

**IDENTIFIKASI KEPAKARAN DOSEN BERDASARKAN  
REKAM JEJAK PUBLIKASI PADA SINTA TERINDEKS  
SCOPUS MENGGUNAKAN *YET ANOTHER KEYWORD  
EXTRACTOR (YAKE)***

**LAPORAN TUGAS AKHIR**

Laporan ini disusun guna memenuhi salah satu syarat untuk menyelesaikan program studi Teknik Informatika S-1 pada Fakultas Teknologi Industri Universitas Islam Sultan Agung



**Disusun Oleh:**

**Muhamad Reynaldi Ilhamsyah**

**NIM 32601900020**

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

**2023**

***IDENTIFICATION OF LECTURER EXPERTISE BASED ON  
PUBLICATION TRACK RECORD ON SCOPUS INDEXED  
SINTA USING YET ANOTHER KEYWORD EXTRACTOR  
(YAKE)***

***FINAL PROJECT***

*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:**

**Muhamad Reynaldi Ilhamsyah**

**NIM 32601900020**

**MAJORING OF INFORMATICS ENGINEERING  
INDUSTRIAL TECHNOLOGY FACULTY  
SULTAN AGUNG ISLAMIC UNIVERSITY  
SEMARANG  
2023**

## LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul **“Identifikasi Kepakaran Dosen Berdasarkan Rekam Jejak Publikasi Pada Sinta Terindeks Scopus Menggunakan *Yet Another Keyword Extractor (YAKE)*”** ini disusun oleh:

Nama Muhamad Reynaldi Ilhamsyah

NIM 32601900020

Program Studi Teknik Informatika

Telah disah oleh dosen pembimbing Pada


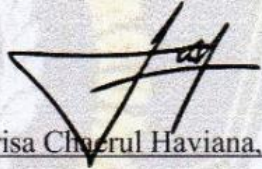
Hari Senin

Tanggal 28 Agustus 2023

Mengesahkan,

Pembimbing I

Pembimbing II

 Imam Much Ibnu Subroto, ST, M.Sc, Ph.D.  Sam Farisa Chyterul 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 Kepakaran Dosen Berdasarkan Rekam Jejak Publikasi Pada Sinta Terindeks Scopus Menggunakan *Yet Another Keyword Extractor (YAKE)*” ini telah dipertahankan di depan tim penguji Tugas Akhir pada.

Hari                                      Senin

Tanggal                                 27 November 2023

### TIM PENGUJI

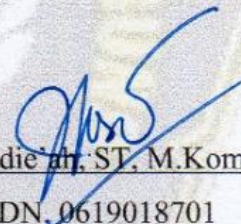
Ketua Penguji



Andi Riansyah, ST, M.Kom

NIDN. 0609108802

Anggota I



Badie An, ST, M.Kom

NIDN. 0619018701



## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini

Nama Muhamad Reynaldi Ilhamsyah

NIM 32601900020

Judul Tugas Akhir Identifikasi Kepakaran Dosen Berdasarkan Rekam Jejak Publikasi Pada Sinta Terindeks Scopus Menggunakan *Yet Another Keyword Extractor (YAKE)*

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,



Muhamad Reynaldi Ilhamsyah

## 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 Kepakaran Dosen Berdasarkan Rekam Jejak Publikasi Pada Sinta Terindeks Scopus Menggunakan *Yet Another Keyword Extractor (YAKE)*” 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. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH, M.Hum. yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T, M.T.
3. Bapak Imam Much Ibnu Subroto, ST., M.Sc.,Ph.D selaku Dosen Pembimbing I yang telah meluangkan waktu dan memberikan ilmu kepada penulis.
4. Bapak Sam Farisa Chaerul Haviana, ST., M.Kom selaku Dosen Pembimbing II yang telah meluangkan waktu dan memberikan ilmu kepada penulis.
5. Orang tua penulis yang telah mengizinkan untuk menyelesaikan laporan ini.
6. Semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan segala kerendahan hati, penulis menyadari masih banyak terdapat kekurangan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini dan masa mendatang.

Semarang, 4 Desember 2023



Muhamad Reynaldi Ilhamsyah

## DAFTAR ISI

<b>LEMBAR JUDUL</b> .....	<b>i</b>
<b>LEMBAR PENGESAHAN PEMBIMBING</b> .....	<b>iii</b>
<b>LEMBAR PENGESAHAN PENGUJI</b> .....	<b>iv</b>
<b>SURAT PERNYATAAN KEASLIAN TUGAS AKHIR</b> .....	<b>v</b>
<b>PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>viii</b>
<b>DAFTAR GAMBAR</b> .....	<b>x</b>
<b>DAFTAR TABEL</b> .....	<b>xi</b>
<b>ABSTRAK</b> .....	<b>xii</b>
<b>BAB I</b> .....	<b>1</b>
<b>PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah.....	2
1.3 Pembatasan Masalah .....	2
1.4 Tujuan.....	2
1.5 Manfaat.....	3
1.6 Sistematika Penulisan.....	3
<b>BAB II</b> .....	<b>5</b>
<b>TINJAUAN PUSTAKA DAN DASAR TEORI</b> .....	<b>5</b>
2.1 Tinjauan Pustaka .....	5
2.2 Dasar Teori .....	7
2.2.1 Kepakaran .....	7
2.2.2 Data Mining .....	7
2.2.3 Scopus .....	9
2.2.4 <i>Yet Another Keyword Extractor (YAKE)</i> .....	10
2.2.5 Sinta .....	19
2.2.6 <i>Mean Average Precision@K (MAP@K)</i> .....	21
<b>BAB III</b> .....	<b>22</b>



<b>METODE PENELITIAN .....</b>	<b>22</b>
3.1 Deskripsi Penelitian.....	22
3.2 Metode Ekstraksi Kepakaran.....	24
3.2.1 Pengumpulan Data .....	25
3.2.2 Ekstraksi <i>Keyword</i> Menggunakan Algoritma YAKE.....	25
3.2.3 <i>Filtering Keyword</i> .....	29
3.3 Metode Pengembangan Sistem .....	31
3.3.1 <i>Requirements analysis</i> .....	32
3.3.2 Desain.....	33
3.3.3 <i>Prototyping</i> .....	41
3.3.4 <i>Customer Evaluation</i> .....	42
3.3.5 <i>Review and Refine</i> .....	43
3.3.6 <i>Develop</i> .....	43
3.3.7 <i>Test</i> .....	43
3.3.8 <i>Release</i> .....	44
3.4 Rancangan Perhitungan MAP@K.....	44
<b>BAB IV .....</b>	<b>46</b>
<b>HASIL DAN ANALISIS PENELITIAN.....</b>	<b>46</b>
4.1 Hasil Ekstraksi Kepakaran .....	46
4.2 Hasil Implementasi Sistem.....	47
4.3 Hasil Pengujian Sistem.....	53
<b>BAB V.....</b>	<b>57</b>
<b>KESIMPULAN DAN SARAN .....</b>	<b>57</b>
5.1 Kesimpulan.....	57
5.2 Saran .....	57
<b>DAFTAR PUSTAKA .....</b>	<b>58</b>
<b>LAMPIRAN.....</b>	<b>60</b>



## DAFTAR GAMBAR

Gambar 3. 1 Flowchart Penelitian.....	22
Gambar 3. 2 Flowchart Ekstraksi Kepakaran .....	24
Gambar 3. 3 Alur pengembangan sistem prototype.....	31
Gambar 3. 4 Flowchart Sistem.....	33
Gambar 3. 5 Use Case Sistem .....	34
Gambar 3. 6 Gambar ERD .....	35
Gambar 3. 7 Rancangan user interface home.....	38
Gambar 3. 8 Rancangan hasil berdasarkan nama dosen .....	39
Gambar 3. 9 Rancangan hasil berdasarkan keyword .....	40
Gambar 3. 10 Rancangan user interface about.....	41
Gambar 4. 1 Hasil ekstraksi kepakaran.....	46
Gambar 4. 2 Hasil kepakaran pada sistem .....	46
Gambar 4. 3 Tampilan menu home.....	48
Gambar 4. 4 Tampilan sistem berdasarkan pencarian nama dosen .....	48
Gambar 4. 5 Tampilan pencarian nama dosen imam.....	49
Gambar 4. 6 Tampilan pencarian nama dosen imam 2.....	50
Gambar 4. 7 Tampilan hasil sistem berdasarkan pencarian kepakaran dosen .....	50
Gambar 4. 8 Tampilan sistem berdasarkan detail .....	51
Gambar 4. 9 Tampilan menu about.....	52
Gambar 4. 10 Tampilan menu about unissula.....	52
Gambar 4. 11 Tampilan hasil perhitungan MAP@K.....	56

## DAFTAR TABEL

Tabel 3. 1 Tabel Pra Pemrosesan teks.....	26
Tabel 3. 2 Tabel deduplikasi data .....	28
Tabel 3. 3 Daftar filtering keyword.....	30
Tabel 3. 4 tabel article_scopus .....	36
Tabel 3. 5 Tabel article_doc_scopus.....	36
Tabel 3. 6 Tabel author .....	36
Tabel 3. 7 Tabel prodi .....	37
Tabel 3. 8 Tabel afiliasi.....	37
Tabel 3. 9 Tabel author_expertise.....	37
Tabel 4. 1 Pengujian blackbox.....	53
Tabel 4. 2 Mean Average Precision.....	56



## ABSTRAK

Perkembangan sistem pendidikan yang semakin digital dan global mempengaruhi berbagai aspek pendidikan, termasuk dalam hal penentuan kepakaran dosen pada perguruan tinggi. Namun, di Universitas Islam Sultan Agung Semarang, sistem pencarian kepakaran dosen masih belum tersistem dengan baik. Tujuan dari penelitian ini adalah mengembangkan sebuah sistem yang dapat mengidentifikasi kepakaran dosen berdasarkan rekam jejak publikasi dosen yang terindeks dalam *scopus* pada SINTA. Metode yang digunakan, yaitu metode *Yet Another Keyword Extractor* (YAKE) yang akan menghasilkan sekumpulan kata kunci dan skor relevansinya yang akan dijadikan data kepakaran dosen secara spesifik. Dalam pengembangan sistem ini setiap dosen akan menghasilkan kata kunci dengan variasi 1 hingga 3 frasa kata kunci yang dihasilkan, dan jumlah kata kunci maksimal dalam penelitian ini adalah 20 kata kunci setiap dosen berdasarkan perhitungan metode *mean average precision@k* (MAP@K). Sehingga, setiap dosen akan memiliki jumlah kata kunci yang berbeda-beda sesuai dengan isi publikasi dan hasil ekstraksi metode *Yet Another Keyword Extractor* (YAKE) pada publikasi dosen tersebut. Jumlah total hasil kata kunci dalam ekstraksi penelitian ini yaitu 3264 kata kunci. Dengan memanfaatkan metode ini dapat secara lebih objektif mengidentifikasi serta mewakili tingkat kepakaran seorang dosen dalam bidang tertentu secara spesifik dan diharapkan dapat meningkatkan efisiensi dalam proses pendidikan, mendorong kolaborasi antar dosen, dan memberikan pandangan yang lebih jelas tentang bidang kepakaran di lingkungan Universitas Islam Sultan Agung Semarang.

Kata Kunci : Sistem Kepakaran Dosen, *Yet Another Keyword Extractor* (YAKE), SINTA, *Scopus*.

## ABSTRACT

*The development of an increasingly digital and global education system influences various aspects of education, including determining the expertise of lecturers in higher education. However, at Sultan Agung Islamic University Semarang, the system for searching for lecturer expertise is still not well systemized. The aim of this research is to develop a system that can identify lecturer expertise based on the track record of lecturers' publications indexed in Scopus on SINTA. The method used is the Yet Another Keyword Extractor (YAKE) method which will produce a set of keywords and their relevance scores which will be used as specific lecturer expertise data. In developing this system, each lecturer will produce keywords with variations of 1 to 3 keyword phrases produced, and the maximum number of keywords in this research is 20 keywords per lecturer based on the calculation of the mean average precision@k (MAP@K) method. So, each lecturer will have a different number of keywords according to the content of the publication and the extraction results of the Yet Another Keyword Extractor (YAKE) method in the lecturer's publications. The total number of keyword results in this research extraction was 3264 keywords. By utilizing this method, we can more objectively identify and represent the level of expertise of a lecturer in a specific field and it is hoped that it can increase efficiency in the educational process, encourage collaboration between lecturers, and provide a clearer view of the field of expertise within the Sultan Agung Islamic University, Semarang.*



*Keywords : Lecturer Expertise System, Yet Another Keyword Extractor (YAKE), SINTA, Scopus.*



# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Perkembangan sistem pendidikan di Indonesia terus mengalami peningkatan diberbagai sektor. Hal itu terjadi karena tuntutan zaman serta mengikuti arus perkembangan globalisasi yang semakin pesat, saat ini hampir semua aspek pendidikan sudah menggunakan sistem digitalisasi. Perubahan ini dinilai positif karena mempermudah dan mengefisiensi proses pembelajaran.

Perubahan sistem itu juga terjadi pada institusi perguruan tinggi, pada institusi ini hampir semua aspek sudah menggunakan sistem digitalisasi dalam menjalankan prosesnya, seperti proses pembelajaran, sistem penilaian, publikasi hasil ilmiah dan sebagainya. Namun, ada beberapa aspek yang belum tersistem dengan baik dalam proses berjalannya seperti penentuan suatu kepakaran dosen dalam hal atau bidang tertentu.

Hal itu juga terjadi pada Universitas Islam Sultan Agung (UNISSULA) pada Universitas ini ada banyak sistem digital dalam proses pembelajarannya. Akan tetapi, belum ada sistem pencari kepakaran dosen. Padahal hal ini dirasa sangat penting karena dapat membantu proses pendidikan seperti mata kuliah yang diampu dosen tersebut akan linier sesuai kepakaran atau keahliannya dan memberikan gambaran mengenai kondisi kepakaran dosen di UNISSULA.

Atas dasar itulah sistem penentuan kepakaran dosen sangat diperlukan, penentuan pakar tersebut harus bersifat objektif berdasarkan parameter yang telah ditentukan. Parameter tersebut dapat berupa riwayat pendidikan, riwayat riset, riwayat proyek, riwayat mengajar dan sebagainya, penulis dalam tugas akhir ini akan berfokus pada rekam jejak publikasi karya ilmiah atau jurnal dosen dalam menentukan suatu kepakaran.

Data yang digunakan berupa publikasi dosen UNISSULA pada SINTA yang terindeks scopus terutama judul publikasinya sebagai acuan untuk menentukan suatu kepakarannya. Data publikasi ini mencakup publikasi dosen mulai pada tahun 1987 hingga tahun 2023, sehingga ada banyak variasi data yang akan dihasilkan pada sistem.

Metode yang digunakan dalam penelitian ini yaitu *yet another keyword extractor* atau YAKE, metode ini dirasa sangat tepat untuk diimplementasikan pada sistem dalam menentukan suatu kepakaran dosen berdasarkan publikasi dosen tersebut.

Harapannya dengan tugas akhir ini akan menciptakan sistem kepakaran dosen berdasarkan rekam jejak publikasi pada SINTA terindeks scopus yang sesuai dengan kepakaran atau keahliannya terkhusus di UNISSULA yang relevan dan dapat dipertanggungjawabkan kevalidan sumber data yang digunakan pada sistem.

## **1.2 Perumusan Masalah**

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah belum adanya sistem pencari kepakaran dosen yang sistematis terutama pada Universitas Islam Sultan Agung sedangkan hal tersebut sangat dibutuhkan untuk membantu proses pendidikan seperti mata kuliah yang diampu dosen tersebut akan linier sesuai kepakaran atau keahliannya, kolaborasi kepakaran antar dosen, dan memberikan gambaran mengenai kondisi kepakaran dosen pada perguruan tinggi tersebut.

## **1.3 Pembatasan Masalah**

Batasan masalah pada tugas akhir ini adalah:

1. Penentuan sistem kepakaran seorang dosen pada lingkup Universitas Islam Sultan Agung Semarang
2. Hasil kepakaran dosen di sini merujuk pada kompetensi yang spesifik, bukan cakupan dalam bidang ilmu, karena informasi ini diperoleh secara langsung dari ekstraksi kata kunci judul-judul publikasi yang dihasilkan oleh dosen.
3. Data bersumber dari SINTA yang terindeks scopus.
4. Validasi keahlian dosen tidak dilakukan karena informasi bersumber dari database Scopus.
5. Kepakaran yang dicari tingkat author.

