

IDENTIFIKASI BIDANG KEPAKARAN PROGRAM STUDI INFORMATIKA DI INDONESIA MENGGUNAKAN RAKE

LAPORAN TUGAS AKHIR

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



Disusun Oleh:

ACHMAD NAUFAL FADHILAH

NIM 32601900003

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

2023

***IDENTIFICATION OF EXPERTISE IN INFORMATICS STUDY
PROGRAM IN INDONESIA USING RAKE***

*Proposed to complete the requirement to obtain a bachelor's degree (S-1)
at Informatics Engineering Department Industrial Technology Faculty*

Sultan Agung Islamic University



Arranged By:

ACHMAD NAUFAL FADHILAH

NIM 32601900003

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

2023

LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul **"Identifikasi Bidang Kepakaran Program Studi Informatika Di Indonesia Menggunakan RAKE"** ini disusun oleh:

Nama Achmad Naufal Fadhilah

NIM 32601900003

Program Studi · Teknik Informatika

Telah disetujui oleh Dosen Pembimbing pada:

Hari Senin

Tanggal : 28 Agustus 2023

Mengesahkan,

Pembimbing I



Pembimbing II



Imam Much Ibnu Subroto, ST.M.Sc,Ph.D. Sam Farisa Chaerul Haviana, ST, M.Kom

NIDN. 0613037301

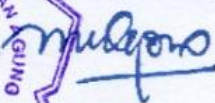
NIDN. 0628028602

Mengetahui,

Ketua Program Studi Teknik Informatika
Fakultas Teknologi Industri

Universitas Islam Sultan Agung




Ir. Sri Mulyono, M.Eng
NIDN. 0626066601

LEMBAR PENGESAHAN PENGUJI

Laporan tugas akhir dengan judul “**Identifikasi Bidang Kepakaran Program Studi Informatika Di Indonesia Menggunakan RAKE**” ini telah dipertahankan di depan tim penguji Tugas Akhir pada:

Hari : Senin

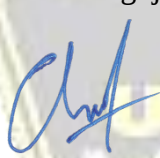
Tanggal : 27 November 2023

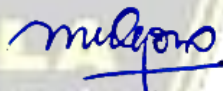


TIM PENGUJI

Ketua Penguji

Anggota I


Andi Riansyah, ST.M.Kom
NIDN 0609108802


Ir. Sri Mulyono, M.Eng
NIDN. 0626066601

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini:

Nama Achmad Naufal Fadhilah

NIM 32601900003

Judul Tugas Akhir Identifikasi Bidang Kepakaran Program Studi Informatika
Di Indonesia Menggunakan RAKE

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

Semarang, 4 Desember 2023

Yang menyatakan,



Achmad Naufal Fadhilah

KATA PENGANTAR

Dengan mengucapkan syukur alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Tugas Akhir dengan judul “Identifikasi Bidang Kepakaran Program Studi Informatika Di Indonesia Menggunakan RAKE” ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang. Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terima kasih kepada :

1. Bapak, Ibu, dan Kakak yang selalu memberikan motivasi, semangat, doa serta dukungan kepada saya.
2. Bapak Dr. Imam M Ibnu Subroto, ST, M.Sc, Ph.D. selaku Dosen Pembimbing I yang telah meluangkan waktu dan memberikan ilmu kepada penulis.
3. Bapak Sam Farisa Chaerul Haviana, ST., M.kom. selaku Dosen Pembimbing II yang telah meluangkan waktu dan memberikan ilmu kepada penulis.
4. Para Dosen FTI Unissula yang telah memberikan banyak ilmu yang bermanfaat.
5. Teman-teman yang telah memberikan semangat kepada penulis.

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

Semarang, 4 Desember 2023



Achmad Naufal Fadhilah

DAFTAR ISI

HALAMAN JUDUL

FINAL PROJECT

LEMBAR PENGESAHAN PEMBIMBING	iii
LEMBAR PENGESAHAN PENGUJI.....	iv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL.....	xii
ABSTRAK	xiii
<i>ABSTRACT</i>	xiv
BAB I.....	15
PENDAHULUAN.....	15
1. 1 Latar Belakang	15
1. 2 Perumusan Masalah.....	16
1. 3 Pembatasan Masalah	16
1. 4 Tujuan.....	17
1. 5 Manfaat.....	17
1. 6 Sistematika Penulisan.....	17
BAB II	19
TINJAUAN PUSTAKA DAN DASAR TEORI	19
2. 1 Tinjauan Pustaka.....	19

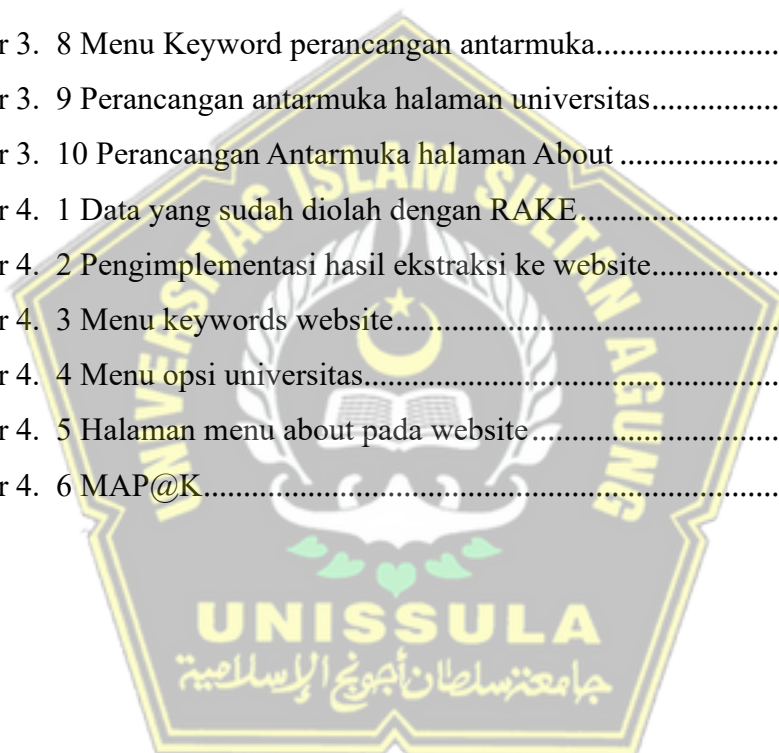
2.2	Dasar Teori	21
2.2.1	Data Mining	21
2.2.2	Scopus	22
2.2.3	RAKE (<i>Rapid Automatic Keyword Extraction</i>)	23
2.2.4	MAP@K (<i>Mean Average Precision@K</i>)	24
2.2.5	SINTA (<i>Science and Technology Index</i>)	24
BAB III		26
METODE PENELITIAN		26
3.1	Deskripsi Penelitian.....	26
3.2	Metode Ekstraksi Kepakaran.....	27
3.2.1	Algoritma RAKE	27
3.3	Metode Pengembangan Sistem	32
3.3.1	<i>Requirements Analysis</i>	33
3.3.2	<i>Design</i>	34
3.3.3	Pembentukan <i>Prototyping</i>	42
3.3.4	<i>Customer Evaluation</i>	43
3.3.5	<i>Review dan Refine</i>	44
3.3.6	<i>Develop</i>	44
3.3.7	Rencana Pengujian Sistem	44
3.3.8	<i>Release</i>	45
BAB IV		46
HASIL DAN ANALISIS PENELITIAN		46
4.1	Analisa Hasil	46
4.2	Implementasi Sistem	48
4.3	Pengujian Sistem	51

BAB V	54
KESIMPULAN DAN SARAN	54
5. 1 Kesimpulan.....	54
5. 2 Saran	54
DAFTAR PUSTAKA.....	55
LAMPIRAN	



DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	26
Gambar 3. 2 Alur Proses Ekstraksi RAKE.....	27
Gambar 3. 3 Model Prototype.....	33
Gambar 3. 4 Flowchart sistem untuk admin	35
Gambar 3. 5 Flowchart sistem untuk user.....	36
Gambar 3. 6 ERD model.....	38
Gambar 3. 7 Use Case sistem.....	39
Gambar 3. 8 Menu Keyword perancangan antarmuka.....	40
Gambar 3. 9 Perancangan antarmuka halaman universitas.....	41
Gambar 3. 10 Perancangan Antarmuka halaman About	42
Gambar 4. 1 Data yang sudah diolah dengan RAKE.....	46
Gambar 4. 2 Pengimplementasi hasil ekstraksi ke website.....	47
Gambar 4. 3 Menu keywords website.....	49
Gambar 4. 4 Menu opsi universitas.....	50
Gambar 4. 5 Halaman menu about pada website.....	51
Gambar 4. 6 MAP@K.....	53



DAFTAR TABEL

Tabel 3. 1 Text Preprocessing	28
Tabel 3. 2 Tabel Delimiters	29
Tabel 3. 3 Perhitungan Score Keywords	30
Tabel 3. 4 Hasil Perhitungan $\deg(w)$ dan $\text{freq}(w)$	31
Tabel 3. 5 Expertise word.....	31
Tabel 3. 6 Tabel database penelitian.....	39
Tabel 4. 1 Data terkumpul.....	47
Tabel 4. 2 Tabel Hasil Pengujian.....	51
Tabel 4. 3 Mean Average Precision.....	53



ABSTRAK

Program studi pada universitas di Indonesia baik universitas swasta maupun universitas negeri terkadang memiliki bidang kepakaran yang berbeda meskipun nama program studi tersebut terlihat sama, hal ini karena program studi diberbagai universitas sangat tergantung dengan kepakaran dosen pada universitas tersebut. Untuk itu, diperlukan sebuah sistem yang memuat informasi tentang bidang kepakaran program studi khususnya Informatika untuk membantu para calon mahasiswa untuk lebih mudah memperoleh informasi program studi informatika dan universitas mana yang sesuai dengan minat para calon mahasiswa tersebut. Dengan memanfaatkan metode RAKE (*Rapid Automatic Keyword Extraction*) dapat mewujudkan kemudahan memperoleh informasi. Metode ini dipilih karena sangat cocok untuk digunakan meranking seberapa tinggi bidang kepakaran program studi informatika di Indonesia. Pembuatan sistem diawali dengan mengumpulkan data dari beberapa *website* seperti SINTA Kemendikbud dan Scopus, implementasi metode dan pembuatan sistem dengan bahasa pemrograman *python* dan MySQL, dan pengujian sistem. Hasil penelitian menunjukkan bahwa sistem informasi dapat berjalan dengan baik setelah pengujian dan menunjukkan hasil berupa informasi kepakaran program studi ketika *user* memasukkan kata kunci ataupun nama program studi yaitu informatika. sistem berhasil menampilkan sistem informasi kepakaran program studi berdasarkan universitas yang dipilih dengan menampilkan kata kunci yang berhasil dihitung berjumlah 2716 dari 5 universitas di Indonesia dan kata kunci setiap universitas hanya ditampilkan sebanyak 20 kata kunci sesuai perhitungan *mean average precision@k*.

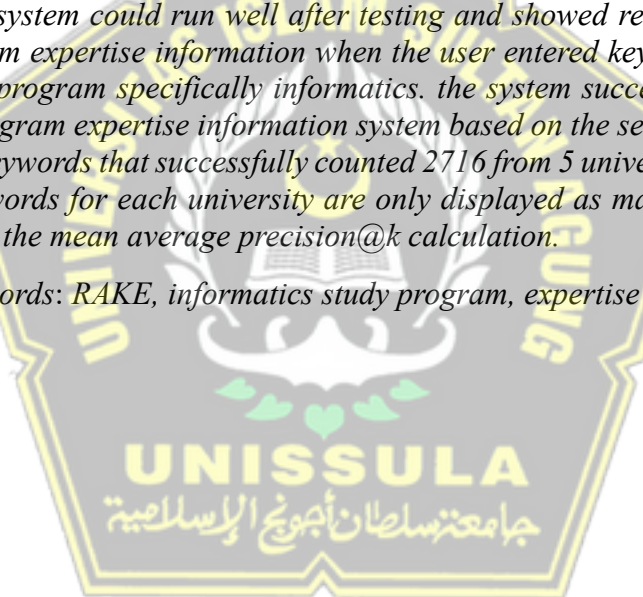
Kata kunci: RAKE, program studi informatika, kepakaran



ABSTRACT

Study programs at universities in Indonesia, both private and state universities, sometimes have different areas of expertise even though the names of the study programs look the same, this is because study programs at various universities depend heavily on the expertise of the lecturers at the university. For this reason, a system is needed that contains information about the area of expertise of the study program, especially Informatics, to help prospective students more easily obtain information on which informatics study programs and universities suit the interests of these prospective students. By utilizing the RAKE (Rapid Automatic Keyword Extraction) method, you can realize the ease of obtaining information. This method was chosen because it is very suitable to be used to rank how high the area of expertise of informatics study programs in Indonesia is. Making the system begins with collecting data from several websites such as SINTA Kemendikbud and Scopus, implementing methods and creating a system with the Python and MySQL programming languages, and system testing. The results showed that the information system could run well after testing and showed results in the form of study program expertise information when the user entered keywords or the name of the study program specifically informatics. the system succeeded in displaying the study program expertise information system based on the selected university by displaying keywords that successfully counted 2716 from 5 universities in Indonesia and the keywords for each university are only displayed as many as 20 keywords according to the mean average precision@k calculation.

