampu bekerja dengan benar sebanyak 969 kali.

c. *Usability*

Indikator *usability* perangkat lunak didapatkan dari metrik *Speed of Operation*. Jika nilai *Speed of Operation* mendekati nilai 0 menandakan *usability* semakin bagus. Sebaliknya, jika nilai *Speed of Operation* semakin mendekati nilai 1, maka *usability* semakin kurang bagus.

$$\text{Speed of Operation} = \frac{\text{Inputs} + \text{Inquires}}{\text{Function Point}}$$

Dengan demikian, *usability* OWL Knowledge Base Generator adalah sebagai berikut.

$$\text{Usability} = 1 - \text{Speed of Operation}$$

d. *Efficiency*

Indikator *efficiency* berkaitan dengan performa, di mana sumber daya yang dikeluarkan diukur sehubungan dengan keakuratan untuk mencapai tujuan. Selain itu, indikator *efficiency* dapat mengukur kemampuan sistem untuk mendukung pengguna dalam melakukan tugasnya dengan menggunakan seberapa banyak waktu yang digunakan untuk melakukan sebuah tugas dibanding dengan alokasi waktu yang ditentukan. Batas waktu untuk memuat halaman *website* menurut Nielsen (2010), adalah 10 detik. Sedangkan menurut Meier et. al (2010), waktu terbaik untuk memuat halaman *website* adalah kurang dari 3 detik dan untuk waktu tunggu yang dapat diterima adalah kurang dari sama dengan 10 detik. Akan tetapi, banyak faktor yang mempengaruhi waktu *load* halaman. Misalnya, proses yang sedang dikerjakan, media halaman, kecepatan jaringan internet, *browser*, hingga spesifikasi komputer. Pengujian *time behaviour* adalah satu-satunya pengujian untuk indikator *efficiency* dalam OWL Knowledge Base Generator. Hal ini merujuk pada seberapa baik perangkat lunak memenuhi tujuannya. Hasil pengukuran *time behaviour* ditunjukkan pada Tabel 4.3.

**Tabel 4.3** Hasil *time behaviour*

<i>Website</i>	<b>Halaman</b>	<b>Atribut</b>	<i>Errors</i>	<i>Avg. Task Duration</i>	<i>Avg. Query</i>	<b>Total</b>

				(per halaman)	<i>Elapsed Time</i>	
Detik.com	3112	778	7	9.890 s	30 ms	9.920 s
Kompas.com	3068	767	7	14.535 s	32 ms	14.567 s

Berdasarkan Tabel 4.3, adanya perbedaan hasil pengujian *time behaviour* disebabkan oleh beberapa faktor seperti *redirecting* halaman, media halaman yang dimuat, ukuran teks yang diekstrak, hingga kesalahan yang terjadi pada sistem (*error*).

$$Efficiency = \frac{\text{Toleran waktu eksekusi}}{\text{Rata - rata waktu eksekusi}}$$

e. *Maintainability*

Semakin banyak perubahan pada sebuah perangkat lunak, maka menandakan bahwa pemeliharaan atau *maintainability* akan semakin sulit untuk dilakukan. Pengukuran nilai *maintainability*, menggunakan metrik *Software Maturity Index* (SMI). Semakin nilai SMI mendekati nilai 1, maka perangkat lunak akan semakin stabil. Sedangkan semakin nilai SMI menjauhi nilai 1 maka, semakin tidak stabil. Variabel-variabel yang dibutuhkan untuk mencari nilai SMI adalah jumlah modul saat ini (MT), jumlah penambahan modul (FA), jumlah modul yang mengalami perubahan (FC), dan jumlah modul yang sudah dihapus sejak perancangan awal (FD).

$$Maintainability = \frac{MT - (FA + FC + FD)}{MT}$$

f. *Portability*

Indikator *portability* perangkat lunak dibagi menjadi tiga yaitu *source code*, *intermediate code*, dan *runnable code*. Setiap bahasa pemrograman memiliki salah satu dari sifat *portability* tersebut. Bahasa pemrograman yang digunakan untuk membangun OWL *Knowledge Base Generator* adalah JavaScript. Perangkat lunak yang dibangun dengan bahasa pemrograman JavaScript merupakan perangkat lunak yang *portable* secara *source code*. *Source code* JavaScript akan diterjemahkan oleh *compiler* Node JS, sehingga dapat digunakan sebagai *web server* dan dapat diakses *client browser*. Pengujian dilakukan pada beberapa *browser* yang berbeda baik di

*desktop* maupun *mobile*. Hasil pengujian dari *portability* dapat dilihat pada Tabel 4.4.