## **1.4 Tujuan**

Tujuan dari tugas akhir ini adalah mengembangkan sistem penentuan



kepakaran dosen berdasarkan rekam jejak publikasi pada sinta terindeks scopus menggunakan metode *Yet Another Keyword Extractor (YAKE)* pada lingkup Universitas Islam Sultan Agung Semarang.

### **1.5 Manfaat**

Manfaat penelitian tugas akhir adalah sebagai berikut:

1. Memudahkan pengguna untuk mencari informasi mengenai kepakaran dosen di lingkup Universitas Islam Sultan Agung Semarang.
2. Penerapan metode *Yet Another Keyword Extractor (YAKE)* untuk ekstraksi kata kunci dari publikasi dosen yang diidentifikasi sebagai kepakaran dosen.

### **1.6 Sistematika Penulisan**

Sistematika penulisan yang digunakan penulis untuk pembuatan laporan penelitian tugas akhir ini adalah sebagai berikut:

#### **BAB 1: PENDAHULUAN**

Dalam tahap ini, akan diuraikan latar belakang, perumusan masalah, pembatasan masalah, tujuan, manfaat, dan sistematika penulisan.

#### **BAB 2: TINJAUAN PUSTAKA DAN DASAR TEORI**

Pada tahap ini menampilkan beberapa studi pustaka yang berhubungan dengan penelitian tugas akhir. Dasar teori juga menjelaskan literatur tentang Identifikasi kepakaran dosen berdasarkan rekam jejak publikasi pada sinta terindeks scopus menggunakan *Yet Another Keyword Extractor (YAKE)*.

#### **BAB 3: METODE PENELITIAN**

Pada tahap ini, akan diungkapkan rangkaian langkah yang digunakan dalam merancang sistem, serta pendekatan yang diambil untuk mencari solusi terhadap permasalahan yang dihadapi. Proses ini melibatkan langkah-langkah seperti analisis perhitungan, analisa alur, penerapan metode, dan merancang sistem yang akan diimplementasikan di kemudian hari. Semua tahapan ini bertujuan untuk mendapatkan pemahaman yang lebih mendalam dan solusi yang tepat terhadap permasalahan yang ada.

#### **BAB 4: HASIL DAN ANALISIS PENELITIAN**

Pada bagian ini, akan disajikan hasil dari penelitian yang telah dilakukan, termasuk

pengujian sistem, data yang telah dirancang, dan diimplementasikan dalam penelitian yang sedang berlangsung.

#### BAB 5: KESIMPULAN DAN SARAN

Bagian ini akan menampilkan ringkasan atau kesimpulan yang diambil dari semua bab, beserta rekomendasi atau saran yang timbul dari hasil yang diperoleh pada penelitian ini untuk pemanfaatan kemajuan sistem yang telah dikembangkan pada masa depan.



## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Kepakaran dosen adalah sebuah keahlian yang dimiliki oleh seorang dosen pada bidang tertentu, kepakaran dosen ini sangat diperlukan untuk dimanfaatkan dalam beberapa hal. Ditangan seorang pakar yang kompeten di bidangnya, penyelesaian masalah dan pencarian solusi dapat dilakukan dengan lebih baik, efektif, dan efisien.

Seorang pakar akan menyelidiki permasalahan dengan menggunakan prinsip-prinsip ilmiah yang dimilikinya dan memanfaatkan pengalaman dalam menghadapi masalah serupa. Hal ini dapat mengurangi kemungkinan kesalahan atau kegagalan dalam menganalisis masalah, merencanakan solusi, merancang implementasi, menentukan rentang waktu penyelesaian, jangka waktu, dan estimasi biaya. Sebagai hasilnya informasi, pandangan, pendapat, dan komentar yang diberikan oleh pakar tersebut dapat diandalkan dan terpercaya (Firdaus, 2020).

Menentukan suatu kepakaran tidaklah mudah diperlukan sebuah sistem yang sistematis. Namun, pada saat ini masih belum ada sistem pencari kepakaran dosen yang sistematis dan dapat dipercaya serta akurat data kepakarannya terutama pada lingkup UNISSULA. Permasalahan itu dapat diatasi dengan membuat sistem pencari kepakaran dosen memanfaatkan data mining.

Dalam beberapa tahun terakhir, penelitian dalam bidang data mining telah mengalami perkembangan yang signifikan. Data mining merupakan proses pengumpulan dan pengolahan data dengan tujuan untuk mengekstrak atau mencari informasi penting dari data tersebut. Proses ekstraksi dan pencarian informasi ini dapat dilakukan dengan menggunakan perangkat lunak yang dibantu oleh perhitungan matematika, statistika, atau teknologi kecerdasan buatan.

Penelitian mengenai data mining banyak digunakan untuk mengatasi berbagai masalah dalam berbagai bidang, termasuk di bidang pendidikan atau akademik, seperti prediksi kelulusan mahasiswa, analisis kelulusan mahasiswa, rekomendasi pemilihan konsentrasi program studi, evaluasi akademik mahasiswa,



dan rekomendasi pemilihan mata kuliah. Pemanfaatan data mining juga dapat diimplementasikan untuk mencari kepakaran dosen (Syahrul dan Solichin, 2022).

Dalam pembuatan sistem ini selain pemanfaatan data mining tentu saja membutuhkan data yang akurat dan terpercaya. Hal itu bisa didapatkan melalui publikasi ilmiah dosen yang terindeks scopus. Scopus bisa menjadi salah satu rujukan untuk membuat sistem ini dan mengetahui kepakaran dosen karena scopus merupakan rujukan untuk memperoleh referensi jurnal internasional berkualitas tinggi.

Pada penelitian ini untuk mencari kata kunci publikasi dosen akan menggunakan metode *Yet Another Keyword Extractor* atau disingkat YAKE. Metode YAKE merupakan metode ekstraksi kata kunci otomatis yang bertumpu pada fitur statistik teks yang diekstraksi dari dokumen tunggal untuk memilih kata kunci teks yang paling penting. YAKE tidak diperlukan pelatihan pada sekumpulan dokumen dan tidak bergantung pada kamus, korpus eksternal, ukuran teks, bahasa, atau domain.

Metode YAKE memungkinkan adaptasi yang mudah ke bahasa lain selain bahasa Inggris, khususnya ke bahasa minor yang alat pemrosesan bahasa sumber terbukanya langka. Ini merupakan keunggulan metode YAKE dibandingkan metode lain yang serupa. Hal tersebut tentu sangat mendukung dalam tugas akhir ini karena data yang digunakan yaitu scopus atau jurnal internasional sehingga banyak bahasa asing jadi penggunaan metode ini lebih relevan (Campos dan Mangaravite, 2018).

Selain penggunaan metode yang tepat pengambilan data juga perlu diperhatikan karena data inilah yang akan menjadi kata kunci yang nantinya diproses menjadi suatu kepakaran dosen, dalam tugas akhir ini akan mengambil data dari SINTA atau *Science and Technology Index*.

SINTA adalah portal yang dikembangkan oleh Kementerian Riset, Teknologi, dan Pendidikan Tinggi (Kemenristekdikti) sebagai sarana untuk mengkomunikasikan karya ilmu pengetahuan dan teknologi dari warga negara Indonesia, termasuk penilaian kinerja dosen dan peneliti, kinerja jurnal, serta kinerja institusi (Saputra, 2020).

SINTA juga berfungsi sebagai database atau pusat data jurnal nasional yang telah terakreditasi, sehingga dapat menjadi sumber referensi bagi mereka yang mencari jurnal nasional yang telah diakui kualitasnya oleh Kementerian Pendidikan dan Kebudayaan (Kemendikbud) (Papatung, 2022).

Jurnal-jurnal yang terdapat di SINTA merupakan sumber referensi yang layak digunakan untuk penelitian, sitasi, dan sejenisnya. Selain itu, SINTA juga terhubung langsung dengan Scopus, yang memberikan reputasi internasional pada jurnal-jurnal tersebut, sehingga kualitasnya tidak perlu diragukan lagi. Atas dasar itulah penggunaan data dari SINTA sangat layak diimplementasikan dalam penelitian tugas akhir ini.

## **2.2 Dasar Teori**

### **2.2.1 Kepakaran**

Secara umum seorang pakar adalah individu yang memiliki pengetahuan khusus, mampu menjelaskan respons atau tanggapan, belajar hal-hal baru tentang topik permasalahan, menyusun kembali pengetahuan yang diperoleh, serta dapat memilah aturan dan menentukan relevansi dalam kepakarannya (Yuliyana dan Sinaga, 2019).

Kepakaran dosen yang dimaksud pada penelitian ini adalah kepakaran atau keahlian dosen pada bidang tertentu secara spesifik bukan kepakaran bidang ilmu secara *general* atau umum, karena kepakaran dosen ini diambil berdasarkan kata kunci yang diekstrak dari judul publikasi dosen secara langsung sehingga menghasilkan kata kunci yang spesifik sesuai isi judul publikasi tersebut. Kata kunci yang dihasilkan dalam proses ini akan diidentifikasi sebagai kepakaran.

### **2.2.2 Data Mining**

Data mining adalah proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. Istilah data mining merujuk pada disiplin ilmu yang tujuan utamanya adalah menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang dimiliki.

Data mining sering disebut juga dengan istilah *Knowledge Discovery in*

*Database* (KDD), yang mencakup kegiatan pengumpulan dan penggunaan data historis untuk menemukan keteraturan, pola, atau hubungan dalam set data berukuran besar (Panggabean dan Silalahi, 2020).

Ada beberapa hal yang penting dan terkait dengan data mining:

1. Data mining adalah proses otomatis terhadap data yang ada.
2. Data yang diproses dalam data mining memiliki skala yang sangat besar.
3. Tujuan utama dari data mining adalah untuk mendapatkan hubungan atau pola yang dapat memberikan indikasi yang bermanfaat.

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Klasifikasi (*Classification*) melibatkan variabel target kategori, seperti mengelompokkan pendapatan menjadi tiga kategori: tinggi, sedang, dan rendah.
2. Deskripsi (*Description*) digunakan untuk menggambarkan pola dan kecenderungan data dengan cara yang sederhana.
3. Estimasi (*Estimation*) mirip dengan klasifikasi, tetapi variabel targetnya adalah numerik daripada kategori.
4. Prediksi (*Prediction*) serupa dengan klasifikasi dan estimasi, tetapi fokus pada nilai di masa mendatang.
5. Pengklusteran (*Clustering*) melibatkan pengelompokan record atau objek berdasarkan kemiripan, membentuk kelompok objek yang serupa satu sama lain.
6. Asosiasi (*Association*) bertujuan untuk menemukan atribut yang sering muncul bersamaan, seperti dalam analisis keranjang belanja dalam dunia bisnis.

Penelitian tentang data mining ini telah banyak dilakukan pada berbagai bidang dengan menggunakan teknik data mining untuk menggali berbagai informasi melalui database tertentu, termasuk menggunakan database dalam bidang akademik. Penerapan data mining dalam bidang akademik salah satu contohnya yaitu untuk menentukan kepakaran seorang dosen melalui publikasi dosen tersebut dengan menggunakan database dari scopus.

### 2.2.3 Scopus

Scopus adalah salah satu jurnal terbesar di dunia yang sering digunakan sebagai referensi dalam penelitian. Untuk mempublikasikan hasil penelitian di jurnal Scopus, penelitian tersebut harus melewati tahap-tahap review yang dilakukan oleh para pakar. Hanya setelah berhasil melewati tahap review tersebut, penelitian dapat dipublikasikan. Hal ini bertujuan untuk memastikan bahwa penelitian yang diterbitkan memiliki kualitas tinggi, termasuk aspek seperti orisinalitas ide, kompleksitas isi, dan kualitas penulisan (Ibrahim, 2019).

Scopus juga merupakan lembaga global yang melakukan indeksasi jurnal di bawah lisensi perusahaan Elsevier, sebuah organisasi penerbit publikasi ilmiah internasional yang berbasis di Amsterdam Belanda dan beroperasi sejak tahun 1880. Scopus menerapkan standar ketat untuk mengindeks jurnal di platformnya. Setiap jurnal juga dipantau secara berkala setiap kuartal untuk memeriksa apakah ada pelanggaran etika publikasi. Jika ditemukan pelanggaran, jurnal tersebut dapat dihapus dari indeks Scopus atau dihentikan.

Saat ini scopus telah mengindeks lebih dari 22.000 judul artikel jurnal dari lebih dari 5.000 penerbit. Selain menyajikan karya ilmiah, Scopus juga menyediakan data hak paten dari berbagai penelitian di seluruh dunia dan memberikan layanan untuk menilai dampak suatu jurnal dalam penelitian ilmiah (Nugroho, 2022).

Scopus memiliki berbagai fungsi, salah satunya adalah menghubungkan semua artikel ilmiah ke dalam portal yang digunakan sebagai alat pengukuran prestasi peneliti di berbagai negara. Scopus juga memiliki sistem penilaian yang disebut *Scimago Journal Rank* (SJR) untuk menilai artikel ilmiah. Selain itu, Scopus digunakan sebagai referensi untuk mendapatkan jurnal internasional berkualitas tinggi.

Ada hal yang menarik mengenai scopus, saat ini scopus telah menjadi standar publikasi jurnal bagi dosen di Indonesia. Dosen yang ingin mencapai jabatan akademik tertentu diwajibkan untuk mempublikasikan jurnal nasional yang memiliki reputasi baik dan terindeks di Scopus.

Beberapa hal itulah yang menjadikan alasan kenapa penulis memilih jurnal



yang sudah terindeks scopus menjadi data identifikasi kepakaran dosen. Selain itu, penentuan kepakaran juga membutuhkan data dengan kualitas tinggi dan valid agar menciptakan sebuah sistem yang memiliki kredibilitas dan dapat dipercaya kebenarannya.

Selain data yang valid penentuan metode yang tepat juga sangat diperlukan untuk mengekstraksi data tersebut menjadi sebuah kata kunci yang nantinya dapat diketahui suatu kepakaran melalui kata kunci tersebut, metode yang tepat dalam permasalahan ini yaitu metode *Yet Another Keyword Extractor*.

#### **2.2.4 Yet Another Keyword Extractor (YAKE)**

Metode *Yet Another Keyword Extractor* atau disingkat YAKE adalah metode ekstraksi kata kunci berdasarkan fitur statistik teks yang diekstrak dari satu dokumen untuk mengidentifikasi dan menentukan peringkat kata kunci yang paling penting dalam teks atau dokumen tersebut. Fitur statistik teks dalam metode *Yet Another Keyword Extractor* antara lain :

1. Frekuensi kata

Metode YAKE memperhitungkan berapa kali sebuah kata muncul dalam teks. Kata-kata yang sering muncul cenderung dianggap lebih penting.

2. IDF Inverse Document Frequency

Metode YAKE menghitung kebalikan frekuensi dokumen untuk setiap kata dalam kumpulan dokumen. Kata-kata yang jarang muncul di seluruh kumpulan dokumen dapat diberi bobot lebih tinggi karena mereka lebih informatif dan berpotensi menjadi kata kunci.

3. Kedalaman Kata atau Term Depth

Metode YAKE dapat menggunakan struktur teks untuk mengidentifikasi kata-kata yang muncul pada kedalaman yang lebih dalam (misalnya, dalam kalimat atau paragraf yang lebih panjang) sebagai kata kunci yang lebih informatif.

4. Panjang Kata

Metode YAKE mungkin memberikan nilai lebih tinggi pada kata-kata

yang panjang karena mereka cenderung mencerminkan konsep yang lebih kompleks atau spesifik.

#### 5. Kekhasan Kata atau Term Uniqueness

Metode YAKE dapat mengukur sejauh mana sebuah kata unik dalam konteks teks tertentu. Kata-kata yang lebih khas atau tidak umum cenderung mendapatkan bobot lebih tinggi sebagai kata kunci.

#### 6. Frekuensi Auksiliar

Metode YAKE mungkin juga menggunakan fitur seperti frekuensi bi-gram (pasangan kata), tri-gram (tiga kata), atau fitur n-gram lainnya untuk mengenali kombinasi kata yang penting. Integrasi fitur statistik teks membantu metode ini dalam mengekstraksi kata-kata kunci yang relevan dan bermakna dari teks tanpa memerlukan pengetahuan domain khusus atau pendekatan manual.

Dengan memanfaatkan karakteristik dan metrik ini, metode *Yet Another Keyword Extractor* dapat menghasilkan ekstraksi kata kunci yang lebih baik dan lebih otomatis dari berbagai jenis dokumen.