Keywords: RAKE, informatics study program, expertise



BAB I

PENDAHULUAN

1.1 Latar Belakang

Menurut kamus *online Webster*, pakar adalah seseorang yang “memiliki, melibatkan, atau menampilkan keterampilan atau pengetahuan khusus yang diperoleh dari pelatihan atau pengalaman”. Istilah keahlian sering digunakan dalam arti relatif dalam penelitian pendidikan. Keahlian dapat didefinisikan dalam hal kinerja luar biasa dalam suatu domain; penelitian kinerja ahli menyelidiki kinerja unggul secara konsisten dari individu yang unggul dalam tugas perwakilan dalam domain itu. (Teodoro dan Schwartz, 2012)

Informatika merupakan ilmu yang berkonsentrasi pada pemanfaatan komputer untuk memilah dan meneliti informasi yang dalam jumlah besar, baik informasi maupun data pada mesin berbasis komputasi. Informatika juga mencakup peninjauan, perencanaan, dan pembuatan sistem komputasi dasar yang dapat berinteraksi dan menyajikan data secara efektif dan efisien.

Ilmu komputer, ilmu informasi, sistem informasi, teknik komputer, dan aplikasi informasi dalam sistem informasi manajemen semuanya termasuk dalam bidang ilmu yang sangat luas dan multidisiplin yang dikenal dengan istilah informatika. Informatika juga mengelola aspek kognitif dan sosial dari penggunaan data dalam lingkungan yang berbeda, misalnya bioinformatika, informatika klinis, dan informatika perpustakaan.

Namun meskipun memiliki nama jurusan yang sama, terkadang bidang kepakaran yang diajarkan berbeda, sehingga membuat calon mahasiswa mengalami kesulitan menentukan program studi apa dan universitas mana yang ingin dituju. Contohnya bidang kepakaran program studi Informatika pada Unissula dan Undip berbeda, karena apa yang diajarkan dosen informatika pada kedua universitas tersebut juga disesuaikan kepakaran dari dosen itu sendiri

Selama ini, belum pernah diketahui *rule* pemilihan program studi yang dilakukan oleh calon mahasiswa. Sehingga belum dapat diketahui strategi khusus yang akan dilakukan pada kegiatan promosi. Metode RAKE (*Rapid Automatic*

Keyword Extraction) sudah digunakan di beberapa penelitian sebelumnya, namun belum ada yang menggunakannya untuk penentuan kepakaran program studi pada universitas di Indonesia.

Pada *website* SINTA (*Science and Technology Index*) sudah mencantumkan kinerja dosen, peneliti, perguruan tinggi, program studi, dan jurnal Indonesia. Karena SINTA merupakan *framework* data *website* yang menawarkan akses cepat, sederhana dan menyeluruh (Wahyudi dan Berlilana, 2022). Namun SINTA masih belum ada sistem informasi untuk bidang kepakaran suatu program studi universitas sehingga para calon mahasiswa tidak cukup informasi.

Dengan alasan itulah, penulis mengusulkan membuat sistem yang bisa membantu para calon mahasiswa untuk lebih mudah mengetahui bidang kepakaran dari suatu program studi khususnya informatika pada universitas di Indonesia, dengan dibuatnya sistem ini diharapkan bisa membantu para calon mahasiswa lebih mudah memperoleh informasi tentang bidang kepakaran suatu program studi informatika dan dapat lebih mudah menentukan program studi informatika dan universitas mana yang sesuai dengan minat para calon mahasiswa tersebut.

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah belum adanya sistem informasi tentang kepakaran bidang studi khususnya informatika di Indonesia yang dimana hal tersebut berdampak negatif untuk para calon mahasiswa untuk memilih program studi sesuai minat dan bakat.

1.3 Pembatasan Masalah

Pembatasan masalah yang diterapkan pada penelitian ini, yaitu:

1. Sistem ini berupa *website* untuk mengidentifikasi bidang kepakaran program studi informatika di Indonesia
2. Kepakaran diambil berdasarkan data dari ekstraksi kata kunci judul jurnal publikasi yang bersumber dari SINTA yang terindeks scopus sehingga tidak dilakukan penelitian untuk validasi data.
3. Penelitian ini mengambil data dari lima universitas di Indonesia yaitu Unissula, Udinus, UKSW, Upgris, UNIKA

1.4 Tujuan

Tujuan tugas akhir ini adalah mengembangkan sistem pengidentifikasi bidang kepakaran program studi informatika di Indonesia dengan mengimplementasikan metode RAKE.

1.5 Manfaat

Manfaat penelitian tugas akhir adalah menerapkan metode RAKE (*Rapid Automatic Keyword Extraction*) ke dalam sistem identifikasi bidang kepakaran program studi informatika di Indonesia untuk mempermudah memperoleh informasi tentang kepakaran bidang studi khususnya informatika di Indonesia.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan dilakukan penulis dalam pembuatan laporan tugas akhir adalah sebagai berikut:

BAB 1 : PENDAHULUAN

Pada bab ini penulis memaparkan latar belakang, perumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan sehingga berguna bagi pembaca untuk mengetahui sekilas tentang penelitian yang akan dibuat.

BAB 2 : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bab ini pembaca dapat lebih memahami bagaimana teori berhubungan dengan data mining, scopus, RAKE, dan SINTA Kemendikbud dengan berdasarkan dari penelitian sebelumnya..

BAB 3 : METODE PENELITIAN

Pada bab ini berisikan proses tahapan-tahapan penelitian mulai dari deskripsi sistem sampai dengan perancangan antarmuka yang berguna untuk mengetahui proses secara terstruktur.

BAB 4 : HASIL DAN ANALISIS PENELITIAN

Pada bab ini penulis mengungkapkan hasil penelitian yakni hasil identifikasi bidang kepakaran program studi informatika menggunakan metode RAKE beserta pengujian sistem

BAB 5 : KESIMPULAN DAN SARAN

Bab ini penulis memaparkan kesimpulan daripada proses penelitian dari awal hingga hasil dari penelitian, juga terdapat saran untuk penelitian selanjutnya agar penelitian ini dapat berlanjut seiring perkembangan jaman.



BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Menurut kamus *online Webster*, pakar adalah seseorang yang “memiliki, melibatkan, atau menampilkan keterampilan atau pengetahuan khusus yang diperoleh dari pelatihan atau pengalaman”. Istilah keahlian sering digunakan dalam arti relatif dalam penelitian pendidikan. Keahlian dapat didefinisikan dalam hal kinerja luar biasa dalam suatu domain; penelitian kinerja ahli menyelidiki kinerja unggul secara konsisten dari individu yang unggul dalam tugas perwakilan dalam domain itu. (Teodoro dan Schwartz, 2012)

Data mining banyak diterapkan untuk menyelesaikan berbagai permasalahan di berbagai bidang, termasuk di bidang pendidikan atau akademik. Banyak peneliti yang telah menerapkan berbagai metode data mining untuk membantu menyelesaikan permasalahan di bidang akademik antara lain untuk menganalisis kelulusan mahasiswa, prediksi kelulusan mahasiswa, evaluasi akademik mahasiswa, rekomendasi pemilihan konsentrasi program studi, penerimaan mahasiswa baru, hingga rekomendasi pemilihan mata kuliah. (Syahrul dan Solichin, 2022)

Dengan mengambil data dari scopus yang akan di olah menggunakan metode RAKE kemudian menghasilkan berupa grafik kepakaran bidang program studi informatika. Scopus adalah kumpulan data teoretis dan referensi Elsevier yang dikirimkan pada tahun 2004. Scopus mencakup hampir 36.377 judul dari sekitar 11.678 penerbit (22.794 judul aktif dan 13.583 judul tidak aktif), 34.346 di antaranya merupakan jurnal *peer-review* dalam bidang subjek tingkat tinggi: kehidupan sains, sosiologi, sains aktual, dan sains kesejahteraan. Ini menggabungkan tiga jenis sumber: jurnal, seri buku, dan jurnal perdagangan. Semua jurnal yang diingat berada dalam *database* Scopus, tanpa memperhatikan di mana identitas mereka didistribusikan, diaudit setiap tahun untuk menjamin prinsip-prinsip terbaik dipatuhi. (Wahyuni, 2018)

Database kutipan dan abstrak yang diawasi oleh para profesional industri

merupakan pengertian dari Scopus. Berdasarkan data situs Scopus (www.scopus.com) yang diperoleh pada 3 Juni 2020, informasi yang terdapat dalam Scopus terdiri dari 39.744 jurnal ilmiah aktif, 29.683 judul, 1.499 judul seri buku dengan 59.698 *volume*, pertukaran distribusi lebih dari 300 judul dan artikel yang didistribusikan oleh distributor di seluruh dunia berjumlah lebih dari 8.000 judul. Terdapat juga 221.169 judul buku monografi atau buku teks di Scopus. Informasi di Scopus berubah secara konsisten untuk menyesuaikan diri dengan informasi tambahan karena distribusi baru atau penurunan informasi karena distribusi yang rusak. (Hakim, 2020)

Sistem ini menggunakan metode *Rapid Automatic Keyword Extraction* (RAKE) adalah algoritma yang digunakan untuk mengekstraksi kata kunci atau frasa penting dari sebuah teks secara otomatis. Algoritma ini dirancang untuk mengidentifikasi kata-kata kunci berdasarkan karakteristik linguistik seperti frekuensi kata, kehadiran kata dalam konteks yang relevan, serta pola penempatan kata yang spesifik. metode tanpa pengawasan, domain-independen, dan bahasa-independen untuk mengekstraksi kata kunci dari dokumen individual. (Rose dan Engel, 2010) Dalam *Rapid Automatic Keyword Extraction* (RAKE), kata kunci memiliki banyak kata konten yang informatif daripada tanda baca dan kata henti. RAKE mengumpulkan kata kerja dalam kandidat kata kunci.

Kinerja RAKE lebih baik dari Textrank. Ini karena kemampuan RAKE untuk merangking kata kunci dalam satu langkah sementara Textrank membutuhkan iterasi terus menerus untuk mendapatkan konvergensi pada skor kata. Keunggulan RAKE adalah efisiensi komputasinya, kecepatannya, presisinya meskipun sederhana dan kemampuannya untuk bekerja pada dokumen individual. (Thushara, Mownika dan Mangamuru, 2019)

Dengan menggunakan matriks kemunculan bersama, metode RAKE menguji asosiasi kata. Kerangka kerja ini digunakan untuk mengukur skor kata yang signifikan untuk dilakukan perankingan. Kata-kata penting adalah bagian kalimat yang menyampaikan gagasan utama sebuah dokumen. Kata penting dimaksudkan paling sedikit satu kata sebagai kunci, sedangkan frase penting paling sedikit dua kata sebagai kunci.. (Fatmawati, 2017)

Algoritma RAKE merupakan salah satu metode yang menggunakan pendekatan unsupervised learning. Dalam proses ekstraksi kata kunci, RAKE menggunakan stoplist untuk mendapatkan daftar kandidat kata kunci dari sebuah dokumen. Kemudian skor kandidat-kata kunci dihitung menggunakan peringkat berbasis grafik.(Muhtar, Gerhana dan Maylawati, 2020)

Sistem identifikasi bidang kepakaran program studi informatika ini akan mengambil data pada *Science and Technology Index* (SINTA) karena SINTA memberikan akses masuk ke referensi dan keahlian di Indonesia. Sebuah *website* untuk mengukur kinerja dosen, analis, perguruan tinggi, program, studi dan jurnal di Indonesia.(Wahyudi dan Berlilana, 2022) Diambil dari situs resmi SINTA(*Science and Technology Index*) SINTA memiliki visi untuk menjadi pusat rujukan kinerja penelitian di Indonesia. yang didukung oleh misi untuk mengembangkan sitasi dan keahlian di Indonesia, menemukan dan mengevaluasi kekuatan penelitian masing-masing institusi, serta mengembangkan sistem analisis tren penelitian dan direktori keahlian Indonesia..(Saputra, 2020)

2.2 Dasar Teori

2.2.1 Data Mining

Machine Learning (ML) dan *Data Mining* (DM) adalah aplikasi muncul di sektor ini hampir dua dekade lalu untuk memecahkan masalah manufaktur. Sistem cerdas untuk mendukung pengambilan keputusan yang efektif program untuk menjadwalkan lini produksi secara bersamaan, pengaturan pemeliharaan mesin dapat dianggap sebagai contoh yang menggunakan metode *Machine Learning* (ML) untuk melakukan tugas manufaktur.(Dogan dan Applications, 2021)