**Tabel 4.4** Hasil pengujian *Portability*

<b>No.</b>	<b>Platform</b>	<b>Browser</b>	<b>Hasil</b>
1.	Google Chrome	<i>Desktop</i>	Sukses
2.	MoziIIa Firefox	<i>Desktop</i>	Sukses
3.	Safari Browser	<i>Desktop</i>	Sukses
4.	Opera Browser	<i>Desktop</i>	Sukses
5.	Firefox Mobile	<i>Mobile</i>	Sukses
6.	Opera Mini	<i>Mobile</i>	Sukses
7.	UC Browser	<i>Mobile</i>	Sukses
8.	DhoIpin Browser	<i>Mobile</i>	Sukses
9.	Chrome Mobile	<i>Mobile</i>	Sukses

#### **4.2. Hasil Uji Coba**

Pada subbab ini di-tampilkannya hasil dari uji coba dengan dilakukan perhitungan dalam mencari nilai *functionality*, *reliability*, *usability*, *efficiency*, *maintainability*, dan *portability* dari sistem OWL *Knowledge Base Generator*, yaitu sebagai berikut.

##### *a. Functionality*

$$FP = 102 * (0.65 + (0.01 * 30)) = 96.9$$

$$FP_{max} = 1.35 * 102 = 137.7$$

$$Functionality = \frac{96.9}{137.7} = 0.7037$$

##### *b. Reliability*

$$ROCOF = \frac{3}{96.9} = 0.0309$$

$$Reliability = 1 - 0.0309 = 0.9691$$

##### *c. Usability*

$$Speed\ of\ Operation = \frac{5 + 13}{96.9} = 0.1863$$

$$Usability = 1 - 0.1863 = 0.8137$$

d. *Efficiency*

$$\text{Efficiency} = \frac{10}{12.243} = 0.8167$$

e. *Maintainability*

$$\text{Maintainability} = \frac{2 - (0 + 1 + 0)}{2} = 0.50$$

f. *Portability*

$$\text{Portability} = \frac{9}{9} = 1.00$$

Berdasarkan hasil uji coba keenam indikator kualitas ISO 9126 yaitu *functionality*, *reliability*, *usability*, *efficiency*, *maintainability*, dan *portability* menghasilkan presentase *Functionality* 70%, *Reliability* 96%, *Usability* 81%, *Efficiency* 81%, *Maintainability* 50%, dan *Portability* 100%, sehingga generalisasi penilaian secara keseluruhan dari pencapaian kualitas perangkat lunak OWL *Knowledge Base Generator* yang dapat dilihat pada Tabel 4.5.

**Tabel 4.5** Pencapaian kualitas OWL *Knowledge Base Generator*

No.	Karakteristik	Nilai
1.	<i>Functionality</i>	0.70
2.	<i>Reliability</i>	0.96
3.	<i>Usability</i>	0.81
4.	<i>Efficiency</i>	0.81
5.	<i>Maintainability</i>	0.50
6.	<i>Portability</i>	1
<b>Total</b>		4.78

Idealnya nilai dari setiap indikator kualitas harus lebih dari sama dengan 0 dan kurang dari sama dengan 1, sehingga nilai maksimal dari keenam kualitas adalah 6. Sedangkan total nilai kualitas OWL *Knowledge Base generator* adalah sebesar 4.78. Dengan demikian presentase pencapaian kualitas adalah sebagai berikut.

$$\text{Presentase Kualitas} = \frac{\text{Pencapaian kualitas}}{\text{Kualitas Maksimum}} * 100\%$$

$$\begin{aligned} &= \frac{4.78}{6} * 100\% \\ &= 79.66\% \end{aligned}$$

### 4.3. Pembahasan

Berdasarkan hasil dari pengujian kualitas perangkat lunak menunjukkan bahwa presentase pencapaian kualitas perangkat lunak adalah sebesar 79.66% di mana angka tersebut menentukan kualitas produk perangkat lunak. Karakteristik kualitas tertinggi dari *OWL Knowledge Base generator* dicapai oleh karakteristik *portability* dan *reliability* dengan nilai 1 dan 0,9691 dan karakteristik terendah dicapai oleh karakteristik *maintainability* dengan nilai 0,50. Berdasarkan pengukuran yang telah dilakukan, maka dapat dilihat bahwa hasil pengukuran dari beberapa metrik berbanding lurus dengan kualitas perangkat lunak dan beberapa metrik lainnya berbanding terbalik. Metrik-metrik yang berbanding lurus dengan kualitas perangkat lunak adalah *Function Point*, *Software Maturity Index (SMI)*, metrik *efficiency*, dan metrik *portability*. Sedangkan metrik-metrik lainnya yang digunakan dalam penelitian ini berbanding terbalik dengan pengukuran kualitas perangkat lunak, sehingga untuk menormalkannya digunakan pengurangan terhadap bilangan 1.