Metode YAKE dapat dengan cepat diterapkan pada dokumen berbagai bahasa atau domain dalam rentang waktu singkat karena sifatnya yang *plug and play*. Metode YAKE tidak memerlukan pelatihan untuk sekumpulan dokumen dan tidak bergantung pada kamus, korpus eksternal, ukuran teks, bahasa, atau domain.

Pada Metode YAKE bebas bahasa, kecuali untuk penggunaan daftar *stopword* yang berbeda tetapi statis untuk setiap bahasa. Ini memungkinkan adaptasi yang mudah dari metode ini ke bahasa lain, khususnya ke bahasa minor yang alat pemrosesan bahasa sumber terbukanya langka. Ini merupakan keunggulan metode YAKE dibandingkan metode lain yang serupa.

Pada metode YAKE semakin rendah nilai sebuah kata kunci, semakin penting kata kunci tersebut. Kata kunci dengan skor lebih rendah cenderung dianggap lebih penting karena dapat menunjukkan karakteristik unik dari teks tersebut. Kata kunci yang muncul lebih sering dalam teks cenderung mendapatkan skor yang lebih tinggi, karena algoritma berasumsi bahwa kata-kata ini lebih mencerminkan topik utama atau fokus teks. Sebaliknya, kata-kata yang jarang muncul mungkin memiliki

skor lebih rendah, tetapi mereka bisa jadi sangat penting untuk mengidentifikasi topik khusus atau istilah teknis yang lebih spesifik dalam teks.

Pada Metode YAKE dengan memberi bobot yang lebih rendah pada kata-kata yang lebih umum dan memberikan nilai lebih tinggi pada kata-kata yang lebih jarang muncul, algoritma ekstraksi kata kunci berupaya mengidentifikasi kata-kata kunci yang benar-benar berarti dalam teks tanpa terpengaruh oleh kata-kata yang lebih umum dan kurang informatif.

Metode YAKE hanya bergantung pada fitur statistik yang diekstrak dari teks itu sendiri memungkinkan penskalaan dengan mudah. Pada Sistem metode ini memiliki enam komponen utama:

1. Teks pra-pemrosesan

Pada langkah pra-pemrosesan metode YAKE membagi teks menjadi istilah individual setiap kali ruang kosong atau karakter khusus misal jeda baris, tanda kurung, koma, titik, dan lain lain.

2. Ekstraksi fitur

Pada ekstraksi fitur metode YAKE terdapat lima fitur untuk menangkap karakteristik dari setiap istilah individu antara lain:

- a. *Casing*

Fitur ini mempertimbangkan *casing* atau kata awalan. Pada fitur ini lebih mementingkan kata-kata dan akronim yang dikapitalisasi seperti "SINTA". Akronim itu sendiri merupakan singkatan yang terbentuk dari huruf-huruf awal atau beberapa suku kata dalam satu frasa atau nama. Biasanya, akronim digunakan sebagai representasi singkat untuk memudahkan komunikasi atau mengingat istilah yang panjang. Contoh akronim yaitu SINTA - *Science and Technology Index*.

Pada langkah pertama ini fitur menghitung berapa kali kata dimulai dengan huruf kapital padahal itu bukan kata awal kalimat. Fitur ini juga menghitung jumlah ketika kata tersebut dalam bentuk akronim. Kemudian, fitur ini mengambil maksimum dari dua hitungan dan menormalkannya dengan log dari jumlah total.

$$casing(w) = \frac{\max(count(w \text{ is capital}), count(w \text{ is acronym}))}{1 + \log(count(w))} \quad (1)$$

Penjelasan rumus :

$$1) \max(\text{count}(w \text{ is capital}), \text{count}(w \text{ is acronym}))$$

$\text{count}(w \text{ is capital})$  ini adalah jumlah kemunculan kata "w" dalam bentuk huruf kapital. Ini menghitung berapa kali kata "w" muncul dalam teks dengan huruf kapital.

$\text{count}(w \text{ is acronym})$  ini adalah jumlah kemunculan kata "w" sebagai akronim. Ini menghitung berapa kali kata "w" muncul sebagai singkatan singkat (biasanya dengan huruf kapital).

2) Fungsi "*max*" digunakan untuk mengambil nilai tertinggi antara kedua jumlah di atas. Ini memberikan informasi tentang bagaimana kata "w" sering muncul dalam bentuk huruf kapital atau sebagai akronim.

3)  $\log(\text{count}(w))$  ini adalah jumlah total kemunculan kata "w" dalam teks. Logaritma digunakan di sini untuk memberikan bobot pada jumlah kemunculan kata "w". Logaritma mengubah skala data dan dapat membantu dalam memberikan penekanan yang lebih baik pada kata-kata yang mungkin lebih jarang muncul tetapi masih penting.

4)  $1 + \log(\text{count}(w))$  merupakan jumlah total kemunculan kata "w" dalam teks. Hasil dari perhitungan logaritma pada langkah sebelumnya ditambahkan dengan 1. Ini adalah bagian dari penyebut dalam perhitungan skor casing. Akhirnya, rumus "*casing(w)*" menghitung rasio antara nilai maksimum dari huruf kapital dan akronim terhadap penyebut yang didapatkan dari penjumlahan 1 dengan logaritma dari jumlah total kemunculan kata "w".

Rumus ini mengukur sejauh mana kata "w" muncul dalam bentuk huruf kapital atau akronim dibandingkan dengan jumlah total kemunculannya, dengan memberikan bobot melalui logaritma. Semakin besar skor *casing*, semakin penting kemungkinan kata tersebut dalam teks.

#### b. *Word Positional*



Fitur ini lebih mementingkan kata-kata yang ada di awal dokumen. Ini didasarkan pada asumsi bahwa kata kunci yang relevan biasanya lebih terkonsentrasi di awal dokumen. Pertama, mendapatkan semua posisi kalimat di mana kata "w" muncul.

$$Sen(w) = \text{position of sentences where } w \text{ occurs} \quad (2)$$

Kemudian, menghitung fitur posisi dengan mengambil posisi median dan menerapkan rumus berikut:

$$position(w) = \log(\log(3 + Median(Sen(w)))) \quad (3)$$

Penjelasan rumus:

Rumus " $position(w)$ " digunakan untuk menghitung skor posisi ( $position\ score$ ) dari kata "w" dalam konteks metode YAKE. Berikut penjelasan komponen dari rumus tersebut:

- 1)  $Sen(w)$ : Ini merujuk pada posisi-posisi kalimat di mana kata "w" muncul dalam teks. Dalam konteks perhitungan ini, kita perlu menghitung berapa kalimat di mana kata "w" muncul.
- 2)  $Median(Sen(w))$ : Ini adalah nilai tengah ( $median$ ) dari posisi kalimat di mana kata "w" muncul dalam teks.  $Median$  adalah nilai yang terletak di tengah-tengah kumpulan nilai.
- 3)  $3 + Median(Sen(w))$ : Nilai  $median$  di atas ditambahkan dengan 3.
- 4)  $\log(3 + Median(Sen(w)))$ : Hasil dari perhitungan langkah sebelumnya diambil logaritma, dengan basis yang umumnya adalah 10, ini bertujuan untuk memberikan penekanan lebih pada kata-kata yang muncul dalam posisi yang lebih awal dalam teks.
- 5)  $\log(\log(3 + Median(Sen(w))))$ : Hasil dari langkah sebelumnya diambil logaritma lagi. Ini adalah bagian penting dari perhitungan skor posisi.

Rumus ini menghasilkan skor yang mencerminkan seberapa awal kata "w" muncul dalam teks, berdasarkan posisinya dalam kalimat. Semakin besar skor posisi, semakin awal kemungkinan kata tersebut muncul dalam teks, yang dalam beberapa konteks mungkin menunjukkan pentingnya kata tersebut dalam konteks yang diberikan.

### c. *Word Frequency*

Fitur ini menghitung frekuensi kata yang dinormalisasi dengan deviasi standar 1 dari rata-rata.

$$frequency(w) = \frac{count\ of\ word\ w}{mean(counts)+standard\ deviation(counts)} \quad (4)$$

Penjelasan rumus :

Rumus "*frequency(w)*" dalam metode YAKE digunakan untuk menghitung skor frekuensi (*frequency score*) dari kata "w". Tujuannya adalah untuk menilai seberapa umum atau langka kemunculan kata tersebut dalam teks, dengan mempertimbangkan distribusi frekuensi kata-kata dalam teks secara keseluruhan.

- 1) *count of word w* merupakan jumlah total kemunculan kata "w" dalam teks.
- 2) *mean(counts)* merupakan rata-rata dari jumlah total kemunculan semua kata dalam teks. Rata-rata ini mengukur frekuensi kata rata-rata dalam teks.
- 3) *Standard deviation(counts)* merupakan *deviasi standar* dari jumlah total kemunculan semua kata dalam teks. *Deviasi standar* mengukur sebaran frekuensi kata-kata dalam teks. *Standar deviasi* adalah ukuran statistik yang mengukur sejauh mana data tersebar atau bervariasi dari rata-rata (*mean*).

Rumus "*frequency(w)*" menghitung rasio antara jumlah kemunculan kata "w" dengan rata-rata frekuensi kata dalam teks, disesuaikan dengan sebaran frekuensi menggunakan *deviasi standar*. Hal ini membantu dalam memberikan penekanan pada kata-kata yang frekuensinya jauh lebih tinggi atau lebih rendah dari rata-rata.

Rumus ini membantu dalam mengidentifikasi kata-kata yang mungkin memiliki makna khusus dalam konteks teks tersebut, tanpa terlalu banyak dipengaruhi oleh kata-kata yang sangat umum atau sangat langka.

d. *Word Relatedness to Context*

Fitur ini menghitung seberapa terkait suatu kata dengan konteksnya. Untuk itu, dihitung berapa banyak istilah berbeda yang muncul di kiri atau kanan kata kandidat. Jika kata tersebut sering muncul dengan kata yang berbeda di sisi kiri atau kanan, kemungkinan besar itu adalah stopwords.

$$\text{relatedness}(w) = 1 + (WR + WL) * \frac{\text{count}(w)}{\text{max count}} + PL + PR \quad (5)$$

Dengan perhitungan sebagai berikut

$$WR = (\text{number of unique words on right}) / (\text{total words on right}) \quad (6)$$

WR merupakan faktor yang merujuk pada bobot yang diberikan pada relevansi dari kata "w" terhadap teks secara keseluruhan. WR adalah suatu nilai yang diasumsikan dan perlu ditentukan sebelumnya. Nilai WR digunakan untuk memberikan tingkat penting pada keterkaitan kata tersebut

$$WL = (\text{number of unique words on left}) / (\text{total words on left}) \quad (7)$$

WL juga merupakan faktor yang merujuk pada bobot yang diberikan pada lokasi atau posisi kata "w" dalam teks. WL juga adalah suatu nilai yang diasumsikan dan perlu ditentukan sebelumnya. Nilai WL digunakan untuk mengevaluasi apakah kata "w" muncul dalam posisi yang penting dalam teks.

$$PL = (\text{total words on left}) / (\text{max count}) \quad (8)$$

PL merupakan faktor yang merujuk pada penyesuaian berdasarkan lokasi kata "w" dalam teks terhadap konteks sebelumnya. Seperti WR dan WL, PL adalah suatu nilai yang perlu ditentukan sebelumnya.

$$PR = (\text{total words on right}) / (\text{max count}) \quad (9)$$

PR merupakan faktor yang merujuk pada penyesuaian berdasarkan lokasi kata "w" dalam teks terhadap konteks setelahnya. Seperti faktor lainnya, PR adalah suatu nilai yang perlu ditentukan sebelumnya.

e. *Word Different Sentence*

Fitur ini mengukur seberapa sering kata kandidat muncul dengan kalimat yang berbeda. Sebuah kata yang sering muncul dalam kalimat yang berbeda memiliki skor yang lebih tinggi.

$$\text{different}(w) = \frac{\text{number of sentences } w \text{ occurs in}}{\text{total sentences}} \quad (10)$$

Penjelasan rumus :

Rumus  $\text{different}(w)$  dalam metode YAKE digunakan untuk menghitung skor perbedaan (*different score*) dari kata "w" dalam teks. Tujuannya adalah untuk mengevaluasi sejauh mana kata tersebut muncul dalam berbagai kalimat yang berbeda dalam teks.

- 1) *Number of sentences w occurs in* merupakan jumlah kalimat di mana kata "w" muncul dalam teks. Rumus ini menghitung berapa kali kata "w" muncul dalam posisi yang berbeda dalam kalimat.
- 2) *Total sentences* merupakan total jumlah kalimat dalam teks.

Rumus  $\text{different}(w)$  mengukur rasio antara jumlah kalimat di mana kata "w" muncul dengan total jumlah kalimat dalam teks. Skor ini memberikan informasi tentang sejauh mana kata "w" tersebar dalam teks dan muncul dalam berbagai konteks kalimat yang berbeda.

Skor perbedaan dapat membantu dalam mengidentifikasi kata-kata yang mungkin memiliki makna yang lebih luas atau lebih serbaguna karena muncul dalam berbagai kalimat, dan ini bisa dianggap sebagai indikasi pentingnya kata tersebut dalam teks secara keseluruhan.

Gabungan *Word Score*

5 fitur ini digabungkan menjadi skor tunggal  $S(w)$  menggunakan rumus:

$$\text{score}(w) = \frac{a*b}{a + \left(\frac{c}{d}\right) + \left(\frac{e}{d}\right)} \quad (11)$$

Keterangan :

$a = \text{casing}$ ,  $b = \text{position}$ ,  $c = \text{frequency}$ ,  $d = \text{relatedness}$ ,  $e = \text{different}$

*Word score*

Dalam metode YAKE menghitung skor kata kunci (*keyword score*) dari suatu kelompok kata kunci (*keyword*). Tujuannya adalah untuk menilai sejauh mana sekumpulan kata kunci tertentu relevan dan penting dalam teks yang diberikan. Skor dihitung menggunakan rumus berikut

$$S(kw) = \frac{\text{product(score of words in keyword)}}{1 + (\text{sum of scores of words}) * \text{count(keyword)}} \quad (12)$$

Rumus  $S(kw)$  menghitung skor kata kunci dengan mengalikan skor



skor komponen kata-kata dalam kelompok kata kunci dan kemudian membaginya dengan faktor normalisasi yang dihasilkan dari jumlah skor komponen dan jumlah kata kunci dalam kelompok, Skor kata kunci ini memberikan nilai yang merefleksikan seberapa penting dan relevan kelompok kata kunci tersebut dalam teks secara keseluruhan.

### 3. Skor istilah individual

Langkah ketiga, pada tahap ini secara heuristik menggabungkan semua fitur ini menjadi satu ukuran.

### 4. Pembuatan daftar kata kunci kandidat

Pada langkah keempat metode YAKE membuat kata kunci kandidat sebagai urutan kata kunci yaitu 1, 2, dan 3-gram yang berdekatan.

### 5. Deduplikasi Data

Pada langkah kelima metode YAKE mengeliminasi kandidat serupa yang berasal dari langkah sebelumnya.

### 6. Pemingkatan

Akhirnya, sistem akan mengeluarkan daftar kata kunci yang relevan, dibentuk oleh 1, 2, 3-gram, sehingga semakin kecil skornya, semakin penting kata kunci tersebut (Campos dan Mangaravite, 2018).

Metode YAKE digunakan untuk menghitung bobot kata dengan memanfaatkan lima fitur dokumen dan menggabungkan bobot dari beberapa kata untuk menghitung bobot frasa. Kelima fitur ini meliputi *Term Casing*, *Term Position*, *Normalized Term Frequency*, *Term Relatedness to Context*, dan *Term Different Sentences*. Persamaan untuk menghitung bobot dapat ditemukan dalam persamaan berikut:

$$S(t) = \frac{T_{rel} \times T_{position}}{T_{case} + \frac{TF_{norm}}{T_{rel}} + \frac{T_{sentence}}{T_{rel}}} \quad (13)$$

$T_{rel}$  adalah cara untuk mengukur seberapa relevan suatu kata dengan konteks teks yang sedang dibahas.  $T_{position}$  adalah penilaian terhadap posisi kata dalam dokumen.  $T_{case}$  menghitung seberapa sering kata muncul ketika frasa tersebut tampil dengan huruf besar di awalnya.  $TF_{norm}$  adalah metode yang mirip dengan *Term Frequency*, namun menggunakan nilai TF yang telah dinormalisasi dalam rumus.

$T_{sentence}$  adalah jumlah kemunculan kata yang bersifat relatif dalam berbagai kalimat yang berbeda (Susanto, 2022).