Data mining adalah strategi yang dapat membantu memprediksi kekhasan suatu *item* yang dikonsentrasikan dengan terlebih dahulu mengetahui atau tidak mengetahui kelas *item* tertentu melalui masukan dari data besar. Suatu masyarakat dapat meningkatkan pengelolaan lingkungannya dengan menggunakan data mining.(Natasuwarna, 2019)

Untuk mencari RAKE dari suatu kumpulan data, pertama-tama harus dicari lebih dulu yang disebut "*frequent itemset*" (sekumpulan *item* yang sering muncul

bersamaan). Setelah semua pola *frequent itemset* ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat yang telah ditentukan.(Arisandi dan Nofriandi, 2019)

2.2.2 Scopus

Database kutipan dan abstrak yang diawasi oleh para profesional industri merupakan pengertian dari Scopus. Berdasarkan data situs Scopus (www.scopus.com) yang diperoleh pada 3 Juni 2020, informasi yang terdapat dalam Scopus terdiri dari 39.744 jurnal ilmiah aktif, 29.683 judul, 1.499 judul seri buku dengan 59.698 *volume*, pertukaran distribusi lebih dari 300 judul dan artikel yang didistribusikan oleh distributor di seluruh dunia berjumlah lebih dari 8.000 judul. Terdapat juga 221.169 judul buku monografi atau buku teks di Scopus. Informasi di Scopus berubah secara konsisten untuk menyesuaikan diri dengan informasi tambahan karena distribusi baru atau penurunan informasi karena distribusi yang rusak.(Hakim, 2020) Pencarian Scopus juga mencakup pencarian *database* paten. Selain itu, Scopus menawarkan empat metrik kualitas yang berbeda untuk setiap judul: yaitu h-Record, CiteScore, SJR (ScImago Diary Rank) dan Clip(Rahmadi Islam, 2018)

Platform Scopus memungkinkan akses data dengan opsi cari, temukan, dan analisis. Opsi pencarian memungkinkan mencari dokumen, penulis, dan pencarian lanjutan. Opsi Temukan memungkinkan pengguna untuk mengidentifikasi kolaborator, organisasi penelitian sehubungan dengan hasil penelitian, menemukan data publikasi terkait melalui berbagai metrik seperti kata kunci penulis, referensi bersama, dan lain-lain.

Opsi Analisis adalah alat untuk melacak kutipan, penilaian hasil pencarian berdasarkan kriteria seperti kebijaksanaan negara, kebijaksanaan afiliasi, kebijaksanaan wilayah penelitian, distribusi data yang dihasilkan. Data yang diunduh dari *database* Scopus untuk publikasi penelitian biasanya terdiri dari 43 bidang, seperti singkatan judul sumber, abstrak, kata kunci penulis, judul sumber (jurnal tempat diterbitkan), jenis dokumen, dan lain-lain.(Singh, Singh dan Karmakar, 2021)

2.2.3 RAKE (*Rapid Automatic Keyword Extraction*)

RAKE (*Rapid Automatic Keyword Extraction*) adalah algoritma yang digunakan untuk mengekstraksi kata kunci atau frasa penting dari sebuah teks secara otomatis. Algoritma ini dirancang untuk mengidentifikasi kata-kata kunci berdasarkan karakteristik linguistik seperti frekuensi kata, kehadiran kata dalam konteks yang relevan, serta pola penempatan kata yang spesifik. metode tanpa pengawasan, domain-independen, dan bahasa-independen untuk mengekstraksi kata kunci dari dokumen individual. (Rose dan Engel, 2010)

Dalam *Rapid Automatic Keyword Extraction* (RAKE), kata kunci memiliki banyak kata konten yang informatif daripada tanda baca dan kata henti. RAKE mengumpulkan kata kerja dalam kandidat kata kunci. Perhitungan dalam RAKE juga tidak sembarangan, dapat menghitung 2 hal terlebih dahulu untuk setiap kata yang ada pada tabel, yaitu:

1. *Word Frequency* ($\text{freq}(w)$)

Perhitungan *word frequency* merupakan perhitungan dari berapa kali kata tertentu muncul di antara semua kata kunci yang ada.

2. *Word Degree* ($\text{deg}(w)$)

Perhitungan *word degree* merupakan jumlah nilai dari baris suatu kata tertentu pada *co-occurrence graph*.

Kemudian untuk menghitung skor kata kunci, dapat digunakan rumus berikut ini :

$$\text{Keyword score} = \frac{\text{deg}(w)}{\text{freq}(w)} \quad (1)$$

Kinerja RAKE lebih baik dari Textrank. Ini karena kemampuan RAKE untuk merangking kata kunci dalam satu langkah sementara Textrank membutuhkan iterasi terus menerus untuk mendapatkan konvergensi pada skor kata. Keunggulan RAKE adalah efisiensi komputasinya, kecepatannya, presisinya meskipun sederhana dan kemampuannya untuk bekerja pada dokumen individual. (Thushara, Mownika dan Mangamuru, 2019)

2.2. 4 MAP@K (*Mean Average Precision@K*)

Mean Average Precision at K (MAP@K) adalah metrik yang digunakan untuk mengevaluasi kinerja sistem rekomendasi atau sistem pengambilan informasi. Ini adalah perluasan dari konsep *Average Precision* (AP) ke sekelompok item teratas (top-K).

Average Precision (AP): AP adalah metrik yang mempertimbangkan presisi suatu sistem pada tingkat recall yang berbeda. Presisi adalah fraksi item yang relevan di antara item yang direkomendasikan. Recall adalah fraksi item yang relevan yang telah diambil dibandingkan dengan jumlah total item yang relevan. AP kemudian adalah rata-rata nilai presisi pada tingkat recall yang berbeda, dengan mempertimbangkan item yang relevan.

Mean Average Precision at K (MAP@K): MAP@K memperluas AP ke sejumlah item teratas tertentu (top-K). Ini mengevaluasi presisi sistem rekomendasi dengan mempertimbangkan rata-rata presisi dari item teratas yang direkomendasikan. Rumus untuk MAP@K adalah:

$$Precision = \frac{| \{relevant\ documents\} \cap \{retrieved\ documents\} |}{| \{retrieved\ documents\} |} \quad (2)$$

Dalam konteks sistem rekomendasi atau pengambilan informasi, MAP@K memberikan evaluasi yang lebih realistis dengan mempertimbangkan skenario praktis di mana pengguna umumnya tertarik hanya pada rekomendasi top-K daripada seluruh daftar item. Ini membantu menilai efektivitas suatu sistem dalam menyediakan item yang relevan dalam rekomendasi top-K. Semakin tinggi nilai MAP@K, semakin baik sistem rekomendasi dianggap.

2.2. 5 SINTA (*Science and Technology Index*)

SINTA adalah metode penyampaian karya ilmu pengetahuan dan inovasi teknologi Indonesia, berbentuk sistem informasi penelitian berbasis *web* yang dimulai oleh Direktur Jenderal Penguatan Karya Inovatif, Kementerian Riset Teknologi dan Dikti Republik Indonesia pada tahun 2016. Portal ini mengevaluasi Institusi akademik, jurnal, dan peneliti di Indonesia. SINTA mengurutkan semua jurnal publik bersertifikat yang telah didistribusikan oleh ARJUNA (Akreditasi

Jurnal Nasional), lembaga yang ditunjuk untuk menyelesaikan evaluasi untuk konfirmasi kualitas jurnal ilmiah melalui kewajaran penyaringan komposisi, kesesuaian para eksekutif, dan kepraktisan distribusi jurnal ilmiah, menjadi enam klasifikasi, terdiri dari SINTA 1 s/d SINTA 6.(Saputra, 2020)

Sistem identifikasi bidang kepakaran program studi informatika ini akan mengambil data pada SINTA (*Science and Technology Index*) karena SINTA memberikan akses masuk ke referensi dan keahlian di Indonesia. Sebuah *website* untuk mengukur kinerja dosen, analis, perguruan tinggi, program, studi dan jurnal di Indonesia. SINTA memberikan standar dan analisis, membedakan kualitas penelitian setiap perguruan tinggi dan program studi untuk membina kerjasama yang kooperatif, menyelidiki pola penelitian dan direktori ahli. Selain itu, SINTA merupakan portal *indexing* jurnal yang saat ini diawasi oleh Dinas Pendidikan dan Kebudayaan Republik Indonesia. Selain itu, ada juga yang menegaskan bahwa SINTA merupakan *database* jurnal nasional yang diakreditasi Arjuna..(Wahyudi dan Berlilana, 2022)

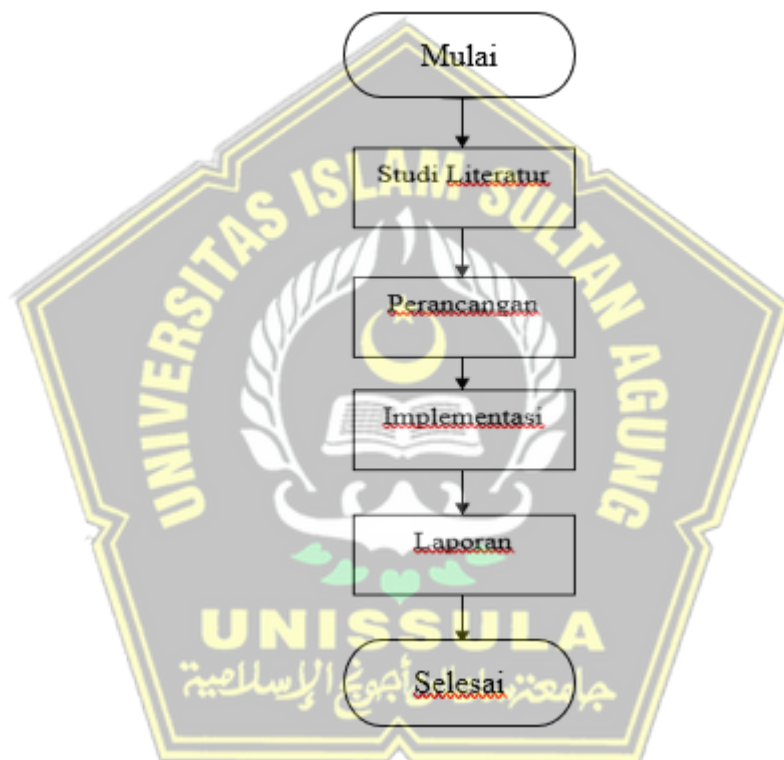
SINTA perlu mengadopsi metode pengukuran yang dinormalisasi. Normalisasi memperhitungkan parameter berikut: jumlah rata-rata kutipan per publikasi (tidak termasuk kutipan sendiri); persentase publikasi tanpa kutipan, jumlah rata-rata kutipan jurnal, dan kinerja unit penelitian di bidang terkait di seluruh dunia. Pengukuran kinerja penelitian yang akan diterapkan dalam SINTA harus mempertimbangkan bidang interdisipliner dan transdisipliner, karena beberapa ukuran mungkin tidak memadai untuk bidang tersebut.(Ibnu Subroto dan Chaerul Haviana, 2018)

BAB III

METODE PENELITIAN

3.1 Deskripsi Penelitian

Peneliti akan membahas proses identifikasi bidang kepakaran program studi menggunakan metode RAKE (*rapid automatic keyword extraction*) , kemudian akan divisualisasikan dengan grafik berbasis *website*.



Gambar 3. 1 Alur Penelitian

Pada gambar 3.1 merupakan alur dari penelitian yang dilakukan diantara lain:

1. Studi Literatur

Peneliti mengambil atau memperoleh data dari sumber-sumber seperti jurnal, artikel, atau dari hasil penelitian terdahulu, serta dari berbagai situs *website* untuk referensi dalam pengerjaan proses penelitian.

2. Perancangan

Untuk perancangan ini sistem akan dikembangkan dengan menggunakan metode *prototype* dimana sistem akan dievaluasi oleh *user* apakah sistem bekerja dengan baik atau tidak.

3. Implementasi

Sistem akan dibuat berbasis *web* dengan mengimplementasikan bahasa pemrograman *python*

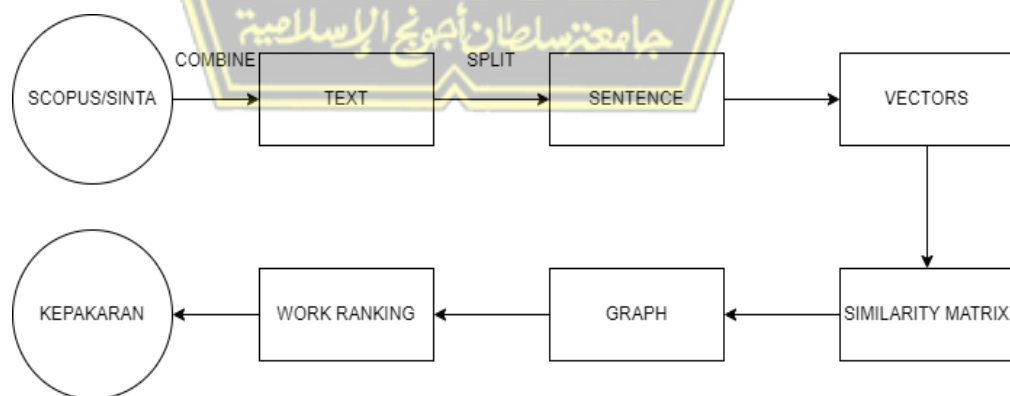
4. Laporan

Setelah penelitian selesai kemudian dibuat laporan untuk dokumentasi apa saja latar belakang, tujuan, manfaat, dan hasil dari penelitian yang sudah dibuat.

