

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Era globalisasi saat ini ditandai dengan perkembangan teknologi yang meningkat pesat. Ini merupakan salah satu penyebab orang-orang lebih suka menggunakan alat komunikasi berbasis komputer dan perkembangan teknologi internet agar lebih praktis dalam pemanfaatannya (Sibagariang, 2017). Kebiasaan baru dalam masyarakat sekarang adalah mencari informasi melalui internet karena data dan fakta pada internet sangat beragam dan dapat diakses kapan saja (Sonya and Prihandoko, 2016). Internet merupakan suatu jaringan yang sangat besar dan membentuk jaringan komputer yang saling terhubung di seluruh dunia. Di dalam internet terdapat *website* yang menjadi salah satu sumber informasi yang dapat memenuhi kebutuhan masyarakat untuk memperoleh informasi secara cepat, tepat dan relevan (Deddy and Mailoa, 2020).

*Website* berperan penting bagi para aktivis akademik untuk mempublikasikan suatu artikel ilmiah yang telah diteliti sebelumnya untuk memberikan informasi yang berhubungan dengan bidang pendidikan. Publikasi tersebut dilakukan oleh para aktivis akademik ke dalam jurnal ataupun prosiding. Jurnal berisi artikel dari hasil penelitian yang sudah selesai, yang diterbitkan oleh suatu lembaga pendidikan, lembaga penelitian atau lembaga lain yang bekerjasama dengan asosiasi profesi tertentu (Budiman, 2015). Terdapat berbagai macam artikel yang telah dipublikasi oleh para sivitas akademik di dalam jurnal salah satunya yaitu artikel ilmiah.

Artikel ilmiah memberikan informasi dibidang akademik, dimana artikel ilmiah bertujuan untuk memberikan informasi yang mengedukasi dan menambah wawasan

bagi para pembaca. Data yang diperoleh pada suatu artikel dapat digunakan sebagai sumber informasi yang dikutip pada artikel yang akan dibuat oleh peneliti lain, atau biasa disebut sitasi. Terdapat berbagai macam artikel yang ada dapat dilihat dari media tulisannya atau sitasi. Sitasi digunakan untuk mengetahui karakteristik komunikasi ilmu pengetahuan dan aspek kualitatif yang berasal dari penelitian dan publikasi (Ilham, 2019) untuk menggunakan informasi dan wawasan dari teks sebagai dasar keputusan dan pemikiran kreatif, serta peneliti juga dapat menghindari diri dari plagiat sebuah karya dan lebih menghargai kutipan-kutipan yang diambil dari karya orang lain, karena nilai kejujuran penting dalam sebuah karya (Ilham, 2019).

Banyaknya informasi artikel ilmiah yang ada di suatu jurnal atau prosiding pada internet membuat para pengguna mengalami kesulitan untuk menentukan referensi yang sesuai dengan apa yang dibutuhkan. Untuk itu dibutuhkan suatu program yang dapat memperoleh informasi yang dibutuhkan sesuai dengan kata kunci yang dimasukkan, agar informasi yang dihasilkan tepat dan sesuai (Ramadhan and Sastramihardja, 2018). Sehubungan dengan keterbatasan sumber daya komputasi dan waktu, maka dibutuhkan suatu cara untuk mengambil konten yang ada di internet dalam waktu yang singkat dan dapat di indeks secara otomatis serta tersimpan pada *database* dengan membangun sebuah sistem untuk menelusuri satu atau lebih halaman dengan pengambilan informasi yang tersebar dan selalu berubah-ubah di internet dalam jumlah besar. Untuk membangun sistem tersebut diperlukan sebuah *web crawling* (Halim *et al*, 2017).

*Web Crawling* merupakan suatu program yang dapat mengunduh seluruh halaman pada suatu *website*. *Web crawling* digunakan untuk aplikasi pencarian artikel

berbasis *website* untuk halaman *website* dengan jumlah berskala besar, dimana *web crawling* akan secara berkala mengumpulkan dan mengarsipkan seluruh *website* tersebut (Deddy and Mailoa, 2020). Hampir seluruh mesin pencari yang ada sekarang menggunakan konsep *crawler* untuk mengumpulkan informasi dari internet sebagai komponen utama mesin pencari tersebut (Ramadhan and Sastramihardja, 2018) sehingga *web crawler* dapat dijadikan sebuah alat untuk mengambil konten-konten yang ditampilkan oleh halaman *website*, kemudian konten-konten tersebut dikelompokkan dalam satu atau lebih atribut (Ilmawan, 2018). *Web crawling* sangat cocok diterapkan pada lembaga pendidikan yang berhubungan langsung dengan pengumpulan informasi publikasi artikel ilmiah di dalam suatu *website* seperti Lembaga Penelitian dan Pengabdian Kepada Masyarakat.

Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) merupakan wadah bagi aktivitas akademik untuk melaksanakan penelitian dan pengabdian kepada masyarakat. Pada tahun 2010 LPPM berada di bawah koordinasi Wakil Rektor Bidang Riset, Inovasi Kemitraan (WRRIM). LPPM dibentuk sebagai upaya untuk mengatasi ketidakseimbangan antara kemampuan IPTEKS di Universitas dan peluang yang ada di masyarakat, selain itu pembentukan LPPM juga didasari untuk memwadahi pelaksanaan tugas TriDharma yang meliputi pendidikan, penelitian, dan pengabdian kepada masyarakat secara seimbang (LPPM ITB, 2018).

Berdasarkan wawancara yang dilakukan pada LPPM Universitas XYZ pihak LPPM membutuhkan informasi artikel ilmiah salah satunya adalah nama artikel, nama penulis artikel, judul artikel dan tahun terbit dari artikel. Informasi data artikel tersebut dibutuhkan untuk dijadikan acuan dalam melakukan proses perekapan data artikel yang

ada pada Universitas XYZ. Pencarian informasi artikel sebelumnya dilakukan secara manual oleh bagian LPPM melalui Google Scholar, SINTA dan Scopus sehingga pihak LPPM mengalami kesulitan dalam proses pencarian dan perekapan data jurnal untuk ranking universitas sesuai dengan periode yang dibutuhkan.

Sebelumnya *web crawling* telah berhasil di implementasi oleh beberapa peneliti yaitu penelitian milik (Nasution, 2020) yang berjudul Sistem Pengumpulan Data Publikasi Ilmiah Menggunakan Metode Web Crawling. Menjelaskan tentang kesulitan universitas dalam membuat *roadmap* penelitian dan publikasi dikarenakan sulitnya mengumpulkan data publikasi ilmiah untuk dilakukan pendataan publikasi ke dalam sistem untuk meningkatkan nilai ranking Universitas. Pada penelitian ini sistem berhasil dengan 10 kali percobaan, dimana 9 kali percobaan berhasil dan 1 kali percobaan berhenti karena koneksi ke server SINTA mengalami gangguan. Penelitian selanjutnya yaitu milik (Sintaro *et all*, 2020) yang berjudul Penerapan Web Walkers Sebagai Media Informasi Untuk Perbandingan Manual Brewing. Web Walkers ini digunakan untuk mengumpulkan data yang didapat dari teks yang diambil dari sosial media yang ada, masyarakat yang menuliskan atau memperbaharui status di sosial medianya akan menjadi informasi yang dapat digunakan. Penelitian ini menghasilkan sebuah mesin pencari *tags* yang akan menampilkan data segmentasi manual brewing berupa grafik. Dan penelitian yang dibuat oleh (Hanifah and Nurhasanah, 2018) yang berjudul Implementasi Web Crawling Untuk Mengumpulkan Informasi Wisata Kuliner di Bandar Lampung. Menjelaskan tentang penggunaan teknik web crawling untuk mengumpulkan informasi wisata kuliner di Bandar Lampung pada situs Trip Advisor untuk dilakukan pemetaan dan pengembangan promosi wisata kuliner yang belum

dilakukan oleh pemerintah kota Bandar Lampung dalam meningkatkan angka kunjungan wisatawan di Bandar Lampung. Hasil rincian informasi terdapat beberapa informasi nomor telepon dan website yang tidak ada, dengan rata-rata informasi yang lengkap dan berhasil di dapat adalah 71,1%.