Dalam pemanfaatan metode ini akan lebih efektif dengan mengambil data yang diproses dari sumber yang terpercaya dan dapat dipertanggungjawabkan kebenarannya. Salah satu sumber yang valid yaitu SINTA.

### 2.2.5 Sinta

SINTA atau *Science and Technology Index* merupakan portal yang dikembangkan oleh Kementerian Riset, Teknologi, dan Pendidikan Tinggi (Kemenristekdikti) sebagai platform untuk mengkomunikasikan kontribusi dalam ilmu pengetahuan dan teknologi dari warga negara Indonesia, termasuk evaluasi kinerja dosen dan peneliti, evaluasi jurnal, dan penilaian kinerja institusi.

Alat untuk mengukur kinerja penelitian ini diusulkan melalui diskusi kelompok terarah tahun 2017. Kelompok ahli mengidentifikasi indikator berdasarkan produk penelitian peneliti dan frekuensi kutipannya di Scopus dan Google Scholar. Indikator dan item evaluasi yang diusulkan untuk mengukur kinerja tidak hanya penulis tetapi juga dari institusi. Kinerja jurnal diukur dengan mengambil data dari jurnal Indonesia yang terindeks di Scopus, sistem akreditasi jurnal nasional Indonesia, dan frekuensi sitasi di Google Scholar.

Berdasarkan hal-hal penilaian tersebut, S-skor diusulkan. Data hasil penelitian dari seluruh dosen dan peneliti di Indonesia dikumpulkan dan dimasukkan ke dalam Iptek dan Indeks Teknologi SINTA portal. Data yang dikumpulkan dari jurnal dicocokkan melalui database jurnal nasional terakreditasi Indonesia ARJUNA (Subroto dan Haviana, 2018).

Berikut adalah fungsi SINTA:

1. Mengumpulkan data publikasi dan sitasi dari akademisi dan peneliti Indonesia melalui ID Google Scholar dan Scopus untuk melacak semua publikasi.
2. Menilai kinerja jurnal berdasarkan standar akreditasi dan sitasi dengan menggunakan data jurnal yang sudah terakreditasi atau dievaluasi dalam peringkat SINTA 1-6.
3. Memberikan analisis tentang profil institusi, penulis, dan jurnal dengan

menampilkan peringkat institusi teratas, penulis teratas, jurnal teratas, dan sitasi institusi terbaik.

4. Menampilkan SINTA *Metrics* yang mencakup peringkat institusi, penulis, dan jurnal berdasarkan perhitungan dan formula tertentu.
5. Memantau kinerja publikasi dosen dan peneliti (Saputra, 2020).

SINTA dikenal sebagai database atau pusat data jurnal nasional yang telah terakreditasi, dapat menjadi acuan utama bagi mereka yang mencari referensi dalam bentuk jurnal nasional yang memiliki kualitas yang diakui oleh KEMENDIKBUD. Jurnal-jurnal yang ada di SINTA memiliki kualitas yang memadai untuk dijadikan referensi, dikutip, dan sejenisnya.

SINTA juga memiliki koneksi langsung dengan Scopus dan Google Scholar, yang memberikan informasi tentang reputasi jurnal secara internasional, sehingga kualitasnya tidak diragukan lagi. SINTA juga dilengkapi dengan berbagai fitur yang mempermudah para dosen dalam memantau perkembangan publikasi mereka dan memudahkan pengguna lain dalam menemukan referensi jurnal yang sesuai dengan harapan dan kebutuhan mereka. Beberapa fitur tersebut mencakup:

1. *Citation*, memberikan informasi tentang posisi h-index dalam satu tahun terakhir di Google Scholar dan Scopus. Ini berarti bahwa h-index dari kedua database tersebut akan terlihat dalam reputasi jurnal di SINTA.
2. *Networking*, adalah fitur yang menjelaskan semua kerjasama yang pernah dilakukan oleh dosen yang menjadi penulis jurnal di SINTA. Ini memungkinkan pengguna untuk melihat jaringan kerjasama penulis dan mengukur kualitas jurnal berdasarkan kerjasama penelitian, karena jurnal yang berisi hasil penelitian kolaboratif dianggap lebih berkualitas.
3. *Research Output* adalah fitur yang menampilkan semua hasil keluaran atau output, termasuk publikasi jurnal, buku ilmiah, dan bentuk lainnya yang telah dihasilkan.
4. Fitur terakhir di SINTA adalah Score, yang menampilkan indeks keseluruhan publikasi jurnal, mulai dari indeks di SINTA, kemudian di Google Scholar, hingga di Scopus, serta hasil indeks di Inasti.

Selain itu, SINTA adalah sebuah portal indeksasi jurnal yang saat ini dikelola oleh Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Ada juga yang menyebut SINTA sebagai database untuk Jurnal Nasional yang telah mendapatkan akreditasi dari ARJUNA.

Atas dasar itulah maka penulis menggunakan SINTA sebagai rujukan data publikasi jurnal dosen karena sudah terindeks scopus dan memiliki banyak fitur yang mendukung tugas akhir ini. Selain itu SINTA juga dikembangkan oleh badan resmi negara yaitu Kementerian Riset, Teknologi dan Pendidikan Tinggi (Kemenristekdikti) sehingga sudah dipastikan keaslian dan keakuratan datanya (Wahyudi, 2022).

### **2.2.6 Mean Average Precision@K (MAP@K)**

*Mean Average Precision@K (MAP@K)* adalah metrik evaluasi yang umum digunakan dalam peringkat dan informasi pencarian. Metrik ini memberikan ukuran seberapa baik suatu sistem peringkat atau pencarian dapat mengurutkan item yang relevan di antara item yang tidak relevan. nilai yang lebih tinggi menunjukkan kinerja yang lebih baik. Metrik ini mempertimbangkan relevansi dan urutan item, sehingga memberikan evaluasi yang lebih holistik terhadap sistem peringkat atau pencarian daripada metrik yang hanya mempertimbangkan presisi atau recall pada suatu titik tertentu (Kohli dan Kar, 2022).

Dalam penelitian ini penggunaan metode atau matriks ini digunakan untuk mencari kata kunci maksimal setiap dosen sehingga hasil kata kunci lebih relevan untuk dijadikan data kepakaran dosen. Rumus metode ini sebagai berikut :

$$\text{Precision} = \frac{|{\text{relevant documents}} \cap {\text{retrieved documents}}|}{|{\text{retrieved documents}}|} \quad (14)$$

Penjelasan rumus:

*Relevant documents* adalah kumpulan dokumen yang dianggap relevan atau penting untuk suatu pertanyaan atau pencarian informasi. Ini biasanya ditentukan berdasarkan kriteria tertentu yang didefinisikan sebelumnya. Sedangkan, *retrieved documents* merupakan kumpulan dokumen yang berhasil diambil atau ditemukan oleh suatu sistem pencarian atau proses pengambilan informasi.



## BAB III METODE PENELITIAN

### 3.1 Deskripsi Penelitian

Metode penelitian adalah suatu pendekatan atau cara sistematis yang digunakan untuk mengumpulkan data, menemukan informasi yang terkait, guna mencapai tujuan dalam suatu penelitian. Dalam Penelitian ini terdapat beberapa Langkah untuk mencapai tujuan penelitian seperti pada gambar 3.1.



#### 1. Studi Literatur

Pada tahap pertama penulis melakukan peninjauan terhadap buku, *e- book*, artikel, journal untuk referensi dalam mengerjakan proses penelitian. Hal ini dilakukan untuk mendapatkan data dan informasi yang berhubungan dengan metode *Yet Another Keyword Extractor*, scopus, SINTA dan sistem kepakaran dosen

## 2. Perancangan

Pada tahap perencanaan penelitian, akan diterapkan pendekatan pengembangan sistem berbasis metode prototype. Keputusan ini didasarkan pada pertimbangan bahwa pendekatan prototype akan memudahkan pengembangan sistem kepakaran ini.

Pendekatan tersebut memungkinkan terjalinnya interaksi antara pengembang sistem dalam hal ini peneliti dengan pengguna dalam konteks ini adalah pengembangan SINTA dan masyarakat secara umum. Pendekatan ini diharapkan dapat mengatasi potensi ketidaksielarasan antara pengembang dan pengguna dalam hal perspektif dan harapan terhadap sistem yang dikembangkan.

## 3. Implementasi

Pada tahap implementasi, peneliti akan menggunakan bahasa pemrograman *Python* untuk mengimplementasi pada data yang diperlukan. Keputusan ini didasarkan pada pertimbangan komprehensif terkait kemampuan *Python* dalam menganalisis data secara efektif. Selanjutnya, dalam merancang dan mengembangkan sistem ini akan berbasis website untuk memudahkan pengguna.

Dalam pengembangannya peneliti menggunakan platform *Visual Studio Code* sebagai platform pengembangan, integrasi Bootstrap CSS untuk antarmuka pengguna, dan pemanfaatan bahasa pemrograman PHP dalam logika sistemnya untuk mencapai implementasi yang optimal dan efisien dari sistem yang dikembangkan dalam lingkup penelitian ini.

## 4. Laporan

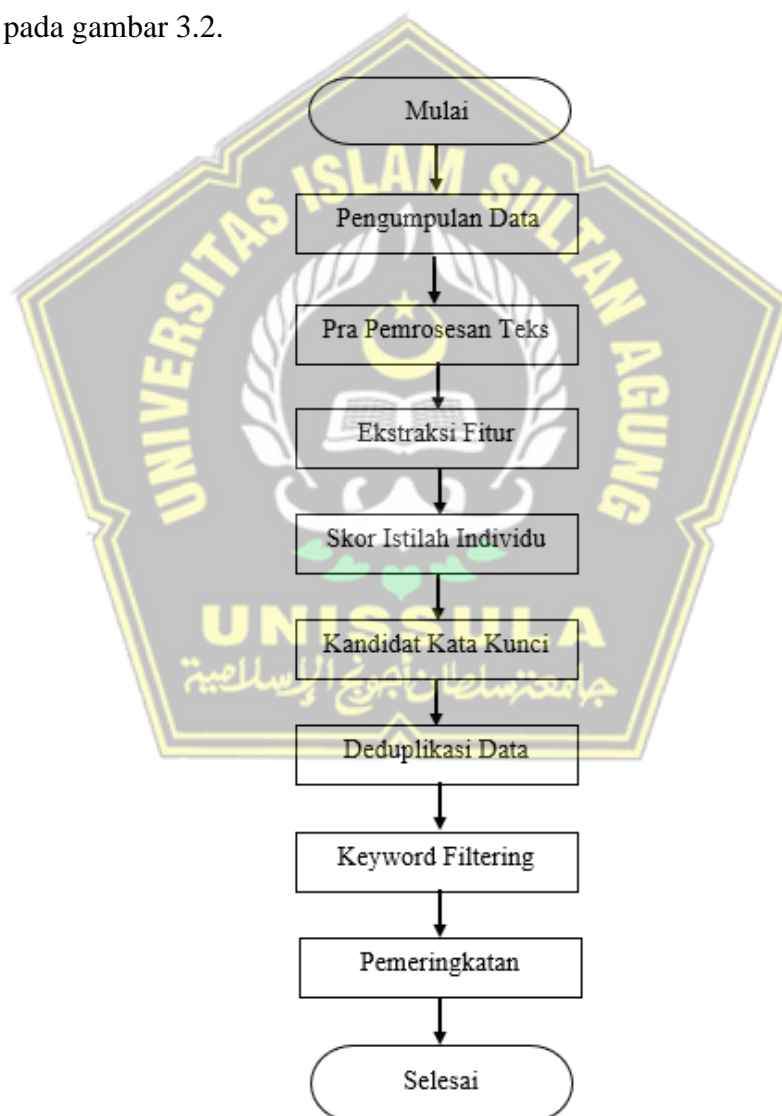
Tahap terakhir dalam rangkaian penelitian ini adalah tahap penyusunan laporan. Pada tahap ini, peneliti akan menyusun laporan yang merangkum seluruh aspek yang relevan dengan penelitian ini. Proses penyusunan laporan mencakup tahapan studi literatur, perancangan, dan implementasi, yang dijelaskan secara berurutan dan sistematis.

Tujuan dari tahap ini adalah untuk menyajikan penelitian secara komprehensif dan terstruktur, dengan maksud untuk memfasilitasi pemahaman

yang holistik bagi pengguna atau pihak yang tertarik terhadap hasil penelitian ini.

### 3.2 Metode Ekstraksi Kepakaran

Pada tahap ekstraksi kepakaran, terdapat beberapa proses yang dilaksanakan secara berurutan. Metode yang digunakan adalah metode *Yet Another Keyword Extractor* atau YAKE untuk mengekstraksi kata kunci dari judul publikasi dosen yang diidentifikasi sebagai bidang kepakaran secara spesifik. Dalam konteks ini, berikut adalah tahapan-tahapan yang terlibat dalam proses ekstraksi kepakaran seperti pada gambar 3.2.



Gambar 3. 2 *Flowchart* Ekstraksi Kepakaran

### 3.2.1 Pengumpulan Data

Pada tahap pengumpulan data, peneliti melakukan pengumpulan data publikasi yang terkait dengan dosen Universitas Islam Sultan Agung melalui database pada SINTA, yang secara spesifik hanya memasukkan data yang terindeks pada Scopus. Total kuantitas data dosen yang diperoleh dari berbagai fakultas yang ada di Universitas ini adalah sebanyak 182 dosen.

Keanekaragaman bidang ilmu di UNISSULA tercermin dalam jumlah tersebut, dengan keterlibatan dosen-dosen dari 11 fakultas yang berbeda. Kondisi ini menghasilkan keragaman dalam data kepakaran, yang mencakup beragam bidang keilmuan yang mencerminkan fokus kegiatan akademik universitas secara menyeluruh.

### 3.2.2 Ekstraksi *Keyword* Menggunakan Algoritma YAKE

Pada penelitian ini kepakaran diambil dari pengalaman publikasi hasil penelitian yang berkualitas yaitu dari publikasi terindeks scopus yang terdata di SINTA. Definisi kepakaran diekstrak dari *keyword* menggunakan metode *Yet Another Keyword Extractor* atau YAKE.

Pada tahap ekstraksi *keyword* bertujuan untuk mengetahui kata kunci yang dihasilkan dari publikasi tersebut. Pada tahap ini juga akan diketahui skor tingkat kepakaran pada kata kunci yang dihasilkan. Berikut merupakan *source code* yang digunakan pada *library* metode YAKE.

```
import yake
```

*Source code* menampilkan kode *library yake*, dengan memanfaatkan *library* ini maka akan didapatkan kata kunci dari database yang ada. Dalam penelitian ini kata kunci tersebut diidentifikasi sebagai kepakaran dosen secara spesifik.

1. `custom_kw_extractor = yake.KeywordExtractor()`
2. `keywords = custom_kw_extractor.extract_keywords(text)`
3. `print(keywords)`

Pada *Source code* menampilkan kode `custom_kw_extractor = yake.KeywordExtractor()` kode tersebut membuat sebuah objek `'custom_kw_extractor'` dari kelas `'KeywordExtractor'` dalam *library yake*. Objek ini akan digunakan untuk melakukan ekstraksi kata kunci dari teks.



Kemudian pada baris kedua terdapat *keywords* = *custom\_kw\_extractor.extract\_keywords(text)*, menggunakan objek *custom\_kw\_extractor* untuk mengekstrak kata kunci dari teks yang telah dikumpulkan sebelumnya dan disimpan dalam variabel `text`. Proses ekstraksi ini akan menganalisis teks dan mengidentifikasi kata-kata atau frasa-frasa yang dianggap penting atau relevan sebagai kata kunci. Pada baris terakhir menampilkan *print(keywords)*, berfungsi mencetak hasil ekstraksi kata kunci ke layar. Hasilnya akan berupa daftar kata kunci dengan skor relevansinya.

Dalam proses pengimplementasian ekstraksi kata kunci menggunakan algoritma *library* metode YAKE seperti diatas terdapat serangkaian langkah-langkah yang dijalankan oleh *library* ini guna menghasilkan kata kunci dan skor relevansinya dari publikasi dosen yang terdaftar pada SINTA terindeks Scopus. Berikut adalah beberapa tahapan yang terlibat dalam proses tersebut.

#### 1. Pra Pemrosesan Teks

Pada langkah pra pemrosesan teks metode YAKE membagi teks menjadi istilah individual setiap kali ruang kosong atau karakter khusus misal jeda baris, tanda kurung, koma, titik, dan lain lain. Berikut beberapa contoh judul publikasi dosen sebelum pemrosesan teks dan sesudah pemrosesan teks pada tabel 3.1.