Data-data masukan diambil melalui studi literatur pada jurnal-jurnal *website* SINTA Kemendikbud dan Scopus yang dikumpulkan di *database* pada *mysql*. Data kemudian di olah dan dianalisis menggunakan metode RAKE di dalam sistem. Hasil yang diharapkan berupa grafik yang menampilkan kepakaran dari suatu program studi dari universitas di Indonesia yang disusun dari tinggi ke rendah yang berarti apabila grafik bar tersebut tinggi atau penuh, maka tingkat kepakaran program studi informatika tersebut sangat tinggi, begitu juga sebaliknya jika grafik bar kosong atau tidak penuh, maka tingkat kepakaran bidang studi tersebut rendah.

3.2 Metode Ekstraksi Kepakaran

3.2.1 Algoritma RAKE



Gambar 3. 2 Alur Proses Ekstraksi RAKE

Pada gambar 3.2 merupakan alur proses dari ekstraksi RAKE diantara lain:

- Data yang digunakan untuk ekstraksi RAKE bersumber dari SINTA yang terindeks scopus yang berjumlah 12.846 dengan data afiliasi, artikel scopus, *author*, dan *prodi*.
- Dari data yang diperoleh tersebut kemudian digabungkan (*combine*) menjadi beberapa teks
- Tahapan berikutnya teks yang sudah dikumpulkan kemudian di pisah(*split*) menjadi beberapa kalimat(*sentence*)
- Kemudian dari beberapa kalimat tersebut dicari beberapa sisipan kata setiap kalimat
- Persamaan antara sisipan kata(*similarity matrix*) akan dihitung dan disimpan didalam matriks
- Kata yang sering muncul bersamaan lebih dari satu kali kemudian akan dihitung dan hasilnya akan di rubah(*converted*) menjadi sebuah grafik
- Terakhir adalah hasil berupa skor kepakaran dari program studi tersebut

Dalam proses ekstraksi kata kunci RAKE menggunakan *stoplist* untuk mendapatkan *list* dari *candidate keywords* dari sebuah dokumen. Kemudian skor dari *candidate keywords* dihitung menggunakan perankingan berbasis graf.

Contoh teks:

“Rendang merupakan hidangan yang begitu lezat dan kaya rempah, adalah contoh gemilang dari kekayaan kuliner Indonesia.”

1. *Text Preprocessing*

Teks diolah dengan cara *tokenizing* dan *case folding* pada tabel 3.1

Tabel 3. 1 *Text Preprocessing*

Setelah <i>Tokenizing</i>
‘Rendang’ ‘merupakan’ ‘hidangan’ ‘yang’ ‘begitu’ ‘lezat’ ‘dan’ ‘kaya’ ‘rempah’, ‘adalah’ ‘contoh’ ‘gemilang’ ‘dari’ ‘kekayaan’ ‘kuliner’ ‘Indonesia’.
Setelah <i>Case Folding</i>

‘rendang’ ‘merupakan’ ‘hidangan’ ‘yang’ ‘begitu’ ‘lezat’ ‘dan’
 ‘kaya’ ‘rempah’, ‘adalah’ ‘contoh’ ‘gemilang’ ‘dari’ ‘kekayaan’
 ‘kuliner’ ‘indonesia’.

2. *Delimiters*

Hasil *preprocessing* diatas dibagi menjadi urutan kata-kata yang dipisahkan dengan *delimiters* (tanda seperti koma (,), titik (.), dsb) dan *stopword* sehingga didapatkan *candidate keywords* seperti pada tabel 3.2

Tabel 3. 2 Tabel *Delimiters*

No	Default Sentence	Split by Delimiters & Stopwords	Candidate Keywords
1	Rendang	Rendang	Hidangan
2	Merupakan		Rendang
3	Hidangan	Hidangan	
4	Yang		
5	Begitu		
6	Lezat	Lezat	Lezat
7	Dan		
8	Kaya	Kaya	Kaya Rempah
9	Rempah	Rempah	
10	Adalah		
11	Contoh	Contoh	Contoh
12	Gemilang	Gemilang	Gemilang
13	Dari		
14	Kekayaan	Kekayaan	Kekayaan
15	Kuliner	Kuliner	Kuliner
16	Indonesia	Indonesia	Indonesia

Pada tabel 3.2 merupakan tabel *delimiters* yaitu kata-kata yang dipisahkan dengan tanda seperti koma(,), titik(.), dsb.(*delimiters*) bagi kata kunci yang dicoret

merupakan *stopwords*. *Stopword* merupakan kata yang dianggap tidak informatif dan tidak mengandung makna jadi tidak diperlukan atau dihapus.

3. Menghitung *Score* Kata Kunci

Perhitungan algoritma RAKE mengambil rasio derajat frekuensi dari kata-kata. Tahap pertama untuk menghitung derajat frekuensi adalah dengan membuat *co-occurrence graph* (graf kemunculan bersama). Berikut *co-occurrence graph* dari kata kunci yang sudah didapatkan.

Tabel 3. 3 Perhitungan *Score* Kata Kunci

	Renda ng	Hidan gan	Lezat	Kaya	Rem pah	Cont oh	Gemil ang	Kekay aan	Kulin er	Indon esia
Renda ng	1	1	0	0	0	0	0	0	0	0
Hidan gan	1	1	0	0	0	0	0	0	0	0
Lezat	0	0	1	0	0	0	0	0	0	0
Kaya	0	0	0	1	1	0	0	0	0	0
Rempah	0	0	0	1	1	0	0	0	0	0
Conto h	0	0	0	0	0	1	1	0	0	0
Gemil ang	0	0	0	0	0	1	1	0	0	0
Kekay aan	0	0	0	0	0	0	0	1	1	1
Kulin er	0	0	0	0	0	0	0	1	1	1
Indon esia	0	0	0	0	0	0	0	1	1	1

Pada tabel 3.3 , kita coba sampel dari kata kunci “kaya”, “rempah” bernilai satu karena “kaya” dan “rempah” muncul bersama sebanyak satu kali diantara semua kata kunci

Selanjutnya untuk menghitung *score*, dapat menghitung dua hal terlebih dahulu yaitu:

a. *Word Frequency* ($\text{freq}(w)$)

Perhitungan *word frequency* merupakan perhitungan dari berapa kali kata tertentu muncul di antara semua kata kunci yang ada. Sebagai contoh “rempah”, maka $\text{freq}(\text{rempah}) = 1$

b. *Word Degree* ($\text{deg}(w)$)

Perhitungan *word degree* merupakan jumlah nilai dari baris suatu kata tertentu pada *co-occurrence graph*. Sebagai contoh “rempah”, maka $\text{deg}(\text{rempah}) = 1+1 = 2$ (nilai dari baris “rempah” pada tabel *co-occurrence graph*)

Hasil perhitungan $\text{deg}(w)$ dan $\text{freq}(w)$ masing-masing *candidate keyword* adalah pada tabel 3.4

Tabel 3. 4 Hasil Perhitungan $\text{deg}(w)$ dan $\text{freq}(w)$

	Degree of word	Word Frequency	Score
Rendang	2	1	2
Hidangan	2	1	2
Lezat	1	1	1
Kaya	2	1	2
Rempah	2	1	2
Contoh	2	1	2
Gemilang	2	1	2
Kekayaan	3	1	3
Kuliner	3	1	3
Indonesia	3	1	3

Dari tabel 3.4 dapat diketahui *score* dari tiap kata kunci. Dimana ‘Kekayaan’ ‘Kuliner’ ‘Indonesia’ memiliki degree skor tertinggi karena presentase kemunculan bersama pada baris kata yaitu 3.

Tabel 3. 5 *Expertise word*

Kekayaan Kuliner Indonesia	Kaya Rempah	Hidangan Rendang	Contoh Gemilang	Lezat
9	4	4	4	1

Dari tabel 3.5 merupakan hasil dari perhitungan dimana kata ‘Kekayaan kuliner Indonesia’ memiliki skor kepakaran tertinggi yaitu $3+3+3 = 9$

4. *Source code* untuk RAKE

Untuk dapat menggunakan perhitungan RAKE, perlu dilakukan instalasi terlebih dahulu yaitu dengan *source code*

```
Pip install rake-nltk
```

Merupakan *source code* untuk penginstalasi RAKE, RAKE bisa diinstall pada *jupyter notebook* maupun menggunakan *command prompt*, apabila sudah menginstall maka secara otomatis dapat menggunakan perhitungan menggunakan metode RAKE menggunakan *source code*

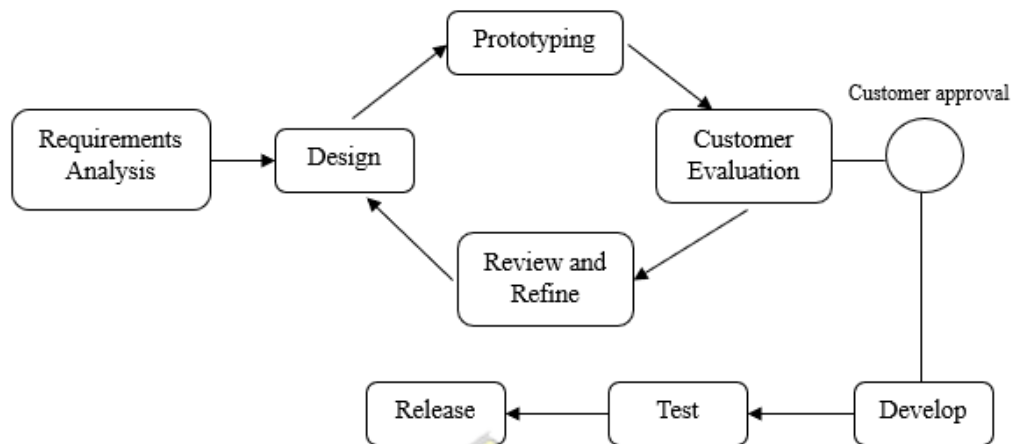
```
1. rake = Rake(max_length=3)
2. rake.extract_keywords_from_text(text)
3. RAKE_keywords = rake.get_ranked_phrases_with_scores()
```

Penjelasan *source code* merupakan penggunaan *library* RAKE yang dimana:

- 1) `RAKE = RAKE(max_length=3)`: kode ini menginisialisasi objek RAKE dari kelas RAKE. Kode argumen `max_length=3` untuk membatasi panjang maksimum frasa yang akan diekstraksi sebagai kata kunci. Ini berarti hanya frasa hingga 3 kata yang akan diperhitungkan sebagai kata kunci.
- 2) `RAKE.extract_keywords_from_text(text)`: Baris ini mengambil teks yang ingin diproses dan mengekstraksi kata kunci dari teks tersebut menggunakan objek RAKE yang telah diinisialisasi sebelumnya.
- 3) `RAKE_keywords = RAKE.get_ranked_phrases_with_scores()`: Setelah ekstraksi selesai, baris ini akan menghasilkan daftar frasa-frasa yang diidentifikasi sebagai kata kunci beserta skor relevansinya. Metode `get_ranked_phrases_with_scores()` mengembalikan daftar *tuple*, di mana setiap *tuple* berisi skor relevansi dan frasa kata kunci yang terkait.

3.3 Metode Pengembangan Sistem

Pengembangan sistem penelitian menggunakan metode *prototype*. Didalam model *prototype* terdapat langkah-langkah pada gambar 3.3



Gambar 3. 3 Model *Prototype*

Pada gambar 3. 3 merupakan model *prototype* yang digunakan untuk metode pengembangan sistem. *Prototype* dipilih karena dengan menggunakan *prototype*, pengguna dapat ikut andil dalam proses pengembangan dengan cara mengevaluasi dan memberikan umpan balik. Model *prototype* yang dipakai yaitu *extreme prototyping*, model ini merupakan salah satu pendekatan dalam pengembangan perangkat lunak yang fokus pada pengembangan *prototype* yang dapat digunakan untuk mendemonstrasikan dan menguji fungsi-fungsi utama dari sebuah aplikasi atau sistem secara sangat cepat. Pendekatan ini sering digunakan dalam pengembangan perangkat lunak berbasis *web*, tetapi juga dapat diterapkan pada proyek-proyek perangkat lunak lainnya. *Extreme prototype* dibagi menjadi tiga tahap yaitu

1. Membuat *prototype* dasar yang berisi semua halaman berformat HTML.
2. Lapisan *prototype service* digunakan untuk mensimulasikan pemrosesan data.
3. Layanan diimplementasikan dan diintegrasikan ke dalam *prototype* akhir.