Teknik pengambilan data dari halaman *website* dilakukan pada dokumen halaman *website* berbasis semantik dalam bahasa *markup* seperti HTML atau XHTML. Teknik ini dilakukan dengan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut. Istilah lain dari kegiatan ini adalah pengambilan konten atau sebagian data dari suatu halaman *website*. *Web Scraping* mempunyai banyak manfaat dan berguna dalam membantu penyelesaian masalah terkait pengambilan data. Salah satunya yaitu untuk konten artikel berita di mana berisi informasi yang langsung diambil dari halaman *website*.

Aktivitas seperti memasukan *input*, klik tombol *submit*, menggerakkan *mouse*, hingga *scroll* pada suatu halaman *website* merupakan salah satu usaha manusia dalam mengeksplor informasi yang tersebar di internet. Pembahasan mengenai proses mencari informasi juga erat kaitannya dengan proses pencarian suatu kebenaran. Mencari kebenaran dalam Islam merupakan bagian dari fitrah dan nurani manusia. Hal ini telah dikemukakan dalam Al-Qur'an pada surat Al-Baqarah ayat 185, yang berbunyi:

أَكْثَرَ وَلَكِنَّ الْقِيَمَ الدِّينُ ۚ ذَٰلِكَ اللَّهُ لَخَلْقِ تَنْدِيلٍ لَا عَلَيْهَا النَّاسُ فَطَرَ الَّتِي اللَّهُ فِطَرَتْ حَنِيفًا لِلَّذِينَ وَجْهَكَ فَاقِمِ  
يَعْلَمُونَ لَا النَّاسِ

“Maka hadapkanlah wajahmu dengan lurus kepada agama (Allah), sebagai fitrah Allah yang telah menciptakan manusia menurut fitrah itu. Tidak ada perubahan pada ciptaan Allah. Itulah agama yang lurus; tetapi kebanyakan manusia tidak mengetahui.” (Qs. Al-Rum/30: 30).

Dari ayat di atas dijelaskan bahwa hak dan mencari kebenaran telah diturunkan oleh Allah Swt dan diletakkan pada fitrah seluruh manusia dan hendaknya taat pada ajaran yang terlepas dari segala bentuk penyimpangan dan kesesatan. Yaitu agama Islam. Agama Islam sejatinya telah mengatur aturan dasar bahwa secara *default* semua hal adalah halal, kecuali secara eksplisit diharamkan dalam Al-Qur'an dan Sunnah.

Teknik pengambilan data diterapkan dalam berbagai bidang, salah satunya adalah bagaimana Google merayapi ataupun menjelajahi berbagai halaman *websit*. Jika *website* tersebut bersifat publik, artinya *website* tersebut mengizinkan akses datanya untuk umum atau publik. Dengan demikian, mengekstraksi data yang tersedia untuk umum, tidak melanggar aturan. Selain itu *Web Scraping* tidak menyebabkan data dihapus dari sumbernya. Akan tetapi, teknik ini dapat disalahgunakan untuk keuntungan komersial. Oleh karena itu, telah dibuat peraturan atau perundangan tentang duplikasi dan distribusi ulang konten, yang diatur di undang-undang ITE atau Undang-undang Hak Cipta. Dengan demikian, jika proses mendapatkan informasi ini dilakukan dengan cara meretas dan mendapatkan informasi yang dilindungi yang tidak boleh diungkapkan kepada publik, maka hal ini adalah haram.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1. Kesimpulan**

Setelah dilakukan penelitian terhadap *OWL Knowledge Base Generator* menggunakan metode *Graph Breadth First Search* menghasilkan menghasilkan presentase kualitas sebesar 79.66% dengan presentase *Functionality* 70%, *Reliability* 96%, *Usability* 81%, *Efficiency* 81%, *Maintainability* 50%, dan *Portability* 100%. Berdasarkan implementasi metode BFS pada proses *crawling* pada halaman *website* dapat diketahui bahwa target tag HTML dapat ditemukan dan hasil ekstraksi data dapat digunakan untuk membuat dokumen RDF sebagai *Knowledge Base*. Berdasarkan evaluasi dari pengujian kualitas perangkat lunak *OWL Knowledge Base Generator* dapat disimpulkan bahwa input parameter yang tidak sesuai dengan struktur DOM halaman *website* akan berpengaruh pada hasil ekstraksi data dalam pembuatan *OWL Knowledge Base* dinamis dari konten halaman *website*.

#### **5.2. Saran**

Berdasarkan penelitian yang telah dilakukan, peneliti menyadari bahwa masih terdapat banyak kekurangan yang peneliti ditemui pada sistem *Web Scraping* ini. Dengan demikian, perlu dilakukan pengembangan sistem agar mendapatkan hasil yang lebih baik. Oleh karena itu, peneliti memberikan saran bagi pihak yang berkepentingan dalam pengembangan penelitian di masa yang akan datang antara lain sebagai berikut.