Tabel 3. 1 Tabel Pra Pemrosesan teks

No.	Sebelum Pemrosesan Teks	Sesudah Pemrosesan Teks
1.	Deep Learning Model for Sentiment Analysis on Short Informal Texts.	1. Deep Learning Model 2. Sentiment Analysis 3. Short Informal Texts
2.	Implementation of 128 bits Camellia Algorithm for Cryptography in Digital Image	1. Bits Camellia Algorithm 2. Algorithm for Cryptography 3. Digital Image
3.	Comparison of Various Similarity Measures for Average Image Hash in Mobile Phone Application	1. Similarity Measures 2. Average Image Hash 3. Mobile Phone Application

No.	Sebelum Pemrosesan Teks	Sesudah Pemrosesan Teks
4.	Analysis of IPv6 jumbogram packages transmission using jumbo frame in mikrotik-based tunneling	1. Jumbogram packages transmission 2. jumbo frame 3. mikrotik-based tunneling
5.	Experimental Analysis of IPv6 Tunneling of Jumbo Frame Transmission using Mikrotik Routers	1. Jumbo Frame Transmission 2. Mikrotik Routers

## 2. Ekstraksi Fitur

Pada metode YAKE terdapat 5 fitur dalam proses ekstraksi kata kunci, fitur fitur tersebut yaitu *casing*, *Word Positional*, *Word Frequency*, *Word Relatedness to Context*, dan *Word Different Sentence*. Untuk rumus perhitungannya sebagai berikut.

$$casing(w) = \frac{\max(count(w \text{ is capital}), count(w \text{ is acronym}))}{1 + \log(count(w))} \quad (15)$$

$$position(w) = \log(\log(3 + \text{Median}(\text{Sen}(w)))) \quad (16)$$

$$frequency(w) = \frac{count \text{ of word } w}{mean(counts) + standard \ deviation(counts)} \quad (17)$$

$$relatedness(w) = 1 + (WR + WL) * \frac{count(w)}{\max count} + PL + PR \quad (18)$$

$$different(w) = \frac{number \ of \ sentences \ w \ occurs \ in}{total \ sentences} \quad (19)$$

## 3. Skor Istilah Individu

Setelah 5 fitur selesai dihitung, Langkah selanjutnya yaitu menggabungkannya menjadi skor tunggal  $S(w)$  atau pemberian skor istilah individu menggunakan rumus:

$$score(w) = \frac{d*b}{a + \left(\frac{c}{d}\right) + \left(\frac{e}{a}\right)} \quad (20)$$

Keterangan :

$a = casing$ ,  $b = position$ ,  $c = frequency$ ,  $d = relatedness$ ,  $e = different$

## 4. Kandidat Kata Kunci

Pada tahap kandidat kata kunci merupakan tahap hasil sementara kata kunci yang sudah melewati proses pra pemrosesan teks, ekstraksi fitur dan skor istilah

individu sebelum melewati beberapa proses lagi sesuai kebutuhan pada sistem. Pada penelitian ini kandidat kata kunci akan melewati proses deduplikasi data atau menghapus data yang sama dan proses *filtering keyword* untuk mengfilter kata kata yang tidak penting atau tidak layak masuk kedalam sistem.

#### 5. Deduplikasi Data

Pada tahap deduplikasi data metode *Yet Another Keyword Extractor* akan menghapus kata atau data yang hampir serupa atau serupa saat mengekstraksi frasa kata kunci. Contoh pada penelitian ini terdapat beberapa publikasi dosen yang sama persis judul publikasinya seperti pada tabel 3.2.

Tabel 3. 2 Tabel deduplikasi data

No.	Data 1	Data 2
1.	Comparison of Various Similarity Measures for Average Image Hash in Mobile Phone Application.	Comparison of various similarity measures for average image hash in mobile phone application.
2.	Root Cause Failure Analysis of Stator Winding Insulation failure on 6.2 MW hydropower generator.	Root Cause Failure Analysis of Stator Winding Insulation failure on 6.2 MW hydropower generator.
3.	The Performance of SISO in Wireless Open-Access Research Platform (WARP) Using QAM Modulation.	The Performance of SISO in Wireless Open-Access Research Platform (WARP) Using QAM Modulation.

Sesuai pada isi tabel 3.2 tentu saja akan menghasilkan kata kunci kepakaran yang sama sehingga diperlukan proses deduplikasi data atau menghapus data yang sama atau mirip. untuk menghilangkan kata duplikat maka proses berikut diterapkan.

Pertama, kata kunci diurutkan dalam urutan menaik berdasarkan skornya dan mempertahankan daftar kata kunci yang dipilih. Kemudian, untuk setiap kata kunci dalam daftar jika kata kunci memiliki jarak Levenshtein yang kecil dengan salah satu kata kunci yang dipilih sejauh ini, kata kunci tersebut akan dilewati. Jika tidak, kata kunci ditambahkan ke daftar kata kunci yang dipilih. Dengan demikian, daftar kata kunci yang dipilih berisi kata kunci akhir yang telah dihapus duplikatnya.

### 3.2.3 Filtering Keyword

Pada penerapannya hasil data kata kunci ini juga terdapat beberapa kata yang tidak sesuai untuk dimasukkan kedalam sistem kepakaran dosen, maka diperlukan *keyword filtering* untuk mengfilter atau menghapus kata yang sekiranya tidak layak atau tidak sesuai dengan kepakaran dosen untuk dimasukkan kedalam sistem. Untuk *source code keyword filtering* sebagai berikut.

```

1. def read_keywords():
2.     with open('keywords.txt', 'r') as f:
3.         keywords = [line.strip() for line in f]
4.     return keywords
5. def is_filtered_keyword(word, keywords):
6.     return word in keywords

```

*source code keyword filtering* berfungsi untuk mengfilter kata kunci yang tidak sesuai dengan kepakaran dosen. Pada kode diatas membuka file *keywords.txt* yang sudah dibuat sebelumnya, pada file tersebut berisi kata kunci yang akan difilter dan tidak dimasukkan kedalam sistem karena tidak sesuai dengan kepakaran dosen.

Kemudian pada kode *is\_filtered\_keyword(word, keywords)*: kode ini berfungsi menerima kata yang ingin diperiksa (*word*) dan daftar kata kunci (*keywords*), lalu memeriksa apakah kata tersebut ada dalam daftar kata kunci yang ada didalam file *keywords.txt*. *return word in keywords*: Mengembalikan *true* jika *word* ditemukan dalam *keywords*, dan *False* jika tidak ditemukan. Kemudian dilanjutkan dengan *source code* untuk memanggilnya.

```

1. filtered_keywords = [(keyword, score) for keyword, score in
    YAKE_keywords
2. if not is_filtered_keyword(keyword, keywords)]
3. if not filtered_keywords:
4.     print("No keywords extracted.")
5. else:
6.     print("Keywords:")
7.     print(filtered_keywords)

```

Menampilkan *Source code* memanggil *filtering keyword*, inti pada *Source code* akan menampilkan kata kunci yang tidak masuk kedalam *keyword filtering*



atau file keywords.txt yang dibuat sebelumnya. Jika kata tersebut ada didalam file maka kata itu akan difilter atau dihapus. *Keyword* yang tidak sesuai ini perlu difilter karena untuk menghindari kesalahan pada data kepakaran dosen.

Pada penelitian ini terdapat banyak kata yang tidak penting atau tidak layak masuk ke data kepakaran dosen seperti pada tabel 3.3 yang akan menjelaskan beberapa contoh *keyword* yang tidak layak masuk dan alasannya.

Tabel 3. 3 Daftar *filtering keyword*

No.	Kata kunci ( <i>keywords</i> )	Keterangan
1.	<i>Process, Impact, Push, Improve</i>	Kata tersebut masuk kedalam jenis kata kerja bukan sebuah <i>keyword</i> yang dicari.
2.	<i>Agung campus, Abdullah Bin, Bin Nuh works, journals in Indonesia, based on knowledge.</i>	Itu merupakan penggalan judul dari data publikasi dosen bukan merupakan <i>keyword</i> yang mempresentasikan sebuah kepakaran.
3.	<i>Study, Programs, Strategy, Capability, ability</i>	Semua kata ini adalah kata benda yang merujuk pada konsep, keadaan, atau objek tertentu, sehingga tidak termasuk kedalam <i>keyword</i> kepakaran.

Pada tabel 3.3 merupakan contoh dari beberapa kata yang dihasilkan oleh metode YAKE dalam pengimplementasiannya kedalam database publikasi dosen. Pada penelitian ini terdapat total 127 *keywords* yang masuk kedalam daftar *keyword filtering* untuk difilter atau dihapus.

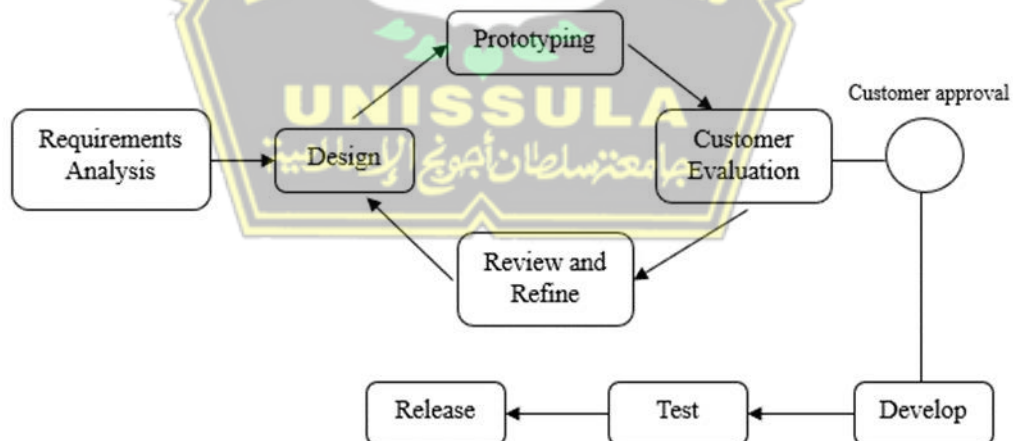
Setelah tahapan-tahapan diatas sudah dilakukan, maka metode YAKE akan memunculkan pemeringkatannya. Pada metode ini semakin kecil skor yang dihasilkan semakin pakar atau semakin tinggi pembobotan kata kunci tersebut karena metode ini beranggapan semakin jarang kata kunci tersebut keluar semakin unik kata kunci tersebut yang secara tidak langsung diidentifikasi semakin penting kata kunci tersebut.

### 3.3 Metode Pengembangan Sistem

Metode pengembangan sistem yang diterapkan dalam penelitian ini adalah metode *Prototype*. Metode *Prototype* adalah suatu pendekatan dalam pengembangan perangkat lunak di mana sebuah *prototype* dibangun, diuji, dan kemudian direvisi secara berulang hingga mencapai hasil yang diinginkan. Pendekatan ini memungkinkan penggambaran awal sistem yang memungkinkan *user* dalam hal ini tim pengembang SINTA memiliki gambaran yang lebih jelas mengenai sistem yang akan dikembangkan oleh tim pengembang yaitu peneliti.

Dengan metode *prototype*, pengembang dan pengguna dapat berinteraksi melalui *prototype* sistem yang dibangun. *Prototype* ini bertindak sebagai jembatan antara tim pengembang dan pengguna dalam memahami dan menggambarkan kebutuhan sistem yang akan dibangun.

Keberhasilan pembuatan *prototype* sangat ditentukan oleh definisi aturan yang jelas pada tahap awal. Artinya, baik pengembang maupun pengguna harus memiliki pemahaman yang seragam bahwa tujuan utama pembangunan *prototype* adalah untuk merumuskan kebutuhan awal sistem dengan lebih baik. Untuk alur dalam metode *prototype* seperti pada gambar 3.3.



Gambar 3. 3 Alur pengembangan sistem *prototype*

Pada gambar 3. 3 merupakan alur pengembangan sistem *prototype* yang akan digunakan dalam pengembangan sistem kepakaran dosen ini. Model *prototype* yang digunakan 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.

### 3.3.1 *Requirements analysis*

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 yang terdiri dari Bapak Imam Much Ibnu Subroto, ST, M.Sc, Ph.D, dan Bapak Sam Farisa Chaerul Haviana, ST, M.Kom. Tujuan komunikasi ini adalah untuk mendiskusikan dan menggambarkan analisa kebutuhan sistem dan rencana pengembangan sistem secara lebih rinci dan menyeluruh.

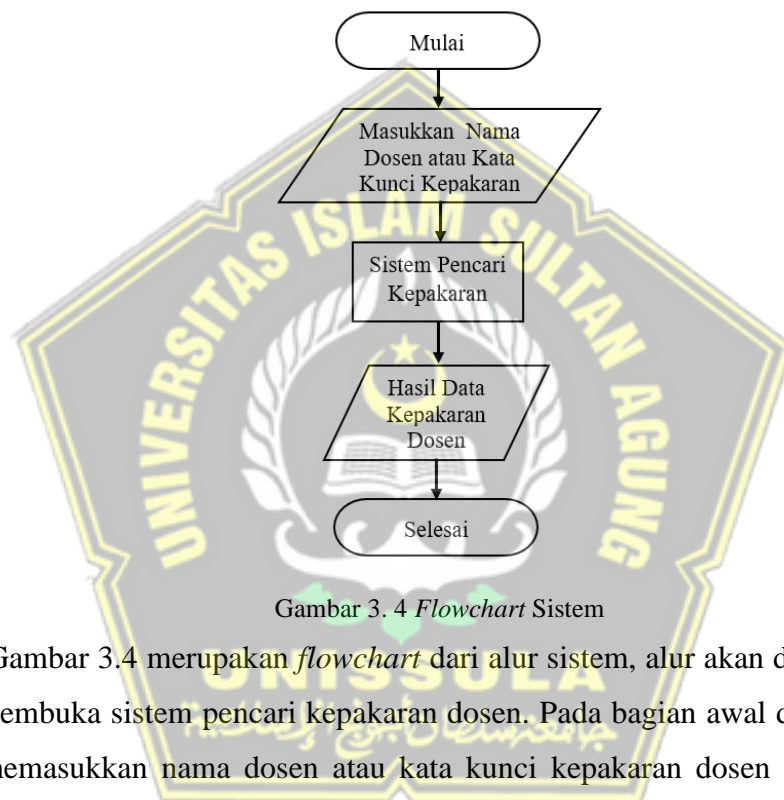
Hal ini juga bertujuan untuk memastikan bahwa persepsi dan pemahaman mengenai analisa kebutuhan sistem dan pengembangan sistem ini dapat diselaraskan di antara semua pihak yang terlibat. Hasil dari tahapan ini peneliti mendapatkan data dari tim pengembang SINTA berupa data judul publikasi dosen yang terindeks scopus. Pada tahap ini juga peneliti mendapatkan gambaran mengenai sistem yang akan dikembangkan yaitu sistem kepakaran dosen di lingkungan Universitas Islam Sultan Agung berdasarkan publikasi pada SINTA terindeks scopus.

Selain itu, Pada tahap ini peneliti menganalisa kebutuhan sistem yang dibutuhkan untuk mengembangkan sistem ini. Adapun kebutuhan sistem atau *tools* yang dibutuhkan antara lain meliputi *python* sebagai bahasa pemrograman untuk mengolah data yang didapatkan, *jupyter notebook* sebagai platform pengolahan data yang diolah menjadi ekstraksi kata kunci, MySQL sebagai platform untuk data *warehousing* atau gudang data, dan *Visual Studio Code* sebagai platform pembuatan sistem untuk menampilkan data yang sudah diekstraksi dengan menggunakan bahasa pemrograman PHP dan HTML.

### 3.3.2 Desain

Tahap kedua dalam pengembangan sistem terdiri dari desain awal atau desain cepat. Selama tahap ini, peneliti menggambarkan secara cepat tentang konsep dasar dari sistem kepakaran dosen agar pengguna dapat memahami gambaran awal dari sistem. Adapun desain dalam sistem kepakaran dosen mencakup beberapa aspek sebagai berikut.