3.3.1 *Requirements Analysis*

Peneliti akan meninjau beberapa artikel, jurnal, hasil penelitian terdahulu seperti tesis dan skripsi, serta berbagai situs *website* untuk referensi dalam mengerjakan proses penelitian. Hal ini dilakukan untuk mendapatkan data dan informasi yang

berhubungan dengan *data mining*, *RAKE*, *scopus*, bidang kepakaran program studi informatika, dan SINTA (*Science and Technology Index*).

Pada tahap awal perencanaan pengembangan sistem ini, peneliti melakukan *requirement analysis* atau analisa kebutuhan dengan melakukan diskusi kepada pihak pengguna atau *user*, dalam hal ini adalah Tim Pengembang SINTA. Tujuan komunikasi ini adalah untuk mendiskusikan dan menggambarkan analisa kebutuhan sistem dan rencana pengembangan sistem secara lebih rinci dan menyeluruh. Hal ini bertujuan untuk memastikan bahwa persepsi dan pemahaman mengenai analisa kebutuhan dan pengembangan sistem ini dapat diselaraskan di antara semua pihak yang terlibat. Hasil dari tahapan ini peneliti mendapatkan data dari tim pengembang SINTA berupa data judul publikasi dosen yang terindeks scopus. Pada tahap ini juga peneliti mendapatkan gambaran mengenai *system* yang akan dikembangkan yaitu sistem kepakaran program studi informatika di Indonesia berdasarkan publikasi pada SINTA terindeks scopus.

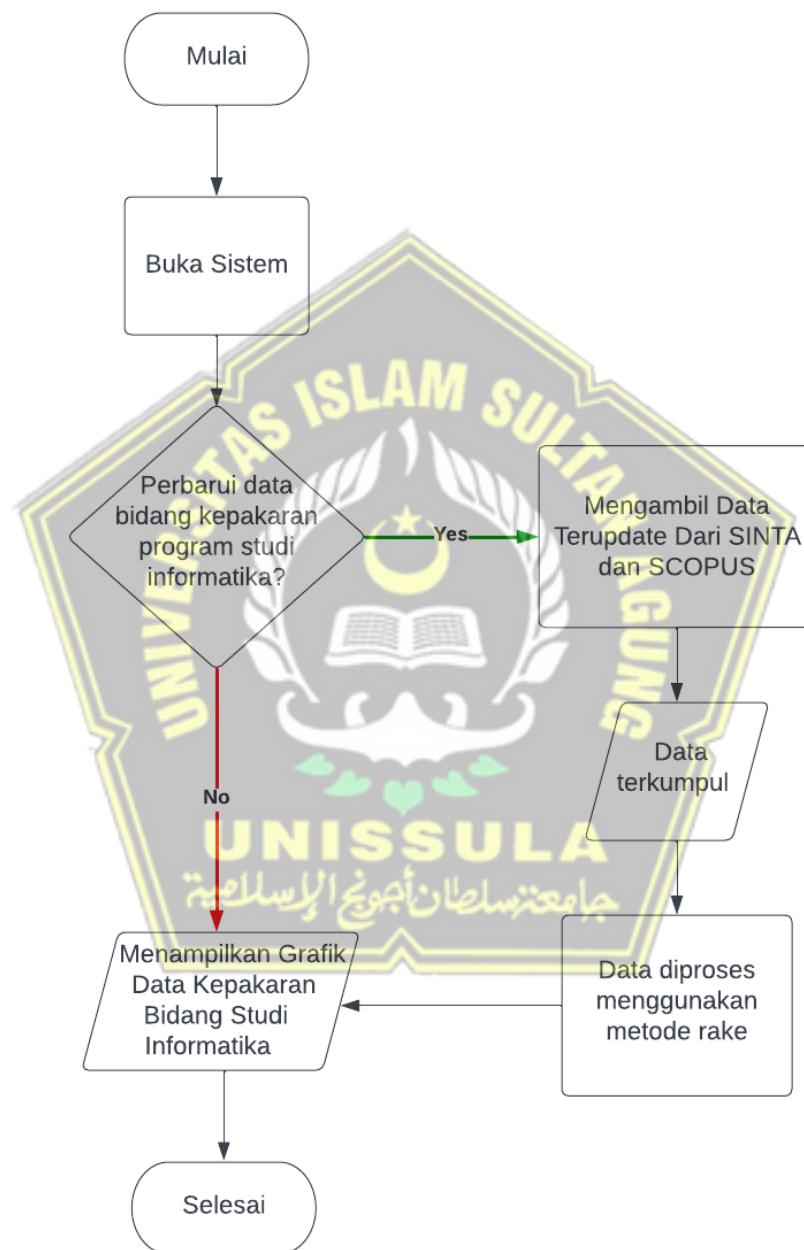
Pada tahap ini juga peneliti menganalisa kebutuhan sistem yaitu apa saja tools yang dibutuhkan untuk mengembangkan sistem ini sehingga proses *input* sampai dengan sistem berjalan sesuai dengan yang diinginkan. Adapun *tools* yang dibutuhkan antara lain meliputi *python* sebagai bahasa pemrograman untuk mengolah data yang didapatkan, *jupyter notebook* sebagai *platform* data diolah menjadi ekstraksi kata kunci, MySQL sebagai *platform* untuk data *warehousing* (gudang data), dan *streamlit* sebagai *platform* untuk membuat *website* sistem untuk menampilkan data yang sudah diekstraksi sehingga *user* dapat mengakses ketika *website* sudah siap.

3.3.2 Design

Pada tahap *design* ini, peneliti menggambarkan secara cepat dan kasar konsep dasar dari sistem kepakaran program studi informatika agar pengguna dapat gambaran awal dari sistem.

1. Flowchart Sistem

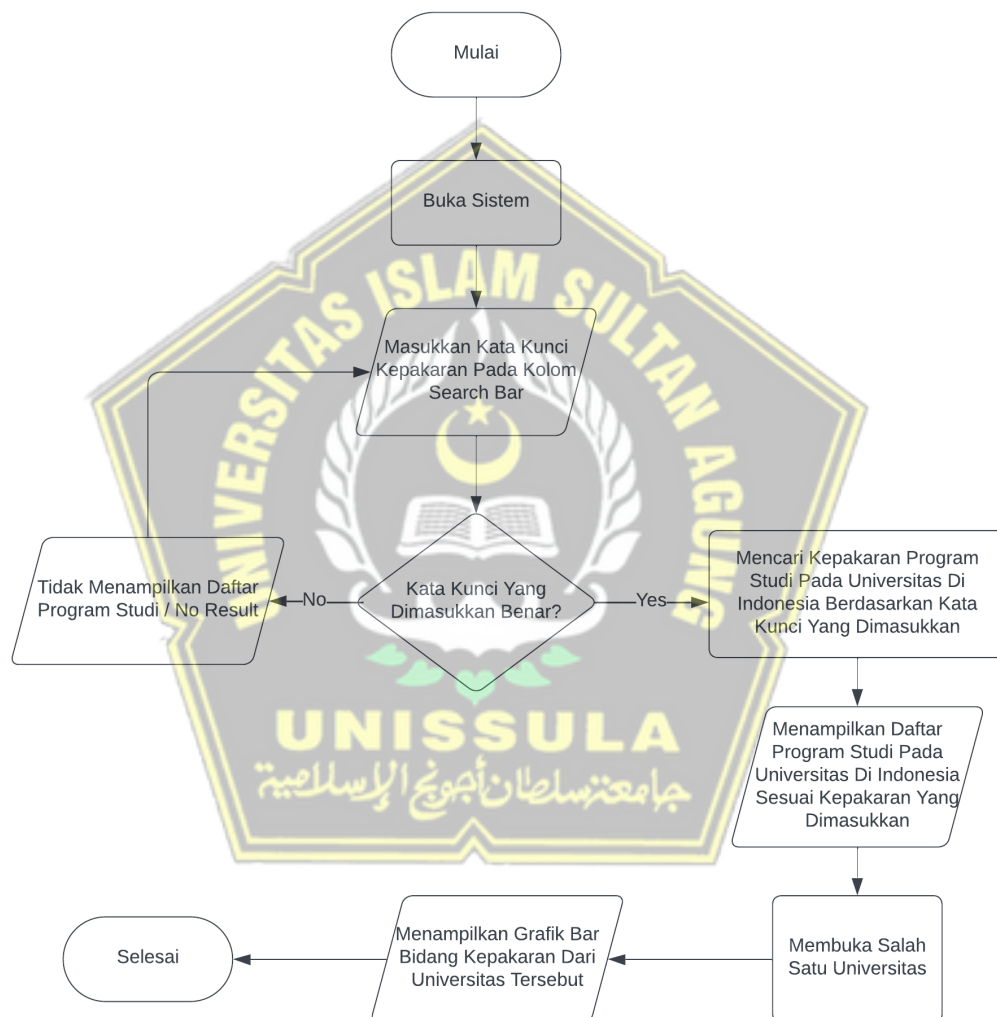
Flowchart sistem berguna untuk menggambarkan sebuah alur pengerjaan maupun proses suatu program, dalam *flowchart* dibawah ini terdapat dua buah *flowchart* yaitu *flowchart user* dan *flowchart admin*



Gambar 3. 4 *Flowchart* sistem untuk *admin*

Gambar 3.4 memperlihatkan *flowchart* sistem untuk *admin* dimana dimulai dari membuka sistem kemudian akan terdapat pilihan apakah ingin memperbarui data

dari bidang kepakaran atau tidak. Jika iya, maka sistem akan mengambil data yang *terupdate* dari SINTA yang terindeks scopus, setelah data terkumpul kemudian diproses menggunakan metode *RAKE* dengan hasil output berupa grafik. Jika tidak, maka hanya menampilkan grafik data kepakaran bidang studi informatika.



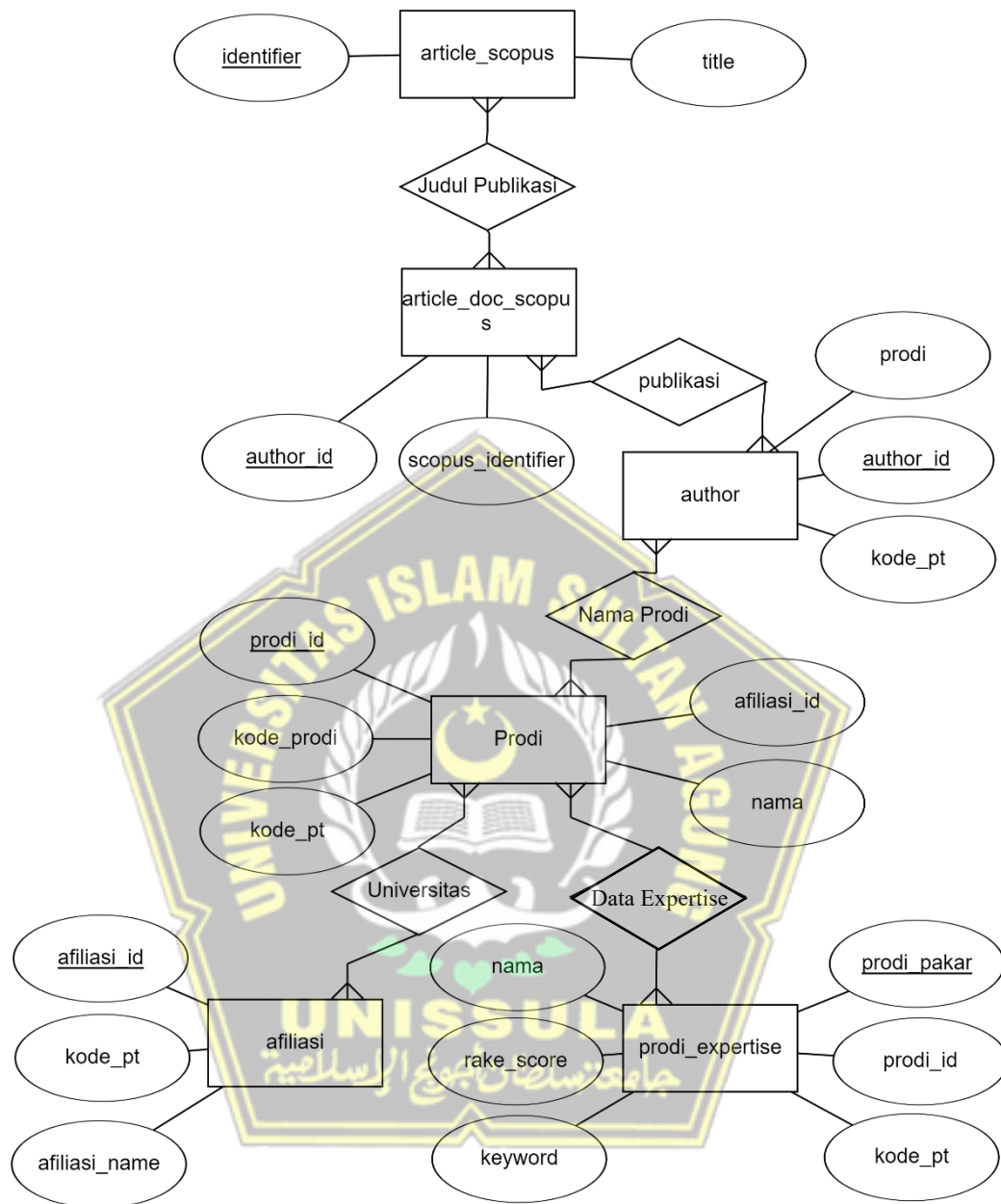
Gambar 3. 5 Flowchart sistem untuk user

Pada gambar 3.5 merupakan *flowchart* alur kerja dari sistem *user* yaitu mulai dengan membuka sistem, kemudian *user* memasukkan kata kunci kepakaran pada kolom *search bar*, apabila kata kunci yang dimasukkan tidak benar, maka tidak akan menampilkan daftar program studi atau pencarian *not result* yang kemudian

akan diminta untuk memasukkan kata kunci kembali pada *search bar*. Jika kata kunci yang dimasukkan benar, maka sistem akan mencari kepakaran program studi pada universitas di Indonesia berdasarkan kata kunci yang dimasukkan, selanjutnya akan ditampilkan daftar program studi pada universitas sesuai kata kunci kepakaran yang dimasukkan, selanjutnya *user* dapat memilih salah satu universitas yang kemudian akan menampilkan grafik bar bidang kepakaran dari universitas yang dipilih.