1. Fungsi pengunduhan file RDF bisa dikembangkan dalam membuat *file* RDF lain dan bisa meningkatkan akurasi isi dan struktur RDF sesuai dengan keinginan.
2. Susunan tag HTML pada suatu *website* dapat berubah kapanpun, sehingga proses pengambilan konten pada halaman *website* dengan implementasi metode *Breadth First Search* diperlukan fungsi maupun inputan lain agar hasil ekstraksi dapat maksimal.
3. *File* RDF hasil *output* dari sistem bisa dikembangkan lebih lanjut dengan mengimplementasikan dalam sebuah program atau aplikasi *multi-platform*.

## DAFTAR PUSTAKA

- (APJII), A. P. (2020). *Laporan Survei Internet APJII 2019-2020 (Q2) [Distributor]*. Diakses dari <https://www.apjii.or.id/survei>.
- Behrens, C. A. (1983). Measuring the productivity of computer systems development activities with function points. *IEEE Transactions on Software Engineering*, 6, 648-652.
- Costa, V. d. (2016). Key knowledge management processes for innovation: A systematic literature review. *VINE Journal of Information and Knowledge Management Systems*, 46(3), 386–410.
- Deutsch, A. F. (1999). A query language for XML. *Computer networks*, 31(11-16), 1155-1169.
- Indonesia, W. B. (2021). *Wage Rudolf Soepratman*. [https://id.wikipedia.org/wiki/Wage\\_Rudolf\\_Soepratman](https://id.wikipedia.org/wiki/Wage_Rudolf_Soepratman) (Diakses 02 November 2021).
- Julian, L. R. (2015). The use of web scraping in computer parts and assembly price comparison. *In 2015 3rd International Conference on New Media (CONMEDIA)*, 1-6.
- Kumawat, P. &. (2018). Design and Development of Cost Measurement Mechanism for Re-Engineering Project Using Function Point Analysis. *In International Conference on Advanced Computing Networking and Informatics: ICANI-2018*, 870, 311.
- Kyriakou, K. I. (2019). Enhancing C/C++ based OSS development and discoverability with CBRJS: A Rust/Node.js/WebAssembly framework for repackaging legacy codebases. *Journal of Systems and Software*, 157, 110395.
- Le, Q. T. (2015). Application of Web Scraping and Google API service to optimize convenience stores' distribution. *In 2015 17th International Conference on Advanced Communication Technology (ICACT)*, 478-482.

- Mahajan, S. &. (2015). A Web Scraping Approach in Node.js. *Bangalore: International Journal of Science, Engineering and Technology*, 4(4), 909–912.
- Meier, J. D. (2010). *Quantifying End-User Response Time Goals*, Microsoft Developer Network. ([www.msdn.microsoft.com/en-us/library/bb924365.aspx](http://www.msdn.microsoft.com/en-us/library/bb924365.aspx), diakses 25 Juni 2021).
- Munthe, I. R. (2020). UML Modeling and Black Box Testing Methods in the School Payment Information System. *Jurnal Mantik*, 4(3), 1634-1640.
- Nandini, D. (2014). *Semantic Web and Ontology*. Bookboon.
- Nielsen, J. (2010). *Website Response Times*. ([www.nngroup.com/articles/website-response-times](http://www.nngroup.com/articles/website-response-times), diakses 25 Juni 2021).
- Rosario, M. P. (2017). Penerapan Web Scraping pada WebsiteCompany Profile. *Seminar Nasional KNTIA*, 37-43.
- Sanfilippo, E. M. (2019). Ontology-based knowledge representation for additive manufacturing. *Computers in Industry*, 109, 182-194.
- Sangeetha, K. &. (2015). A review paper on software effort estimation methods. *International Journal of Scientific Engineering and Applied Science (IJSEAS)*, 1(3), 163-169.
- Satria, A. H. (2016). Analisis Keterhubungan Ontology Pada Web Semantik Menggunakan Semantic-based Ontology Matching. *eProceedings of Engineering*, 3(3), 5345.
- Sundaramoorthy, K. D. (2017). Newsone—an aggregation system for news using web scraping method. *In 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC)*, 136-140.
- Supriyono, S. (2019). Penerapan ISO 9126 Dalam Pengujian Kualitas Perangkat Lunak pada E-book. *MATICS*, 11(1), 9-13.
- Vishwakarma, D. K. (2019). Detection and veracity analysis of fake news via scraping and authenticating the web search. *Cognitive Systems Research*, 58, 217-229.

Alexa.com. (2021). *Top Sites in Indonesia*. [www.alexacom/topsites/countries/ID](http://www.alexacom/topsites/countries/ID), diakses 23 Mei 2021.