#### 1. *Flowchart*

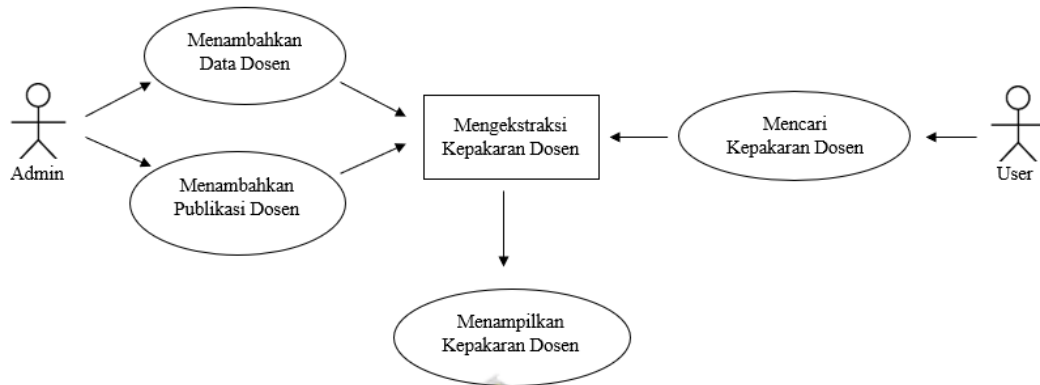


Gambar 3. 4 *Flowchart* Sistem

Gambar 3.4 merupakan *flowchart* dari alur sistem, alur akan diawali dengan *user* membuka sistem pencari kepakaran dosen. Pada bagian awal dari sistem ini, *user* memasukkan nama dosen atau kata kunci kepakaran dosen sesuai dengan kebutuhan kedalam sistem. Setelah itu sistem akan memprosesnya dan sistem akan menampilkan hasil data kepakaran dosen sebagai outputnya, maka alur proses dari sistem ini telah selesai.



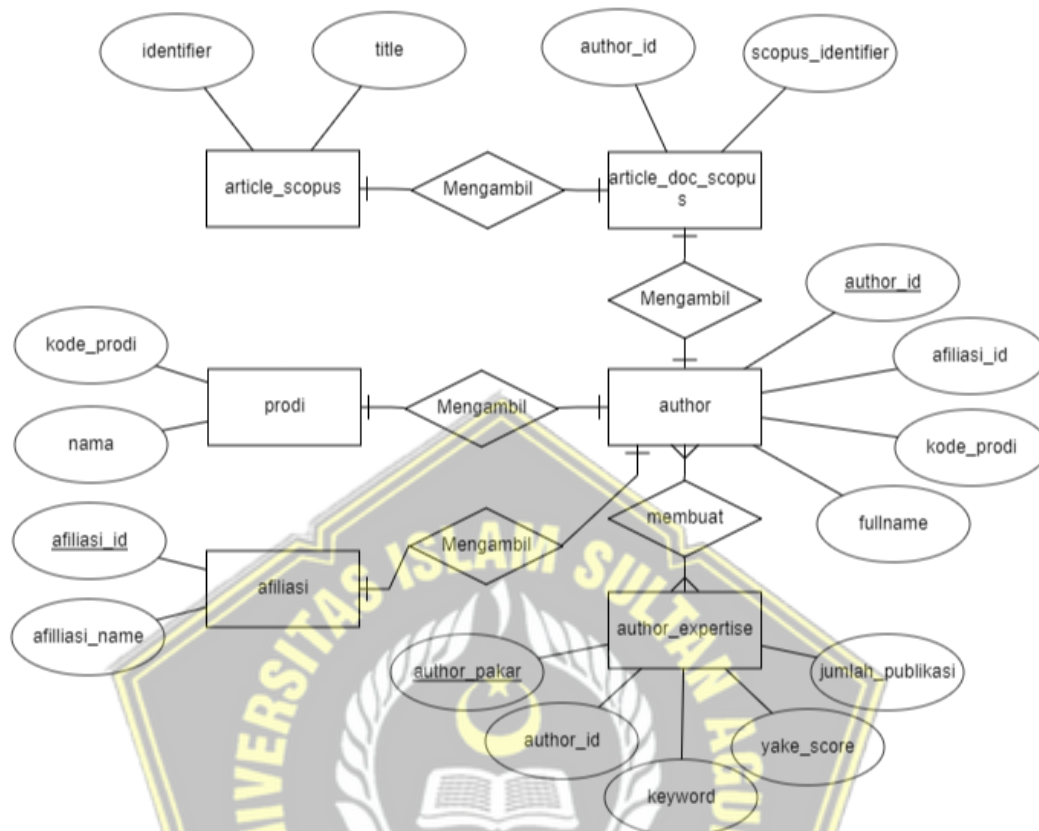
## 2. Use case



Gambar 3.5 Use Case Sistem

Dalam pengembangan sistem *use case* akan dirancang seperti gambar 3.5 dimana admin akan menambahkan data dosen dan publikasi dosen dari kepakaran dosen yang nantinya akan diproses oleh sistem yang akan menghasilkan suatu kepakaran dosen dalam bidang tertentu, tetapi penambahan data ini dilakukan oleh admin SINTA karena peneliti mendapat data dari SINTA. Sedangkan untuk *user* akan mencari kepakaran dosen pada sistem kemudian diproses oleh sistem tersebut dan akan menghasilkan data kepakaran dosen yang dicari oleh *user* atau pengguna.

### 3. Entity Relationship Diagram (ERD)



Gambar 3. 6 Gambar ERD

Pada *entity relationship diagram* (ERD), tergambar entitas dan hubungan data yang digunakan dalam sistem ini. ERD ini mencakup beberapa tabel, yaitu "author," "author\_expertise," "afiliasi," "prodi," "article\_doc\_scopus," dan "article\_scopus." Tabel "article\_scopus" memiliki keterhubungan dengan "article\_doc\_scopus" menggunakan hubungan kolom *identifier* dan kolom *scopus\_identifier* untuk mengambil kolom *title* atau judul publikasi.

Kemudian tabel "article\_doc\_scopus" dihubungkan dengan tabel "author" menggunakan kolom "author\_id" untuk mengambil publikasi dosen tersebut. Selanjutnya, terdapat relasi antara tabel "prodi" dan tabel "author" menggunakan kolom *kode\_prodi* yang memungkinkan pencarian nama prodi untuk seorang dosen. Selain itu, terdapat juga relasi antara tabel "afiliasi" dan tabel "author" menggunakan kolom *afiliasi\_id* untuk mencari nama universitas tempat seorang dosen berafiliasi.

Dalam tahap terakhir, terdapat proses pembuatan tabel baru yang disebut tabel "*author\_expertise*." Tabel ini dibuat dengan memanfaatkan *library* metode YAKE serta informasi dari hubungan antara tabel-tabel yang telah disebutkan sebelumnya. Hasilnya adalah tabel "*author\_expertise*" yang berisi kolom-kolom seperti "*author\_pakar*," "*author\_id*," "*keyword*," "*yake\_score*," dan "jumlah publikasi."

#### 4. Tabel

Dalam tahap pembuatan tabel penerapannya yaitu menginput data kedalam database *phpMyAdmin* untuk nantinya akan diproses oleh sistem *front end* nya, Pada sistem ini terdapat 6 tabel yaitu "*article\_scopus*," "*article\_doc\_scopus*," "*author*," "prodi," "afiliasi," dan "*author\_expertise*." Untuk detail tabel-tabel tersebut seperti pada tabel berikut :

Tabel 3. 4 tabel *article\_scopus*

Kolom	Tipe Data	Null	Default	Extra
identifier	bigint(30)	No	None	
title	varchar(50)	No	None	

Pada tabel 3.4 merupakan tabel *article\_scopus* yang berisikan kolom *identifier* dan kolom *title*. Terdapat juga tabel *article\_doc\_scopus* seperti pada tabel 3.5 berikut.

Tabel 3. 5 Tabel *article\_doc\_scopus*

Kolom	Tipe Data	Null	Default	Extra
author_id	bigint(20)	No	None	
scopus_identifier	bigint(20)	No	None	

Pada tabel 3.5 merupakan tabel *article\_doc\_scopus* yang berisikan data kolom *author\_id* dan *scopus\_identifier* yang digunakan dalam sistem ini. Terdapat juga tabel *author* seperti pada tabel 3.6

Tabel 3. 6 Tabel *author*

Kolom	Tipe Data	Null	Default	Extra
author_id	bigint(20)	No	None	Auto_increment, Primary Key
afiliasi_id		No	None	

Kolom	Tipe Data	Null	Default	Extra
kode_prodi	varchar(20)	No	None	
fullname	varchar(100)	No	None	

Pada tabel 3.6 merupakan tabel *author* yang berisikan data kolom *author\_id*, *afiliasi\_id*, *kode\_prodi* dan *fullname* yang digunakan dalam sistem ini. Terdapat juga tabel *prodi* seperti pada tabel 3.7.

Tabel 3. 7 Tabel *prodi*

Kolom	Tipe Data	Null	Default	Extra
kode_prodi	varchar(10)	No	None	
nama	varchar(50)	No	None	

Pada tabel 3.7 merupakan tabel *prodi* yang berisikan data kolom *kode\_prodi* dan kolom *nama* yang digunakan dalam sistem ini. Terdapat juga tabel *afiliasi* seperti pada tabel 3.8.

Tabel 3. 8 Tabel *afiliasi*

Kolom	Tipe Data	Null	Default	Extra
afiliasi_id	int(11)	No	None	Auto_increment, Primary Key
afiliasi_name	varchar(150)	No	None	

Pada tabel 3.8 merupakan tabel *afiliasi* yang berisikan data kolom *afiliasi\_id* dan *afiliasi\_name* yang digunakan dalam sistem ini. Terdapat juga tabel *author\_expertise* seperti pada tabel 3.9.

Tabel 3. 9 Tabel *author\_expertise*

Kolom	Tipe Data	Null	Default	Extra
author_pakar	bigint(20)	No	None	Auto_increment, Primary Key
author_id	Int(11)	Yes	Null	
Keyword	Varchar(255)	Yes	Null	
yake_score	Float	Yes	Null	
jumlah_publicasi	Int(11)	Yes	Null	

Pada tabel 3.9 merupakan rancangan tabel baru yang nantinya akan berisi tentang database *keywords* dan diberi nama tabel *author\_expertise*.

## 5. Perancangan *User Interface*

Desain awal juga mencakup pembuatan *User Interface* yang sederhana. *User Interface* pada sistem kepakaran dosen ini mungkin belum sepenuhnya sempurna, tetapi mencerminkan cara pengguna akan berinteraksi dengan sistem. Pada tahap perencanaan *user interface* ini akan menampilkan desain *mockup* yang akan dibuat pada sistem. Berikut merupakan perancangan *user interface* dari sistem identifikasi kepakaran dosen:

### a. Halaman *Home*

Halaman *Home* atau halaman utama pada website penelitian ini menampilkan tombol untuk *search* atau mencari data kepakaran dosen secara spesifik, input pencarian bisa menggunakan kata kunci kepakaran atau bisa juga menggunakan nama dosen. Rancangan *user interface* bisa dilihat dalam gambar 3.7.



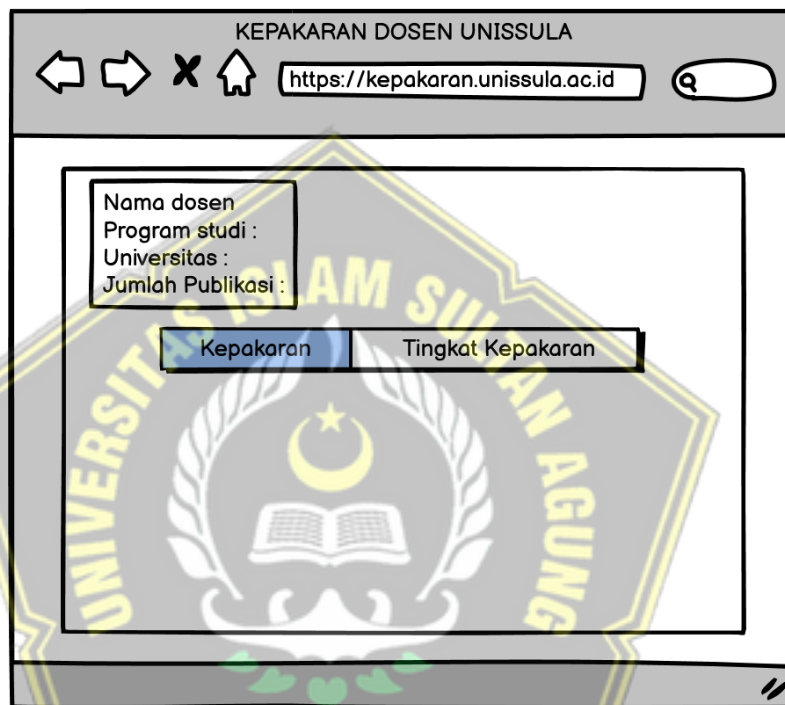
Gambar 3. 7 Rancangan *user interface home*

Pada gambar 3.7 merupakan rancangan *user interface* yang akan dikembangkan dalam penelitian ini, *user interface* dibuat secara *simple* untuk mempermudah *user* atau pengguna dalam mengoperasikannya, akan tetapi tetap terlihat menarik untuk memuaskan para *user* dalam penggunaannya.



b. Rancangan Hasil Berdasarkan Nama Dosen

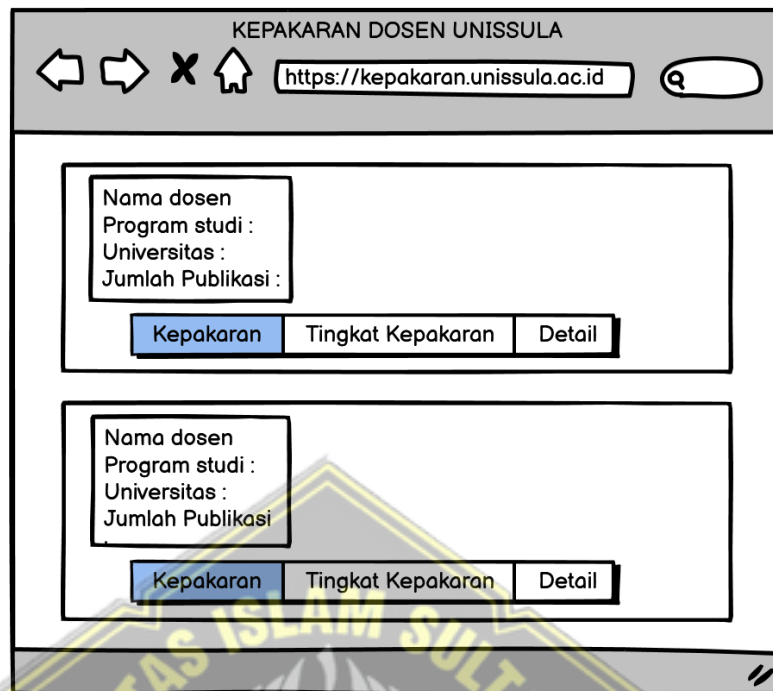
Pada tahap ini akan menampilkan rancangan hasil tampilan berdasarkan nama dosen yang dicari pada sistem, pada tampilan ini akan menampilkan nama dosen, program studi, universitas, dan jumlah publikasi pada dosen yang terkait dan terdapat tabel yang berisi informasi kepakaran dan tingkat kepakaran, seperti pada gambar 3.8.



Gambar 3. 8 Rancangan hasil berdasarkan nama dosen

c. Tampilan Hasil Berdasarkan *Keyword*

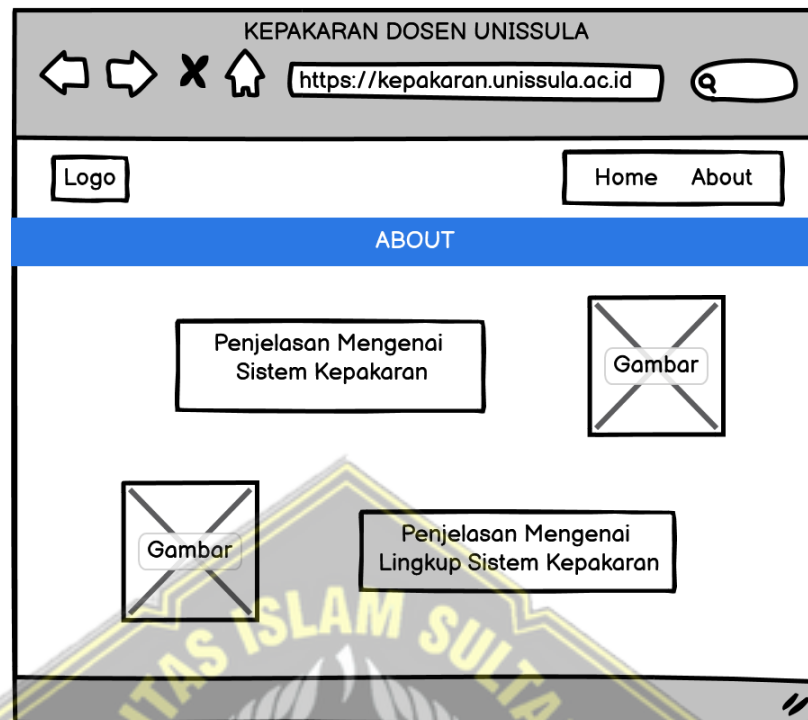
Pada tahap ini akan menampilkan rancangan hasil tampilan berdasarkan *keyword* yang dicari pada sistem, pada tampilan ini akan menampilkan nama dosen, program studi, universitas, dan jumlah publikasi pada dosen yang terkait pada *keyword* yang dicari, diurutkan berdasarkan skor kepakarannya dan terdapat tabel yang berisi informasi kepakaran, tingkat kepakaran, dan detail. Pada kolom detail ini terdapat tombol jika diklik akan menampilkan informasi secara lengkap mengenai dosen yang terkait. seperti pada gambar 3.9.