2. ERD

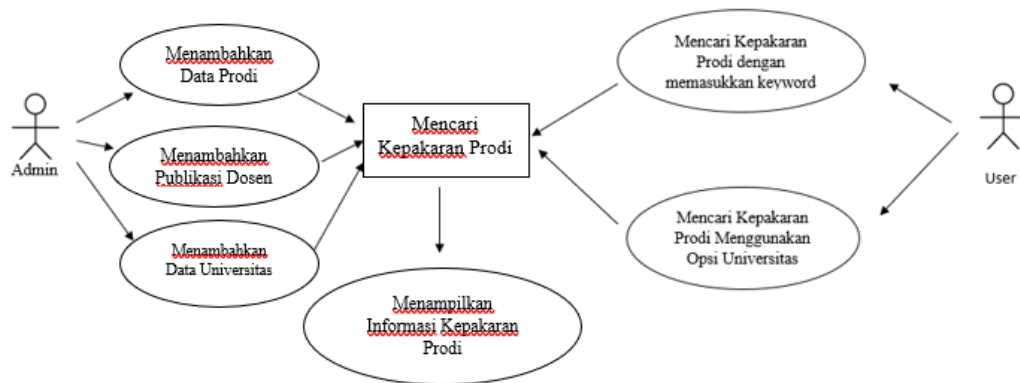




Gambar 3. 6 ERD model

Pada gambar 3.6 merupakan ERD dari penelitian yang akan dilakukan, jadi mengambil judul publikasi dari tabel artikel scopus berdasarkan publikasi dari author. Prodi mengambil data dari author dan dibuat tabel *prodi expertise* sebagai tabel penyimpanan data, dan data dari afiliasi digunakan untuk mengambil afiliasi nama agar dapat diinput bersama dengan data pada tabel *prodi expertise*

3. Use Case



Gambar 3. 7 Use Case sistem

Pada gambar 3.7 merupakan gambar dari gambar *use case* sistem dimana *admin* dapat menambahkan data prodi, menambahkan publikasi dari dosen, dan juga menambahkan data universitas tetapi penambahan ini dilakukan oleh *admin* SINTA karena peneliti mendapatkan data dari SINTA yang terindeks scopus. Kemudian diproses oleh sistem akan menghasilkan informasi kepakaran prodi tertentu. Sedangkan *user* dapat mencari kepakaran prodi dengan memasukkan kata kunci atau mencari kepakaran menggunakan opsi universitas. Kemudian sistem akan memproses dan menghasilkan data kepakaran prodi yang dicari *user*.

4. Struktur Tabel

Struktur tabel adalah kumpulan data terstruktur. Data terstruktur tersusun dari kolom dan baris yang tersimpan di suatu *database* atau basis data. Data dapat ditambah, diubah, dihapus, dan dilihat untuk menghasilkan informasi. Tabel 3. 5 merupakan struktur tabel pada penelitian ini.

Tabel 3. 6 Tabel *database* penelitian

Kolom	Tipe Data	<i>Null</i>	<i>Default</i>	<i>Extra</i>
Prodi_pakar	Bigint 20	<i>No</i>	<i>None</i>	<i>Auto_increment, Primary Key</i>
Prodi_id	Int 11	<i>Yes</i>	<i>Null</i>	
Kode_pt	Varchar(255)	<i>Yes</i>	<i>Null</i>	

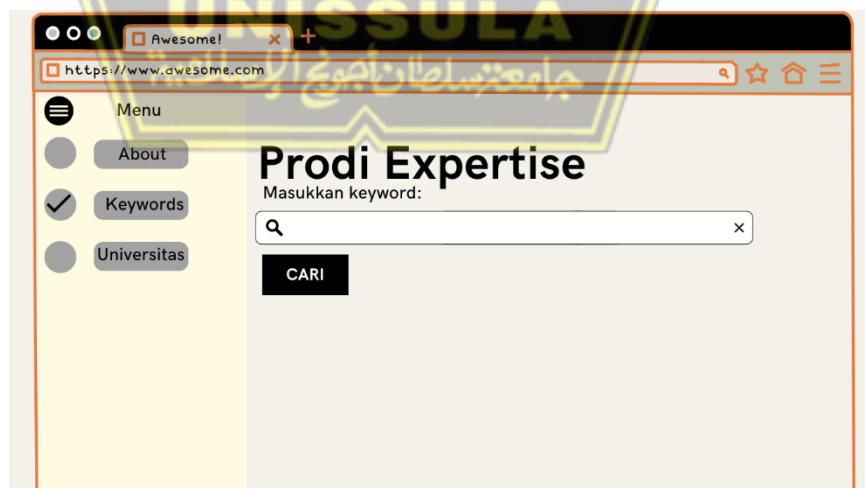
Kolom	Tipe Data	<i>Null</i>	<i>Default</i>	<i>Extra</i>
Nama	Varchar(255)	<i>Yes</i>	<i>Null</i>	
Keyword	Varchar(255)	<i>Yes</i>	<i>Null</i>	
RAKE_score	Float	<i>Yes</i>	<i>Null</i>	

5. Perancangan *User Interface*

Perancangan antarmuka merupakan tampilan awal dari *website* sebelum diterapkan pada bahasa pemrograman. Pada tahap ini peneliti merancang UI secara cepat sebagai gambaran awal sistem tetapi belum sempurna. Pada perancangan antarmuka meliputi tata letak komponen-komponen yang akan ditampilkan pada *website*.

a. Halaman *Expertise*

Halaman *expertise* pada perancangan antarmuka sistem ini menampilkan judul dari *website* yaitu *prodi expertise*, juga terdapat kolom *search bar* untuk memasukkan kata kunci untuk mencari data kepakaran program studi secara spesifik, pengguna dapat memasukkan kata kunci berupa kepakaran program studi maupun dari program studi tersebut.

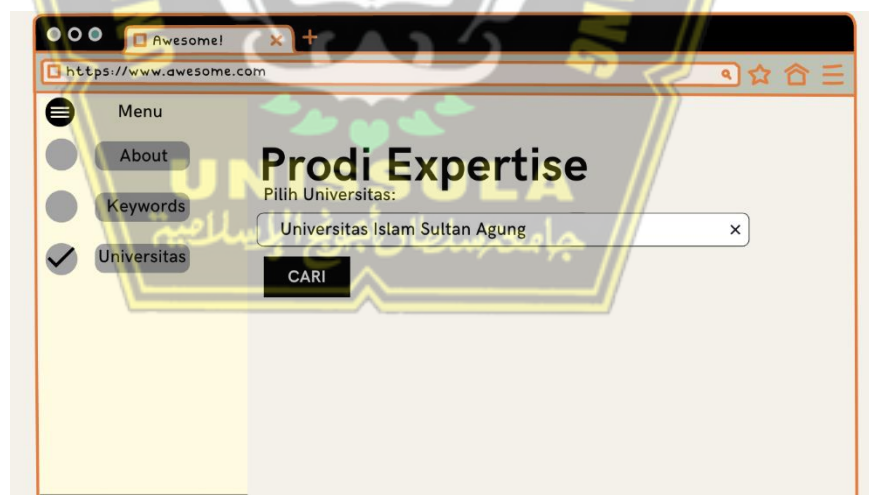


Gambar 3. 8 Menu *Keyword* perancangan antarmuka

Pada gambar 3.8 menampilkan perancangan antarmuka pada *website*. Komponen-komponen *website* antara lain sebagai berikut:

- a) *Header* yang merupakan judul dari sistem
- b) Halaman *search bar* yang berguna sebagai kolom pencarian pada *website* untuk mencari kata kunci.
- c) Tombol cari digunakan untuk memproses sistem agar dapat mencari kata kunci yang telah dimasukkan
- d) Pilihan menu untuk mengetahui fitur-fitur lain dari *website* terdapat *about* untuk penjelasan singkat tentang sistem, juga menu universitas untuk mengetahui kepakaran program studi dengan memilih opsi dari universitas
- b. Halaman Universitas

Pada halaman ini akan menampilkan kolom opsi pilihan berdasarkan universitas di Indonesia. Berikut perancangan antarmuka dari opsi pilihan universitas.

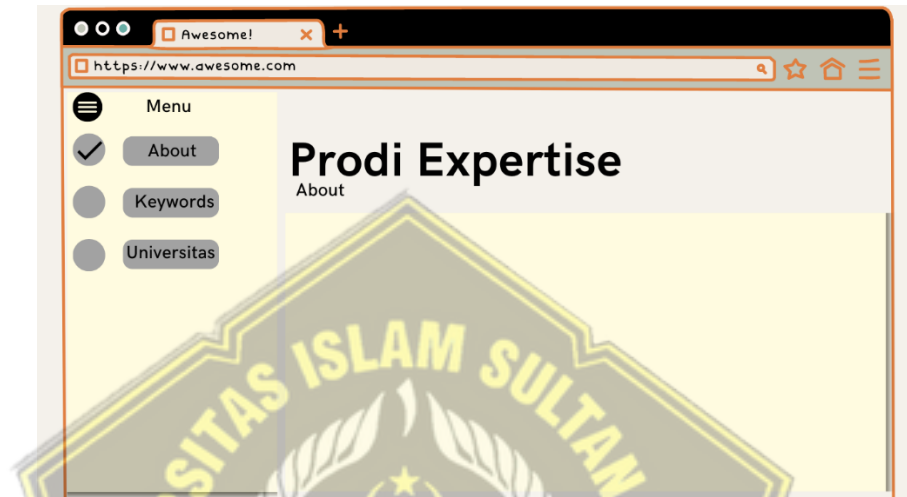


Gambar 3. 9 Perancangan antarmuka halaman universitas

Pada gambar 3.9 merupakan perancangan antarmuka pada opsi pilihan universitas dimana terdapat opsi pilihan untuk mengetahui kepakaran program studi pada universitas tersebut dengan klik cari pada universitas yang dipilih.

c. Halaman *About*

Pada halaman ini akan menampilkan informasi mengenai sistem kepakaran secara spesifik dan juga fitur-fitur yang ada pada *website* ini. Berikut merupakan gambar dari halaman *about*:



Gambar 3. 10 Perancangan Antarmuka halaman *About*

Pada gambar 3.10 merupakan perancangan antarmuka pada *about website* dimana disana akan dijelaskan secara singkat tentang kepakaran program studi dan tujuan utama *website* dibuat.

Dengan langkah-langkah ini diharapkan implementasi sistem dapat berjalan lancar sesuai dengan tujuan dan kebutuhan penelitian yang dilakukan.

3.3.3 Pembentukan *Prototyping*

Setelah analisis dan desain, maka dilakukan Pembentukan *Prototype* yang berguna untuk mendukung desain yang telah dirancang sebelumnya. Pembentukan *prototype* ini bertujuan memberikan gambaran lebih konkret tentang bagaimana sistem dibuat dan dioperasikan. Berikut adalah tahapan-tahapan dalam pembentukan *prototype*:

1. Pengumpulan Data

Setelah melakukan wawancara dengan tim pengembang SINTA, didapatkan data yang bersumber dari SINTA yang terindeks scopus, data yang diperoleh mencakup judul publikasi dosen dari 5 universitas yang ada di Indonesia.