Solusi dari masalah yang dihadapi oleh pihak LPPM, maka pada penelitian ini akan menerapkan program dengan menggunakan *web crawling* untuk mengumpulkan informasi artikel ilmiah, dengan adanya program *web crawling* ini proses pencarian informasi artikel ilmiah dapat dilakukan secara otomatis oleh program dan hasil yang didapat dari program *web crawling* berupa informasi halaman web yang diambil secara otomatis seperti *text* yaitu judul artikel, nama penulis, dan tahun artikel yang diinformasikan pada masing-masing artikel. Serta mempermudah proses pencarian artikel ilmiah menjadi lebih efisien daripada proses pencarian data yang dilakukan secara manual. Selain itu keunggulan *web crawling* adalah jumlah dan akurasi informasi yang didapat. Karena dengan menggunakan program *web crawling* ini data artikel yang didapat lebih banyak dan akurat sehingga data yang diperoleh dapat diolah lebih lanjut.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka didapat suatu masalah yaitu bagaimana penerapan *web crawling* untuk mengumpulkan informasi artikel ilmiah pada LPPM Universitas XYZ yang telah dipublikasi oleh para dosen sesuai dengan periode yang dibutuhkan?

### 1.3 Batasan Masalah

Setiap penelitian yang sifatnya ilmiah sudah tentu memiliki batasan masalah. Batasan masalah dalam penelitian ini antara lain adalah :

1. Sumber data pada penelitian diambil dari *google scholar*.
2. Informasi yang ditampilkan berupa nama penulis, tahun, judul artikel, nama jurnal atau prosiding, info jurnal dan jumlah sitasi.
3. Sistem aplikasi hanya dapat melakukan pencarian berdasarkan nama dosen, fakultas, dan prodi.
4. Metode pengembangan sistem yang digunakan adalah *Extreme Programming (XP)*.
5. Bahasa pemrograman yang digunakan adalah *PHP* untuk tampilan hasil *crawling*.
6. Bahasa pemrograman yang digunakan untuk proses *crawling* adalah *Python*.
7. Library *python* yang digunakan *beautifulsoup4*.
8. Database yang digunakan adalah *MySQL*.

### 1.4 Tujuan Penelitian

Setiap penelitian yang sifatnya ilmiah sudah tentu mempunyai tujuan dan sasaran. Tujuan dalam penelitian ini antara lain adalah:

1. Mempermudah pihak LPPM dalam membuat rekapitulasi artikel.
2. Memberikan informasi seputar artikel ilmiah yang telah dipublikasi di *Google Scholar*.

### 1.5 Manfaat Penelitian

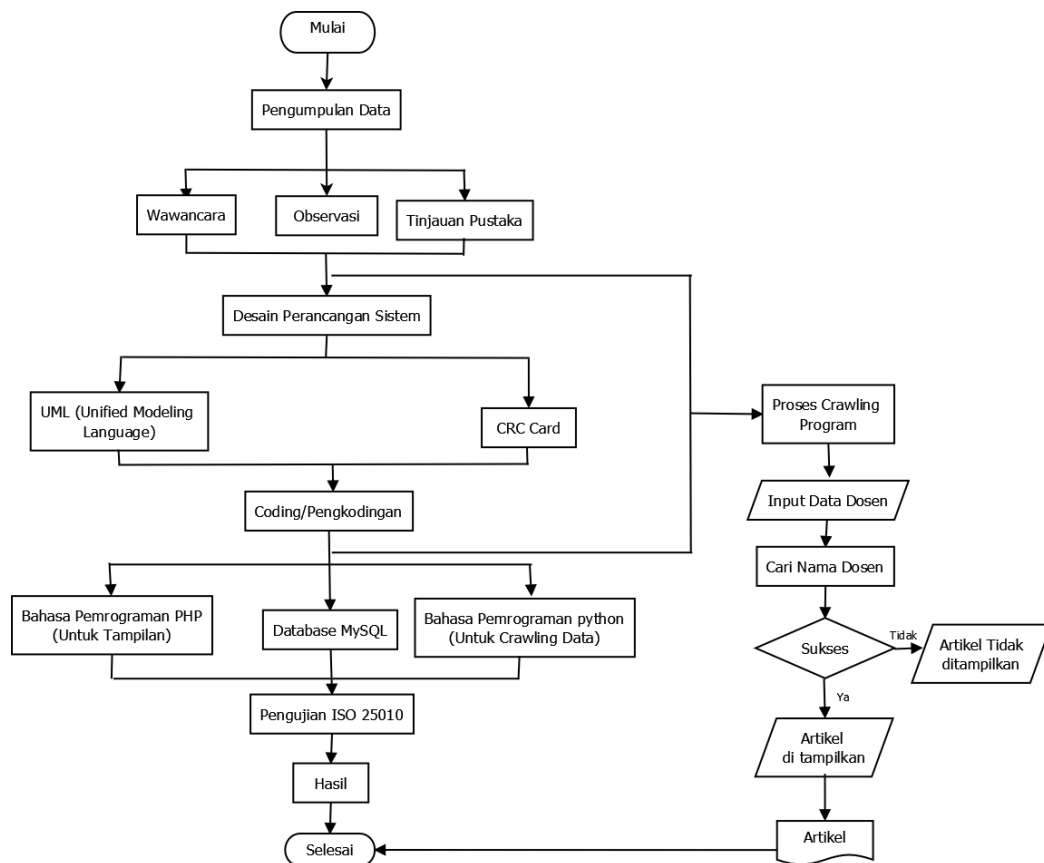
Manfaat yang diharapkan dari hasil penelitian ini adalah :