Gambar 3. 9 Rancangan hasil berdasarkan *keyword*

d. Halaman *About*

Pada halaman *about* akan menampilkan informasi secara lengkap mengenai sistem kepakaran dosen secara spesifik dan lingkup kepakaran dosen yang dicari. Rancangan *user interface* untuk halaman *about* bisa dilihat pada gambar 3.10.



Gambar 3. 10 Rancangan *user interface about*

Pada gambar 3.10 merupakan tampilan rancangan *user interface about* yang akan dikembangkan dalam sistem ini, rancangan akan berisi gambar dan penjelasannya untuk membuat *user* atau pengguna tidak bosan ketika membaca penjelasan tentang sistem kepakaran dan lingkup sistem kepakaran dosen tersebut.

### 3.3.3 *Prototyping*

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

1. Pengumpulan Data: Data akan diambil dari SINTA yang terindeks di scopus. Data ini diperoleh dari tim pengembang SINTA, mencakup publikasi para dosen, khususnya dari lingkungan Universitas Islam Sultan Agung Semarang, yang relevan dengan fokus penelitian ini.
2. Pengolahan Data: Setelah data terkumpul, dilakukan pengolahan menggunakan bahasa pemrograman *Python*. Pemilihan *Python* didasarkan

pada kemampuannya dalam melakukan analisis data secara efektif.

3. Analisis Kepakaran: Dalam pengolahan data, metode *Yet Another Keyword Extractor* (YAKE) akan digunakan untuk menentukan kata kunci serta skor relevansi dari publikasi dosen yang dilakukan pada jupyter notebook. Hal ini akan membantu mengidentifikasi kepakaran para dosen.
4. Setelah mendapatkan data kata kunci dan skor kepakarannya, data akan disimpan kedalam database baru pada MySQL.
5. Pengembangan Sistem: Data kepakaran yang dihasilkan akan menjadi dasar bagi pengembangan sistem. Sistem ini akan berbasis situs web agar mudah diakses oleh pengguna. Pengembangan akan dilakukan menggunakan platform *Visual Studio Code*. Penggunaan platform ini didasarkan pada kemampuannya dalam pengembangan web dan mendukung banyak bahasa pemrograman populer sehingga memudahkan dalam penggunaannya.
6. Implementasi Logika Sistem: Logika utama dari sistem akan diimplementasikan dengan menggunakan bahasa pemrograman PHP. Pendekatan ini dipilih guna memastikan implementasi sistem berjalan secara optimal dan efisien sesuai dengan konteks penelitian.
7. Antarmuka Pengguna: Antarmuka pengguna akan diwujudkan menggunakan *Bootstrap* CSS, sehingga pengguna akan merasakan tampilan yang responsif dan menarik saat menggunakan sistem.

Dengan mengikuti langkah-langkah *prototyping* ini, diharapkan pembentukan *prototype* dapat berjalan lancar sesuai dengan tujuan dan kebutuhan penelitian yang dilakukan.

#### **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 ada kekurangan dalam *prototype*. Jika *prototype* sudah sesuai dengan kebutuhan *user* maka akan lanjut ke tahapan *customer approval*, tetapi jika belum sesuai maka akan dilanjutkan ke tahap *review and refine*.

### **3.3.5 Review and 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*. Tetapi pada proses penelitian ini hanya melakukan sekali iterasi untuk menghasilkan *prototype*.

Pada proses iterasi tersebut ketika *prototype* sudah sesuai dengan keinginan *user* maka dilakukan tahapan *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 sebelumnya. Proses pada tahapan ini meliputi implementasi pengolahan data, kode sistem dan pengembangan website dibuat berdasarkan *prototype* yang sudah ada sebelumnya.

### **3.3.7 Test**

Pada tahap *test* peneliti merancang rencana pengujian pada sistem kepakaran dosen berupa fungsionalitas sistem yang akan digunakan untuk menguji sistem secara menyeluruh. Rencana pengujian sistem ini berguna untuk membantu memastikan bahwa sistem dapat berfungsi sesuai dengan tujuan dan kebutuhan *user*. Pada tahap ini rencana pengujian akan menggunakan metode *blackbox testing*. *Blackbox testing* adalah metode pengujian perangkat lunak yang dilakukan tanpa memerhatikan struktur internal atau logika kode pada sumber programnya. Melainkan, pengujian difokuskan pada input yang diberikan dan output yang



dihasilkan oleh perangkat lunak tanpa memperhatikan bagaimana perangkat lunak tersebut bekerja di dalamnya. Rencana pengujian ini meliputi beberapa hal sebagai berikut.

1. Input atau pengujian yang dilakukan dalam pengujian ini berupa pencarian kepakaran dosen berdasarkan nama lengkap dosen, kata kunci nama dosen, kata kunci kepakaran dosen, detail dosen dan menu *about*.
2. Kriteria keberhasilan, pada tahap ini peneliti menilai apakah sistem telah berhasil dalam pengujian atau tidak. Hal ini dapat ditentukan dari output yang dihasilkan apakah sesuai dengan hasil yang diharapkan pada proses pengujian sistem.

### 3.3.8 Release

Dalam pengembangan sistem *release* merupakan tahapan akhir yang bertujuan untuk dirilis atau diperkenalkan kepada pengguna akhir. *Release* merupakan titik penting dalam siklus pengembangan sistem karena jika sudah sampai tahap ini maka menandakan sistem siap digunakan, diuji, dan diterapkan oleh pengguna akhir.

### 3.4 Rancangan Perhitungan MAP@K

Dalam penelitian ini perhitungan *Mean Average Precision@K* (MAP@K) digunakan untuk mencari kata kunci maksimal setiap dosen sehingga hasil kata kunci lebih relevan untuk dijadikan data kepakaran dosen. Rumus metode ini sebagai berikut :

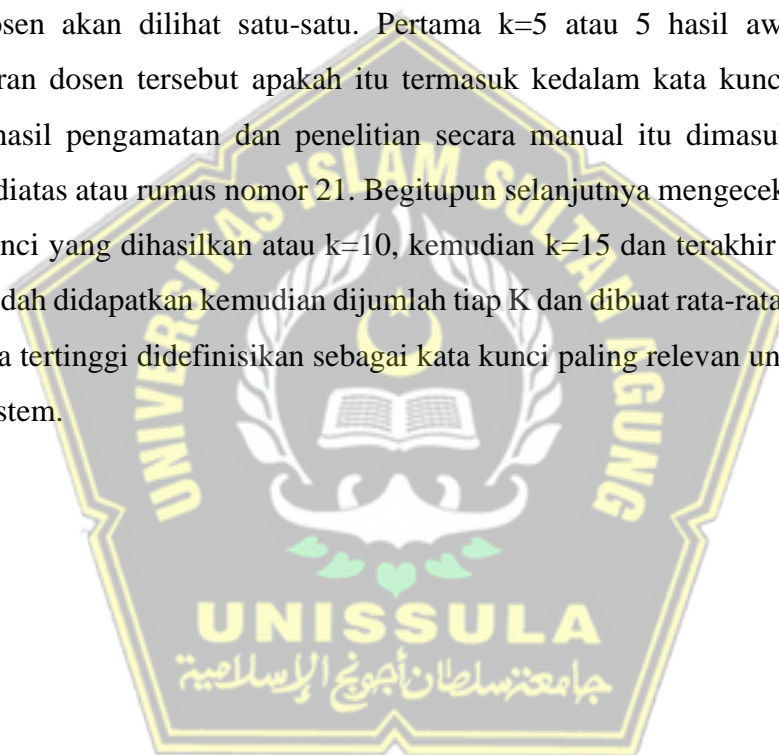
$$Precision = \frac{| \{relevant\ documents\} \cap \{retrieved\ documents\} |}{| \{retrieved\ documents\} |} \quad (21)$$

Penjelasan rumus :

$| \{relevant\ documents\} \cap \{retrieved\ documents\} |$  ini mengacu pada jumlah dokumen yang secara bersamaan termasuk dalam kedua set dokumen, yaitu dokumen relevan dan dokumen yang diambil. Operasi  $\cap$  (irisan) menyatakan elemen-elemen yang dimiliki oleh kedua himpunan tersebut kemudian dibagi  $| \{retrieved\ documents\} |$  ini merupakan jumlah total dokumen yang berhasil diambil atau ditemukan oleh sistem.

Dalam pehitungannya akan menggunakan 5 dosen sebagai sampel perhitungan, dosen tersebut yaitu Heru Sulisty, Pratikso, Sri Hartono, Azizah Hikma Safitri, dan Gunawan Dewantoro. *relevant documents* didapatkan dari penelitian dan pengamatan secara manual dengan memperhatikan kata kunci yang dihasilkan tersebut apakah relevan dengan kepakaran seorang dosen atau tidak. Sedangkan *retrieved documents* yang diambil dalam perhitungan ini yaitu  $k=5$ ,  $k=10$ ,  $k=15$ , dan  $k=20$ .

Sistematika perhitungannya pada hasil kata kunci kepakaran yang dihasilkan tiap dosen akan dilihat satu-satu. Pertama  $k=5$  atau 5 hasil awal kata kunci kepakaran dosen tersebut apakah itu termasuk kedalam kata kunci relevan atau tidak, hasil pengamatan dan penelitian secara manual itu dimasukkan kedalam rumus diatas atau rumus nomor 21. Begitupun selanjutnya mengecek 10 hasil awal kata kunci yang dihasilkan atau  $k=10$ , kemudian  $k=15$  dan terakhir  $k=20$ . Setelah hasil sudah didapatkan kemudian dijumlah tiap  $K$  dan dibuat rata-rata. Hasil dengan rata-rata tertinggi didefinisikan sebagai kata kunci paling relevan untuk digunakan pada sistem.



## BAB IV

### HASIL DAN ANALISIS PENELITIAN

#### 4.1 Hasil Ekstraksi Kepakaran

Pada tahap ini akan menampilkan hasil ekstraksi kepakaran dalam bentuk kata kunci dari publikasi dosen Universitas Islam Sultan Agung yang terdaftar pada SINTA terindeks scopus menggunakan *library* metode *Yet Another Keyword Extractor*. Hasil dari ekstraksi kepakaran berupa daftar kata kunci beserta skor relevansinya seperti pada gambar 4.1.

```
[('Jumbo Frame Transmission', 0.005534696449558763), ('jumbogram packages transmission', 0.008204342507138102), ('jumbo frame', 0.015265134741917267), ('Frame Transmission', 0.030530269483834534), ('Transmission using Mikrotik', 0.03791822816957832), ('packages transmission', 0.038994085285395254), ('mikrotik-based tunneling', 0.04204967756599691), ('Mikrotik Routers', 0.04403547002840346), ('jumbogram packages', 0.04484533716784565), ('Business System Model', 0.0487582473098678), ('System Modeling Approach', 0.05824125372195508), ('Model', 0.06296912191751827), ('transmission using jumbo', 0.07079799374867275), ('Position Controller Model', 0.07204269221074766), ('System', 0.07251657041482558), ('performance', 0.07699413395096077), ('Analysis', 0.0780342045881824), ('Furniture in Indonesia', 0.07882833587274148), ('System Model', 0.0793669869726929), ('transmission', 0.08606071439116615)]
```

Gambar 4. 1 Hasil ekstraksi kepakaran

Pada gambar 4.1 menampilkan contoh hasil dari proses ekstraksi kepakaran dengan menggunakan *library* metode YAKE. Hasil ekstraksi tersebut nantinya yang akan digunakan pada sistem seperti pada gambar 4.2.



Kepakaran	Tingkat Kepakaran
Jumbo Frame Transmission	98.74%
jumbogram packages transmission	98.11%
jumbo frame	96.70%

Gambar 4. 2 Hasil kepakaran pada sistem

Dalam gambar 4.2 menampilkan sistem kepakaran dosen, pada tampilan tersebut terdapat profil dosen yang berisi nama lengkap, program studi, universitas, jumlah publikasi dan terdapat tabel kepakaran dosen beserta tingkat kepakarannya yang diimplementasikan menggunakan gambar lingkaran sebagai representasi persen, jika gambar lingkaran penuh maka tingkat kepakaran dosen pada bidang tersebut 100% atau maksimal, dibawah gambar lingkaran juga dilengkapi keterangan persen untuk memudahkan *user* dalam memahami tingkat kepakaran dosen.

Tingkat kepakaran dosen yang berupa persen didapatkan dari perhitungan YAKE *score* yang didapatkan masing-masing kata kunci dengan pemanfaatan perhitungan *library* YAKE dan score tertinggi yaitu 0.511057 sebagai pembagi untuk setiap perhitungan kata kunci, sehingga akan didapatkan nilai persen tiap kata kunci yang dihasilkan.

Jumlah kata kunci yang dihasilkan setiap dosen berbeda beda tergantung banyaknya jumlah publikasi dosen tersebut dan hasil ekstraksi *library* YAKE terhadap judul publikasi dosen yang bersangkutan. Jumlah data dosen yang terdapat dalam sistem ini yaitu 182 dosen dengan total publikasi sebanyak 728 judul publikasi dosen pada SINTA terindeks scopus.

Hasil kepakaran dari sistem ini telah diurutkan sesuai dengan prinsip implementasi dari metode YAKE di mana kata kunci yang memiliki skor lebih rendah diberikan bobot yang lebih tinggi. Dengan kata lain, kata kunci dengan skor relevansi atau skor kepakaran yang lebih kecil akan ditampilkan lebih awal sedangkan kata kunci yang memiliki skor relevansi yang lebih tinggi, atau kata-kata yang muncul secara berulang, akan diberikan bobot yang lebih rendah dan ditampilkan diakhir.

#### **4.2 Hasil Implementasi Sistem**

Pada tahap implementasi sistem akan menampilkan tampilan kepada *user* atau biasa disebut *user interface*. Pada sistem ini akan terbagi menjadi 2 tampilan utama yaitu *home* dan *about*. Pada tampilan *home* akan berisi menu pencarian kepakaran dosen berdasarkan nama dosen atau kepakaran dosen yang dicari dan pada tampilan *about* akan berisi informasi mengenai sistem dan lingkup pencarian

sistem kepakaran dosen tersebut. Seperti pada gambar 4.3.



Gambar 4. 3 Tampilan menu *home*

Pada gambar 4.3 merupakan tampilan *home* atau tampilan utama pada sistem ini. Pada tampilan ini terdapat menu untuk memilih opsi pencarian bisa berdasarkan nama dosen atau kepakaran dosen, disebelahnya ada menu memasukan kata yang terkait opsi pencarian, kemudian ada juga tombol cari untuk menemukan hasil dari kata yang dimasukkan pada kolom pencarian mengenai kepakaran dosen yang ada di Universitas Islam Sultan Agung Semarang. Untuk tampilan hasil pencarian berdasarkan nama dosen seperti pada gambar 4.4.