2. Data yang sudah diperoleh kemudian diolah menggunakan bahasa pemrograman *python*. *Python* dipilih sesuai dengan kemampuan dalam melakukan analisis data.
3. Data diolah dengan mengimplementasikan metode RAKE(*Rapid Automatic Keyword Extraction*) untuk menentukan kata kunci dan skor dari jurnal publikasi dari *author* pada *jupyter notebook*
4. Setelah mendapatkan skor kepakaran dari ekstraksi kata kunci, kemudian data terbaru disimpan didalam *database* baru pada *MySQL*.
5. Data yang tersimpan pada *database* dijadikan dasar utama untuk ditampilkan pada *website* yang akan dibuat pada *visual studio code*
6. Pengembangan sistem yaitu dengan membuat sebuah sistem dan data kepakaran tersebut sebagai dasar bagi pengembangan sistem. Sistem yang dibuat berbasis *website* agar mudah diakses oleh pengguna. Pengembangan sistem akan menggunakan *platform streamlit*. Penggunaan *platform* ini didasarkan kemampuan dalam membuat *website* dan sangat praktis bagi *developer*.

3.3.4 Customer Evaluation

Pada tahap *customer evaluation* atau evaluasi *user* dalam hal ini tim pengembang SINTA. Rancangan sistem yang diusulkan atau dikembangkan disajikan kepada *user* untuk diminta memberikan *feedback* atau komentar mereka baik secara tertulis maupun melalui wawancara. Pada tahap ini bermanfaat untuk mengetahui kelebihan dan kelemahan *prototype* yang sudah dibuat pada tahap sebelumnya. Umpan balik dan saran dari *user* dikumpulkan dan diteruskan ke pengembang dalam hal ini peneliti. Tujuannya dari tahap ini untuk perbaikan lebih lanjut mengenai pengembangan sistem tersebut apabila *prototype* masih belum memenuhi kebutuhan *user* maka diproses lagi pada tahap *review* dan *refine*, apabila *prototype* sudah sesuai dengan kebutuhan *user* dan mendapatkan *approve*, maka sistem bisa langsung dapat dikembangkan.

3.3.5 *Review dan Refine*

Setelah mendapatkan *feedback* pada tahapan sebelumnya, maka dilakukan tahapan *review and refine*, Pada tahap ini dilakukan proses peninjauan ulang berdasarkan hasil evaluasi atau *feedback user*. Setelah itu, dilakukan tahap penyempurnaan untuk memperbaiki sistem tersebut, tahapan ini dapat melibatkan beberapa iterasi. Setiap iterasi menghasilkan versi *prototype* yang lebih matang dan lebih dekat dengan ekspektasi *user*.

Pada proses iterasi tersebut ketika *prototype* sudah sesuai dengan keinginan *user* dan mendapatkan *customer approval* atau persetujuan *user* untuk melanjutkan proses pengembangan sistem pada tahap selanjutnya.

3.3.6 *Develop*

Pada tahap ini peneliti mengembangkan sistem sesuai *prototype* yang sudah dibuat dan *feedback* dari *user* sebelumnya. Develop ini meliputi implementasi dimana kode sistem dan *website* dibuat berdasarkan *prototype* yang sudah ada sebelumnya.

3.3.7 *Rencana Pengujian Sistem*

Ide di balik rencana pengujian adalah untuk menguji fungsi-fungsi sistem yang sedang dibangun untuk melihat apakah berfungsi sesuai dengan yang diharapkan. Rencana pengujian yang dilakukan pada *framework* ini adalah uji kegunaan dimana pengujian kegunaan menggunakan strategi *black box*. Persyaratan fungsional sistem adalah fokus utama pengujian ini. Motivasi dibalik pengujian dengan menggunakan teknik penemuan adalah untuk menemukan kesalahan yang berguna dalam kerangka yang telah disusun. Selain itu, pengujian ini dilakukan dengan mencoba semua prospek yang terjadi dan dilakukan lebih dari satu kali. Selama pengujian, jika terjadi kesalahan, pencarian dan perbaikan akan dilakukan untuk memperbaiki kesalahan tersebut.

3.3.8 *Release*

Pada tahap ini, *prototype* tersebut dinyatakan sebagai versi yang siap untuk diperkenalkan atau dibagikan kepada *user* yang akan menguji atau mengevaluasi *prototype* tersebut.



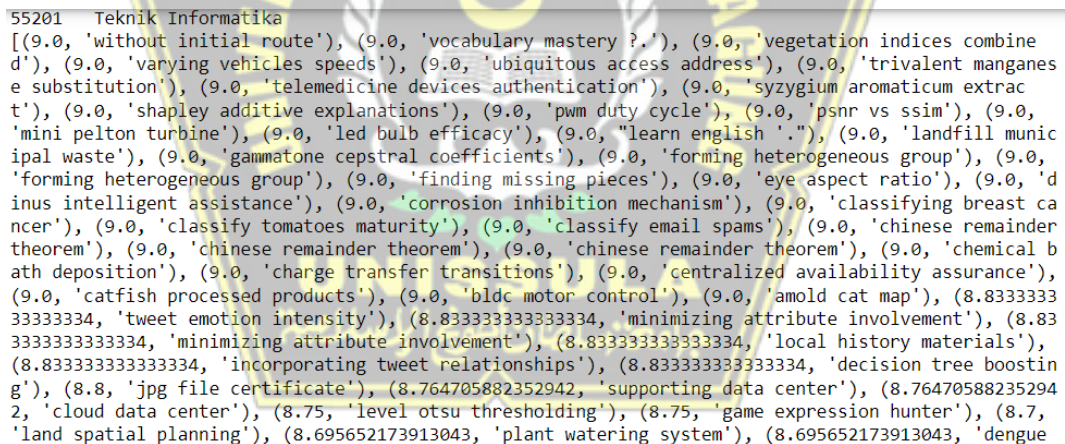
BAB IV

HASIL DAN ANALISIS PENELITIAN

4.1 Analisa Hasil

Analisis hasil adalah proses menganalisis dan mengevaluasi data yang diperoleh dari hasil penelitian. Proses ini bertujuan untuk menentukan apakah hasil penelitian tersebut memenuhi tujuan atau tidak. Hasil dari penelitian pengidentifikasi bidang kepakaran program studi informatika di Indonesia menggunakan RAKE telah memenuhi akhir dengan jumlah data yang berhasil dikumpulkan yaitu sebanyak 2716 dengan mengambil 5 universitas di Indonesia.

Berikut merupakan contoh data yang sudah diolah menggunakan RAKE dan python dan hasilnya seperti gambar 4.1



```
55201 Teknik Informatika
[(9.0, 'without initial route'), (9.0, 'vocabulary mastery ?.'), (9.0, 'vegetation indices combine
d'), (9.0, 'varying vehicles speeds'), (9.0, 'ubiquitous access address'), (9.0, 'trivalent manganes
e substitution'), (9.0, 'telemedicine devices authentication'), (9.0, 'syzygium aromaticum extrac
t'), (9.0, 'shapley additive explanations'), (9.0, 'pwm duty cycle'), (9.0, 'psnr vs ssim'), (9.0,
'mini pelton turbine'), (9.0, 'led bulb efficacy'), (9.0, "learn english '.'"), (9.0, 'landfill munic
ipal waste'), (9.0, 'gammatone cepstral coefficients'), (9.0, 'forming heterogeneous group'), (9.0,
'forming heterogeneous group'), (9.0, 'finding missing pieces'), (9.0, 'eye aspect ratio'), (9.0, 'd
inus intelligent assistance'), (9.0, 'corrosion inhibition mechanism'), (9.0, 'classifying breast ca
ncer'), (9.0, 'classify tomatoes maturity'), (9.0, 'classify email spams'), (9.0, 'chinese remainder
theorem'), (9.0, 'chinese remainder theorem'), (9.0, 'chinese remainder theorem'), (9.0, 'chemical b
ath deposition'), (9.0, 'charge transfer transitions'), (9.0, 'centralized availability assurance'),
(9.0, 'catfish processed products'), (9.0, 'bldc motor control'), (9.0, 'amold cat map'), (8.8333333
33333334, 'tweet emotion intensity'), (8.833333333333334, 'minimizing attribute involvement'), (8.83
3333333333334, 'minimizing attribute involvement'), (8.833333333333334, 'local history materials'),
(8.833333333333334, 'incorporating tweet relationships'), (8.833333333333334, 'decision tree boostin
g'), (8.8, 'jpg file certificate'), (8.764705882352942, 'supporting data center'), (8.76470588235294
2, 'cloud data center'), (8.75, 'level otsu thresholding'), (8.75, 'game expression hunter'), (8.7,
'land spatial planning'), (8.695652173913043, 'plant watering system'), (8.695652173913043, 'dengue
```

Gambar 4. 1 Data yang sudah diolah dengan RAKE

Pada gambar 4.1 merupakan data program studi informatika unissula yang sudah diolah dengan menggunakan metode RAKE dan *python* dan diperoleh skor kepakaran beserta kata kunci

Anda dapat memilih opsi universitas yang tersedia:

Universitas Islam Sultan Agung



Cari

Profil Universitas Islam Sultan Agung

Program Studi Dengan Rake Score Terbesar:

Keyword	Donut Chart	Keterangan
various similarity measures		● Expertise ● None
Keyword	Donut Chart	Keterangan
support vectors machine		● Expertise ● None

Gambar 4. 2 Pengimplementasi hasil ekstraksi ke website

Pada gambar 4. 2 merupakan pengimplementasian hasil ekstraksi ke *website* dalam hal ini diambil ekstraksi kata kunci dari program studi unissula, dimana divisualisasikan dengan donat chart dimana skor menunjukkan 100% menandakan kata kunci tersebut kepakarannya sangat tinggi.

Data yang dikumpulkan berupa prodi id, kode pt, nama program studi, kata kunci yang dihitung, dan skor perhitungan dengan RAKE seperti pada tabel berikut diambil data 10 sampel.

Tabel 4. 1 Data terkumpul

Prodi_id	Kode_pt	Nama	Keyword	RAKE_Score
55201	061002	Teknik Informatika	<i>Average image hash</i>	8.5
55201	061001	Teknik Informatika	<i>without initial route</i>	9

Prodi_id	Kode_pt	Nama	Keyword	RAKE_Score
55201	061001	Teknik Informatika	<i>financial data integration</i>	8.85714
55201	061038	Informatika	<i>teaching using moodle</i>	7.4
55201	061038	Informatika	<i>transition metal ions</i>	7.66667
55201	061038	Informatika	<i>transaction patterns</i>	4
55201	061002	Teknik Informatika	<i>short informal texts</i>	9
55201	061031	Teknik Informatika	<i>methyl butanoate adsorption</i>	8.2
55201	061031	Teknik Informatika	<i>authenticity</i>	1
55201	061031	Teknik Informatika (Kampus Kota Kediri)	<i>ubiquitous access address</i>	9.0

Pada tabel 4.1 merupakan data yang sudah berhasil dikumpulkan dan dihitung dengan menggunakan metode RAKE. Analisis hasil menunjukkan pada kolom kode pt, nama, *keyword*, dan skor sesuai dengan universitas. Skor kepakaran terbesar yaitu dengan nilai 9.0 dan skor kepakaran terkecil dengan nilai 1.0. kata kunci yang diekstraksi diambil satu sampai tiga kata setiap judul publikasi scopus.

4.2 Implementasi Sistem

Implementasi antarmuka merupakan tahap pemaparan tampilan sistem dan kegunaan fungsi dari *form* yang tersedia. Pada tahap ini, implementasi hasil rancangan dan analisa telah menghasilkan sebuah *website* identifikasi bidang

kepakaran program studi informatika dengan bahasa pemrograman *python* dan *mysql*.

Pencarian Expertise

Masukkan keyword atau program studi yang ingin dicari:

data

Cari

Anda mencari keyword: data

Hasil pencarian:

Universitas	Kode pt	Nama	Keyword	Diagram Donat	Keterangan
Universitas Kristen Satya Wacana	061001	Teknik Informatika	data gathering architecture		● Expertise ● None

Gambar 4. 3 Menu *keywords website*

Pada gambar 4.3 merupakan tampilan menu *keywords* dari *website* identifikasi kepakaran, dimana ada kolom *search bar* untuk memasukkan kata kunci yang ingin dicari, *user* juga dapat memasukkan nama program studi pada *search bar* tersebut. Terdapat tombol cari untuk melakukan pencarian kata kunci yang sudah dimasukkan. Hasilnya berupa tabel informasi mengenai universitas, program studi dan skor kepakaran yang divisualisasikan dengan diagram donat chart yang dimana jika skor kepakaran tinggi, maka akan penuh dan mencapai 100%.

Selain dengan kolom *search bar*, terdapat juga sistem pencarian dengan pilihan opsi dari universitas, jadi *user* memilih terlebih dahulu universitas apa yang ingin dicari dengan mengklik pilihan opsi universitas kemudian disana juga terdapat opsi pilihan program studi beserta kata kuncinya, jadi *user* dapat memilih universitas apa dan program studi mana yang akan dicari kepakarannya.