1. Mempercepat proses pelaporan artikel ilmiah.
2. Membantu pihak LPPM dalam mendapatkan informasi yang akan digunakan untuk dijadikan acuan untuk membantu proses bisnis secara baik.
3. Mempermudah pihak LPPM dalam penginformasian identitas artikel ilmiah yang dibutuhkan oleh Universitas XYZ.

### 1.6 Tahapan Penelitian

Tahapan penelitian merupakan tahapan yang berisi dasar pemikiran dalam pembuatan suatu penelitian. Tahapan penelitian dapat dilihat pada **Gambar 1.1** yang menunjukkan bahwa proses awal pada tahapan penelitian dimulai dari pengumpulan data dengan melakukan wawancara oleh pihak LPPM, observasi, dan tinjauan pustaka. Selanjutnya dilakukan perancangan desain sistem yang akan dibuat dengan menggunakan UML (*Unified Modeling Language*) dan *CRC Card*. Setelah desain dibuat maka selanjutnya ke tahap pengkodean dengan menggunakan bahasa pemrograman *python* untuk melakukan *crawling* data, menggunakan bahasa pemrograman *PHP* untuk tampilan program dan *database* penyimpanan menggunakan *MySQL*. Dari tahap perancangan desain sistem dan pengkodean menghasilkan sebuah sistem dengan menggunakan *crawling*, dengan kerja sistem yaitu melakukan input nama dosen, kemudian sistem akan mencari nama dosen yang diinputkan, jika terdapat kesalahan pada nama dosen yang diinputkan maka artikel tidak akan ditampilkan, sebaliknya jika nama dosen terdapat di dalam sistem maka artikel akan

ditampilkan dengan bentuk informasi data artikel. Tahap selanjutnya yaitu dilakukan pengujian sistem dengan menggunakan ISO 25010. Setelah dilakukan pengujian maka sistem pengumpulan informasi artikel ilmiah telah berhasil dibuat dan dapat digunakan oleh pihak LPPM.



**Gambar 1.1 Tahapan Penelitian**

## 1.7 Sistematika Penulisan

Sistematika penulisan berisi pedoman dalam penulisan skripsi yang dibuat untuk mempermudah dalam mengetahui pembahasan pada skripsi secara menyeluruh. Sistematika penulisan pada skripsi ini, disusun sebagai berikut :



## 1. Bagian Awal Skripsi

Bagian awal skripsi mencakup halaman sampul depan judul, lembar pengesahan, lembar pernyataan, lembar pernyataan persetujuan publikasi, kata pengantar, daftar isi, daftar tabel, daftar gambar, daftar lampiran, dan intisari dari skripsi.

## 2. Bagian Utama Skripsi

Dalam bagian utama skripsi mencakup bab dan sub bab yaitu sebagai berikut :

### BAB I PENDAHULUAN

Dalam bagian pendahuluan mencakup latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tahapan penelitian, dan sistematika penulisan.

### BAB II LANDASAN TEORI

Dalam bagian landasan teori di BAB II berisi tinjauan pustaka yang digunakan sebagai referensi pada penelitian. Landasan teori juga berisi teori-teori pendukung yang digunakan pada penelitian.

### BAB III ANALISIS DAN PERANCANGAN SISTEM

Pada analisis dan perancangan sistem mencakup metode pengumpulan data pada penelitian, analisis kebutuhan sistem, rancangan desain program, dan kerangka pengujian sistem.

### BAB IV IMPLEMENTASI SISTEM

### BAB V HASIL DAN PEMBAHASAN

Hasil dan pembahasan mencakup hasil penelitian, pengujian, dan pembahasan / evaluasi pengujian sistem.

### BAB VI KESIMPULAN DAN SARAN