Anda dapat memilih opsi universitas yang tersedia:

Universitas Islam Sultan Agung



Cari

Profil Universitas Islam Sultan Agung

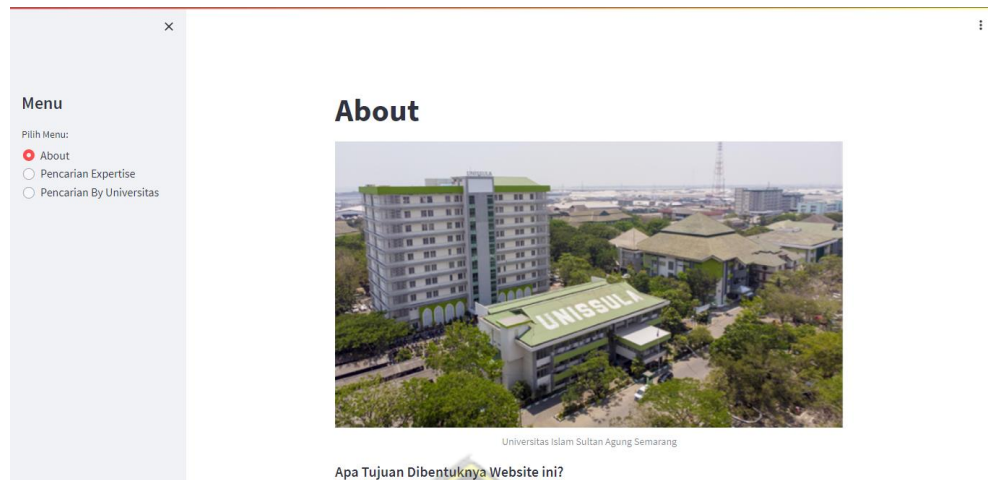
Program Studi Dengan Rake Score Terbesar:

Keyword	Donut Chart	Keterangan
various similarity measures		● Expertise ● None
Keyword	Donut Chart	Keterangan
support vectors machine		● Expertise ● None

Gambar 4. 4 Menu opsi universitas

Pada gambar 4.4 merupakan menu opsi pencarian kepakaran berdasarkan opsi universitas, jadi terdapat menu memilih universitas dan juga opsi program studi untuk dicari kepakarannya dengan klik tombol cari sebagai *button*nya. Hasil dari pencarian ini juga berupa tabel informasi mengenai skor kepakaran pada program studi tersebut dengan divisualisasikan dengan diagram donat *chart*

Pada *website* juga terdapat menu *about* sebagai sistem informasi kepada *user* mengenai kegunaan *website* dan tujuan dibuatnya *website* seperti gambar 4.5



Gambar 4. 5 Halaman menu *about* pada *website*

Pada gambar 4.5 merupakan tampilan menu *about* dimana bertujuan untuk memberikan informasi tentang cara kerja sistem dan juga tujuan pembuatan sistem.

4.3 Pengujian Sistem

Metode pengujian sistem yang digunakan dalam penelitian ini adalah pengujian *blackbox*. Pengujian difokuskan pada *input* yang diberikan dan dihasilkan oleh perangkat lunak untuk memastikan apakah berjalan sesuai yang diharapkan. Proses uji coba akan dilakukan dengan mengacu pada kriteria-kriteria sebagaimana tertera dalam tabel 4.2.

Tabel 4. 2 Tabel Hasil Pengujian

No.	<i>Input</i>	Hasil yang diharapkan	<i>Output</i>	Hasil
1.	Mencari kepakaran berdasarkan kata kunci kepakaran program studi	Menampilkan skor kepakaran berdasarkan kata kunci beserta keterangan universitas	Menampilkan berupa tabel hasil beserta skor kepakaran berdasarkan kata kunci yang dicari beserta nama universitas	Sesuai

No.	<i>Input</i>	Hasil yang diharapkan	<i>Output</i>	Hasil
2.	Mencari kepakaran berdasarkan nama universitas	Menampilkan pilihan universitas dan hasil skor kepakaran prodi berdasarkan universitas yang dipilih	Menampilkan opsi universitas dan menampilkan profil universitas yang dicari beserta tabel kepakaran dari universitas tersebut	Sesuai
3.	Memilih Menu <i>About</i>	Menampilkan halaman <i>about</i>	Menampilkan menu <i>about</i> yang berisi informasi mengenai sistem kepakaran program studi	Sesuai
4.	Memilih Menu Pencarian <i>Expertise</i>	Menampilkan Halaman pencarian kepakaran berdasarkan kata kunci	Menampilkan halaman pencarian <i>expertise</i>	Sesuai
5.	Memilih Menu Pencarian <i>By Universitas</i>	Menampilkan Halaman pencarian kepakaran berdasarkan universitas	Menampilkan halaman pencarian kepakaran berdasarkan universitas beserta profil universitas yang dipilih	Sesuai

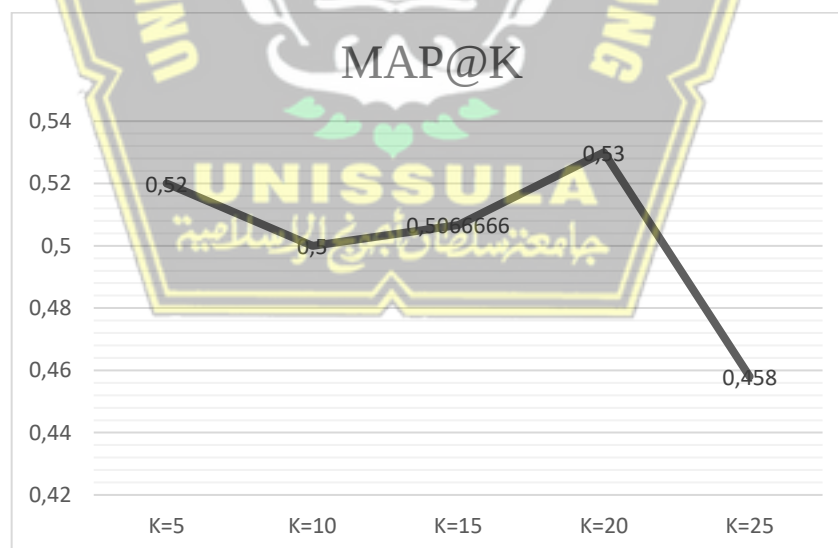
Pada tabel 4.2 merupakan tabel hasil dari penelitian menggunakan metode *blackbox testing* dimana dari tabel tersebut diuji apakah sistem bekerja sesuai apa yang diperintahkan.

Untuk menentukan keakuratan kata kunci yang diuji, telah dihitung dengan metode *Mean Average Precision@K* (MAP@K) dengan hasil seperti pada tabel dengan hasil seperti pada tabel 4.3

Tabel 4. 3 *Mean Average Precision at K*

	Unissula	Upgris	UKSW	UNIKA	Udinus	Rata-rata
K=5	0,4	0,6	0,2	0,8	0,6	0,52
K=10	0,4	0,6	0,4	0,6	0,5	0,5
K=15	0,333333	0,733333	0,466667	0,6	0,4	0,506667
K=20*	0,45	0,7	0,55	0,5	0,45	0,53
K=25	0,4	0,56	0,56	0,52	0,25	0,458

Pada tabel 4.3 merupakan tabel hasil dari *Mean Average Precision@K* (MAP@K) untuk menentukan kerelevan kata kunci ketika menggunakan metode RAKE.



Gambar 4. 6 MAP@K

Gambar 4.6 merupakan gambar dari perhitungan *Mean Average Precision@K* (MAP@K) dimana nilai tertinggi yang mendekati satu yaitu K=20 dengan nilai 0,53.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa sistem kepakaran program studi di Universitas di Indonesia dapat berjalan dengan baik tanpa adanya kesalahan dan dapat menghasilkan informasi mengenai kepakaran program studi secara spesifik sesuai dengan publikasi dosen pada SINTA yang terindeks scopus. Dalam pengembangan sistem ini, metode yang digunakan adalah dengan mengimplementasikan metode RAKE. Metode RAKE telah berhasil diimplementasikan untuk menghitung kata kunci secara otomatis. Kata kunci yang berhasil dihitung berjumlah 2716 dari 5 universitas di Indonesia. Setiap program studi akan menghasilkan kata kunci dengan variasi 1 hingga 3 kata kunci yang dihasilkan, dan jumlah kata kunci maksimal dalam penelitian ini adalah 20 kata kunci setiap program studi sesuai dengan perhitungan menggunakan metode *mean average precision@k* (MAP@K). Sehingga, setiap program studi akan memiliki jumlah kata kunci yang berbeda-beda sesuai dengan isi publikasi dan hasil ekstraksi metode RAKE pada publikasi dosen tersebut. *Website* berhasil memunculkan sistem informasi kepakaran program studi ketika *user* memasukkan kata kunci ataupun nama program studi atau informatika. *Website* berhasil menampilkan sistem informasi kepakaran program studi berdasarkan universitas yang dipilih.

5.2 Saran

Berdasarkan kesimpulan dan analisis laporan, saran dari peneliti untuk penelitian lebih lanjut yaitu untuk penelitian berikutnya, disarankan mengklasifikasikan hasil identifikasi kepakaran yang spesifik menjadi lebih luas berdasarkan program studi masing-masing, jadi tidak hanya terbatas pada satu program studi saja yaitu informatika, dan dapat secara luas meneliti program studi yang lain dari universitas yang berbeda-beda.

DAFTAR PUSTAKA

- Arisandi, D. dan Nofriandi (2019) “Analisa Pola Pemilihan Program Studi Bagi Calon Mahasiswa Di Universitas Abdurrah Menggunakan Association Rule,” *jurnal.univrab.ac.id* [Preprint].
- Dogan, A. dan Applications, D.B. (2021) “Machine learning and data mining in manufacturing,” *Elsevier* [Preprint].
- Fatmawati, M. (2017) “Pengklasteran Laporan Tugas Akhir Berdasarkan Abstrak Menggunakan Metode Rapid Automatic Keyphrase Extraction Dan Average Linkage Hierarchical Clustering.”
- Hakim, L. (2020) “Analisis Bibliometrik Penelitian Inkubator Bisnis pada Publikasi Ilmiah Terindeks Scopus,” *Procuratio: Jurnal Ilmiah Manajemen*, 8(2), hal. 176–189.
- Ibnu Subroto, I.M. dan Chaerul Haviana, S.F. (2018) “Proposal of the S-score for measuring the performance of researchers, institutions, and journals in Indonesia,” *Chemical and Engineering News*, 73(13), hal. 42–49.
- Muhtar, R., Gerhana, Y. dan Maylawati, D. (2020) “Multinomial Naïve Bayes and Rapid Automatic Keywords Extraction for Taharah (Purify) Law Chatbot.”
- Natasuwarna, A.P. (2019) “Tantangan menghadapi era revolusi 4.0 - Big data and data mining,” *Seminar Nasional Hasil Pengabdian Kepada Masyarakat*, hal. 23–27.
- Rahmadi Islam (2018) “Analisa Integrasi Data SINTA (Science and Technology Index) Menggunakan Website Internasional Dengan Manajemen Sistem Informasi EIS (Executive Information System),” *Analisa Integrasi Data SINTA (Science and Technology Index) Menggunakan Website Internasional Dengan Manajemen Sistem Informasi EIS (Executive Information System)*, (3), hal. 1–13.

- Rose, S. dan Engel, D. dkk (2010) “Automatic Keyword Extraction from Individual Documents,” *Text Mining: Applications and Theory*, (October 2017), hal. 1–20.
- Saputra, A. (2020) “Pemanfaatan Science and Technology Index (SINTA) untuk Publikasi Karya Ilmiah dan Pencarian Jurnal Nasional Terakreditasi,” *Media Pustakawan*, 27(1), hal. 56–68.
- Singh, V.K., Singh, P. dan Karmakar, M. (2021) “The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis,” *Scientometrics*, 126(6), hal. 5113–5142.
- Syahrul, A. dan Solichin, A. (2022) “Rekomendasi Pemilihan Mata Kuliah Dalam Pengisian Rencana Studi Mahasiswa dengan Penerapan Algoritma Apriori,” *eltikom.poliban.ac.id*, 79(1). Tersedia pada: <https://doi.org/10.31961/eltikom.v6i1.549>.
- Teodoro, V.D. dan Schwartz, J.L. (2012) “Encyclopedia of the Sciences of Learning,” *Encyclopedia of the Sciences of Learning* [Preprint], (March).
- Thushara, M.G., Mownika, T. dan Mangamuru, R. (2019) “A comparative study on different keyword extraction algorithms,” *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, (Iccmc), hal. 969–973.
- Wahyudi, R. dan Berlilana (2022) “Meningkatkan Visibilitas Program Studi melalui Update Profile SINTA (Science and Technology Index),” *uby.ac.id*, 3(2), hal. 32–38.
- Wahyuni, T. (2018) “Analisis Integrasi Data pada Pangkalan Data Kekayaan Intelektual (PDKI) Indonesia (Nasional) dan Scopus (Internasional) dengan SINTA 2 (Science and Technology Index),” *Analisis Integrasi Data pada Pangkalan Data Kekayaan Intelektual (PDKI) Indonesia (Nasional) dan Scopus (Internasional) dengan SINTA 2 (Science and Technology Index)*, (3), hal. 1–13.

LAMPIRAN