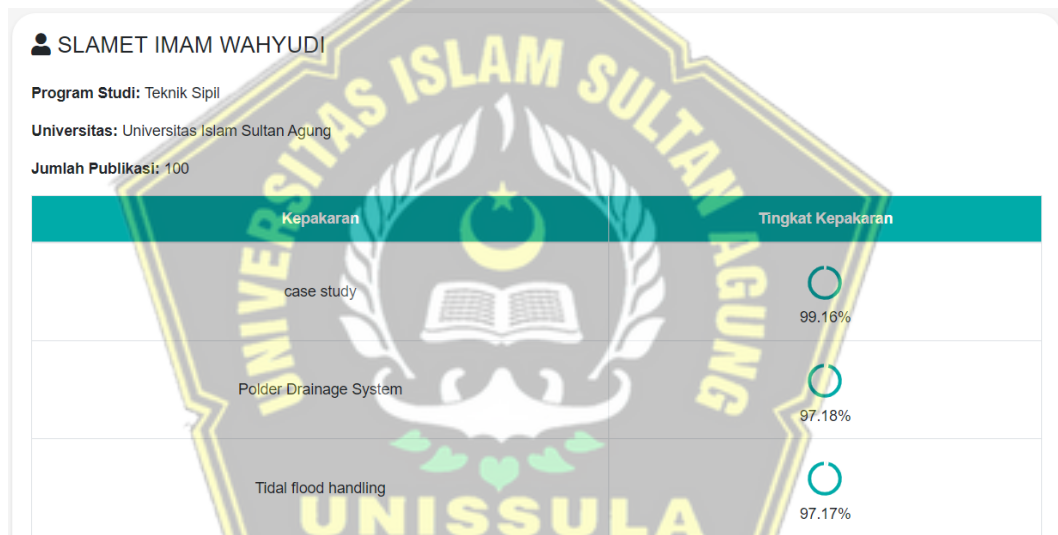
Kepakaran	Tingkat Kepakaran
Jumbo Frame Transmission	98.74%
jumbogram packages transmission	98.11%
jumbo frame	96.70%

Gambar 4. 4 Tampilan sistem berdasarkan pencarian nama dosen

Gambar 4.4 menampilkan hasil kepakaran berdasarkan pencarian nama dosen. Pada contoh tampilan tersebut, pencarian dilakukan dengan menggunakan



nama "Imam Much Ibnu Subroto." Hasil tampilan menampilkan informasi terkait dosen tersebut, termasuk nama lengkap, program studi, universitas, serta jumlah publikasi. Terdapat juga tabel yang memuat informasi mengenai kepakaran dosen dan tingkat kepakarannya yang ditampilkan berupa *chart* berbentuk lingkaran. Pada sistem ini, pencarian berdasarkan nama dosen juga mampu untuk menghasilkan lebih dari satu hasil. Sebagai contoh, ketika melakukan pencarian dengan kata kunci "imam," sistem akan menampilkan beberapa nama dosen yang mengandung kata "imam." Di UNISSULA, terdapat lebih dari satu nama dosen dengan kata "imam," sehingga sistem akan menampilkan semua nama dosen yang sesuai. Hal ini dapat dilihat seperti pada gambar 4.5.



Gambar 4. 5 Tampilan pencarian nama dosen imam

Gambar 4.5 merupakan tampilan pencarian nama dosen imam dengan menampilkan nama dosen slamet imam wahyudi, selain nama tersebut sistem juga menampilkan nama lain yaitu imam kusmaryono seperti pada tampilan gambar 4.6.

**IMAM KUSMARYONO**

**Program Studi:** Pendidikan Matematika  
**Universitas:** Universitas Islam Sultan Agung  
**Jumlah Publikasi:** 12

Kepakaran	Tingkat Kepakaran
Social Science Research	98.30%
Likert Scale Education	98.13%
students' mathematical knowledge	97.34%

Gambar 4. 6 Tampilan pencarian nama dosen imam 2

Berdasarkan tampilan gambar 4.6 akan memudahkan pengguna untuk mencari nama dosen dengan tidak mencarinya berdasarkan nama lengkap dosen, melainkan berdasarkan kata kunci nama dosen tersebut. Selain pencarian berdasarkan nama dosen, sistem ini juga memiliki kemampuan untuk melakukan pencarian berdasarkan kepakaran dosen. Tampilan pencarian berdasarkan kepakaran dosen dapat dilihat dalam gambar 4.7.

**TRI WIKANINGRUM**

**Program Studi:** Manajemen  
**Universitas:** Universitas Islam Sultan Agung  
**Jumlah Publikasi:** 17

Kepakaran	Tingkat Kepakaran	Detail
Support Knowledge Management	99.79%	<a href="#">Detail</a>

**NURHIDAYATI**

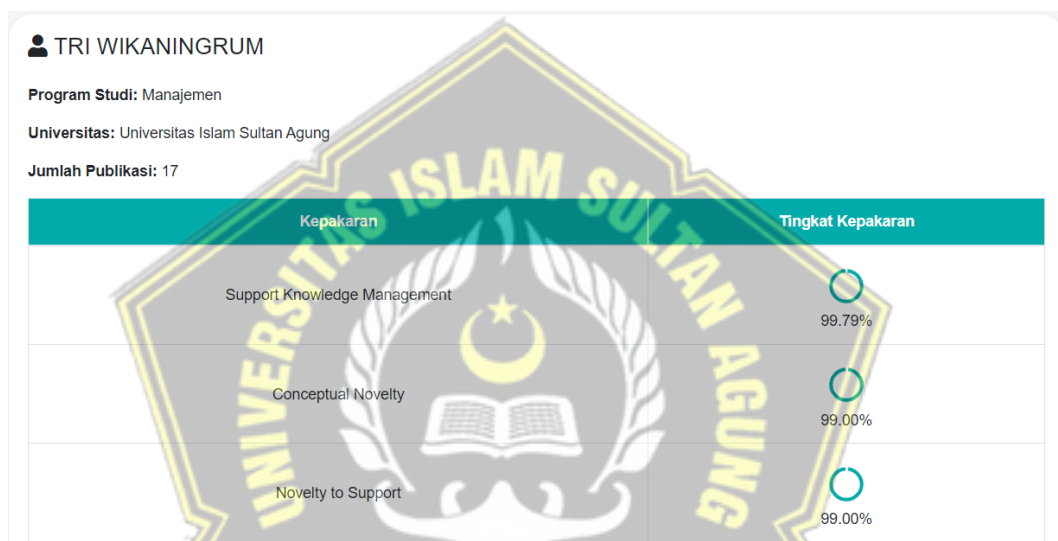
**Program Studi:** Akuntansi  
**Universitas:** Universitas Islam Sultan Agung  
**Jumlah Publikasi:** 17

Gambar 4. 7 Tampilan hasil sistem berdasarkan pencarian kepakaran dosen

Pada Gambar 4.7 ditampilkan hasil tampilan pencarian berdasarkan kepakaran dosen. Pada contoh di atas, ditampilkan hasil pencarian untuk kata kunci "management." Hasil pencarian berupa tabel yang memuat kata kunci atau *keyword*

"*management*," beserta nama dosen, program studi, universitas, dan jumlah publikasi terkait. Seperti yang terlihat pada gambar 4.7, sistem akan mengurutkan hasil sesuai dengan skor kepakaran dosen dari yang tertinggi. Untuk setiap dosen, hanya akan ditampilkan satu kepakaran yang paling sesuai dengan pencarian, yaitu yang memiliki skor tertinggi.

Untuk memperoleh informasi lengkap mengenai kepakaran dosen tersebut, pengguna dapat mengklik tombol "detail" pada kolom yang sesuai dengan dosen yang ingin dilihat. Tampilan detail dosen seperti pada gambar 4.8.



Kepakaran	Tingkat Kepakaran
Support Knowledge Management	99.79%
Conceptual Novelty	99.00%
Novelty to Support	99.00%

Gambar 4. 8 Tampilan sistem berdasarkan detail

Pada gambar 4.8 menampilkan detail dosen yang terkait pada hasil pencarian. Tampilan akan menampilkan profil dosen berupa nama lengkap, program studi, universitas dan jumlah publikasi disertai tabel yang berisi kepakaran dan tingkat kepakarannya yang ditampilkan berupa *chart* lingkaran pada dosen yang terkait.

Selain tampilan *home* yang berisi pencarian, pada sistem ini juga terdapat tampilan *about* yang berisi informasi mengenai sistem kepakaran dan lingkup pencarian kepakaran dosen. Untuk tampilan *about* seperti pada gambar 4.9.

## ABOUT

### SISTEM KEPAKARAN

Sistem ini merupakan sistem pencari kepakaran dosen berdasarkan rekam jejak publikasi pada sinta terindeks scopus. Kepakaran disini merupakan pakar secara spesifik bukan bidang ilmu secara umum. Kepakaran diambil dari kata kunci judul publikasi dosen menggunakan metode yet another keyword extractor. Jumlah kata kunci pada sistem ini yaitu 3264 kata kunci yang diidentifikasi sebagai kepakaran dosen



Gambar 4. 9 Tampilan menu *about*

Pada gambar 4.9 merupakan tampilan *about* dari sistem yang berisi informasi mengenai sistem kepakaran, pada bagian ini dijelaskan secara rinci mengenai sistem kepakaran yang dikembangkan. Pada menu *about* juga terdapat informasi mengenai lingkup sistem kepakaran dosen yaitu Universitas Islam Sultan Agung seperti pada gambar 4.10.



### UNISSULA

Fokus sistem ini adalah di lingkungan Universitas Islam Sultan Agung Semarang meliputi dosen dari 11 fakultas yaitu Fakultas Kedokteran, Fakultas Teknik, Fakultas Hukum, Fakultas Psikologi, Fakultas Ekonomi, Fakultas Agama Islam, Fakultas Bahasa dan Ilmu Komunikasi, Fakultas Teknologi Industri, Fakultas Ilmu Keperawatan, Fakultas Kedokteran Gigi, dan Fakultas Keguruan dan Ilmu Pendidikan. Jumlah data dosen pada sistem ini yaitu 182 dosen.

Copyright 2023 By Muhamad Reynaldi Ilhamisyah

Gambar 4. 10 Tampilan menu *about* unissula

Pada gambar 4.10 merupakan tampilan menu *about* pada sistem tetapi pada bagian UNISSULA. Pada bagian ini menjelaskan secara rinci tentang fokus pengembangan sistem kepakaran dosen yaitu di lingkungan Universitas Islam Sultan Agung Semarang.

### 4.3 Hasil Pengujian Sistem

Metode pengujian sistem yang digunakan dalam penelitian ini adalah pengujian *blackbox*. Dalam proses uji coba akan dilaksanakan dengan mengacu fungsionalitas pada input dan output yang dihasilkan apakah sesuai dengan hasil yang diharapkan, pengujian ini tertera dalam tabel 4.1

Tabel 4. 1 Pengujian *blackbox*

No.	Input	Hasil yang diharapkan	Output	Hasil
1.	Mencari kepakaran berdasarkan nama lengkap dosen. pada halaman home, kata nama lengkap dosen dimasukkan kedalam kolom masukkan kata setelah itu klik tombol cari	Sistem menampilkan hasil kepakaran dosen yang sesuai dengan nama lengkap dosen terkait. Tampilan berupa nama dosen, program studi, universitas, dan jumlah publikasi disertai tabel yang berisi kepakaran dan tingkat kepakarannya berupa <i>chart</i> lingkaran dilengkapi keterangan persen pada dosen yang terkait.	Menampilkan hasil kepakaran berdasarkan nama lengkap dosen disertai profil dosen terkait	Berhasil
2.	Mencari kepakaran berdasarkan kata kunci nama dosen, pada halaman home, kata kunci dosen dimasukkan kedalam kolom masukkan kata setelah itu klik tombol cari	Sistem menampilkan hasil kepakaran dosen yang sesuai dengan kata kunci nama dosen terkait, pada hasil pencarian akan menampilkan beberapa data kepakaran dosen jika terdapat beberapa nama dosen yang terkait pada pencarian. Tampilan	Menampilkan hasil kepakaran berdasarkan kata kunci nama dosen disertai profil dosen terkait	Berhasil



No.	Input	Hasil yang diharapkan	Output	Hasil
		berupa nama dosen, program studi, universitas, dan jumlah publikasi disertai tabel yang berisi kepakaran dan tingkat kepakarannya berupa <i>chart</i> lingkaran dilengkapi keterangan persen pada dosen yang terkait.		
3.	Mencari kepakaran berdasarkan <i>keyword</i> kepakaran dosen. pada halaman home, <i>keyword</i> kepakaran dosen dimasukkan kedalam kolom masukkan kata setelah itu klik tombol cari	Sistem menampilkan hasil kepakaran dosen yang sesuai dengan <i>keyword</i> atau kata kunci yang terkait dengan pencarian. Tampilan berupa nama dosen, program studi, universitas, dan jumlah publikasi disertai tabel yang berisi kepakaran dan tingkat kepakarannya berupa <i>chart</i> lingkaran dilengkapi keterangan persen pada dosen yang terkait pada <i>keyword</i> yang dicari.	Menampilkan hasil kepakaran berdasarkan <i>keyword</i> kepakaran dosen disertai profil dosen terkait	Berhasil
4.	Klik tombol detail pada tabel hasil	Sistem menampilkan detail dosen yang terkait	Menampilkan detail	Berhasil

No.	Input	Hasil yang diharapkan	Output	Hasil
	pencarian berdasarkan <i>keyword</i> kepakaran dosen	pada hasil pencarian. Tampilan akan menampilkan profil dosen berupa nama lengkap, program studi, universitas dan jumlah publikasi disertai tabel yang berisi kepakaran dan tingkat kepakarannya yang ditampilkan berupa <i>chart</i> lingkaran dilengkapi keterangan persen pada dosen yang terkait.	kepakaran dosen berdasarkan tombol detail yang diklik disertai profil dosen terkait	
5.	Klik menu <i>about</i>	Sistem menampilkan menu <i>about</i> yang berisi informasi lengkap mengenai sistem kepakaran yang dikembangkan dan lingkup pencarian sistem kepakaran dosen.	Menampilkan menu <i>about</i> yang berisi informasi mengenai sistem kepakaran dosen dan lingkup pencarian sistem kepakaran dosen	Berhasil

Berdasarkan tabel 4.1 hasil pengujian dari sistem ini, Sistem berjalan sesuai dengan hasil yang diharapkan. Sistem dapat digunakan untuk mencari kepakaran dosen berdasarkan nama lengkap dosen, nama kata kunci dosen, keparakan dosen,

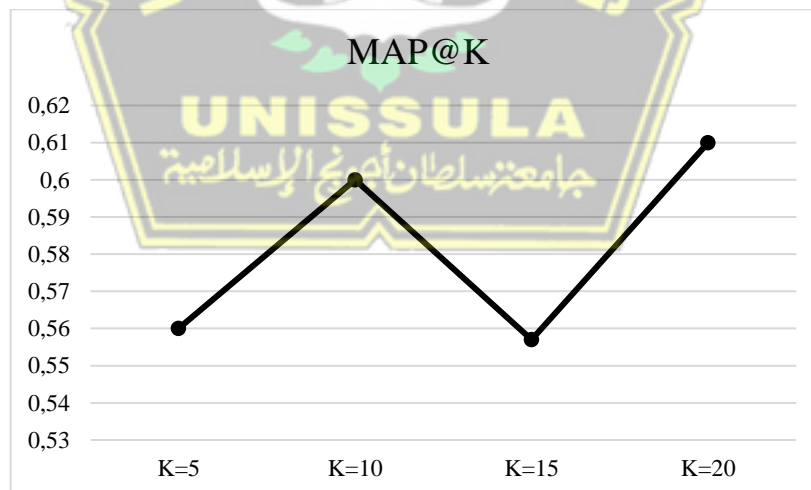
detail informasi dosen, dan informasi mengenai sistem tersebut. Sehingga sistem pencarian kepakaran dosen di Universitas Islam Sultan Agung Semarang sudah layak untuk digunakan oleh *user*.

Menentukan keakuratan kata kunci yang diuji dan menentukan jumlah kata kunci maksimal yang dipakai dalam penelitian ini menggunakan metode *Mean Average Precision @K* (MAP@K) diimplementasikan pada data yang ada dan mengambil *sample* 5 data dosen sebagai bahan penelitian dengan hasil seperti pada tabel 4.2

Tabel 4. 2 *Mean Average Precision*

	Heru Sulistyoyo	Pratikso	Sri Hartono	Azizah Hikma Safitri	Gunawan Dewantoro	Rata-rata
K=5	0,6	0,6	0,6	0,6	0,4	0,56
K=10	0,5	0,6	0,7	0,6	0,6	0,6
K=15	0,466	0,533	0,66	0,6	0,53	0,557
K=20*	0,45	0,6	0,7	0,7	0,6	0,61

Tabel 4.2 merupakan tabel hasil dari perhitungan *Mean Average Precision@K* (MAP@K) untuk menentukan relevansi kata kunci yang digunakan sebagai data kepakaran dosen.



Gambar 4. 11 Tampilan hasil perhitungan *MAP@K*

Berdasarkan Gambar 4.11 perhitungan *Mean Average Precision@K* (MAP@K), nilai tertinggi yaitu K=20 dengan nilai 0,61. Sehingga kata kunci yang diambil dari setiap dosen adalah 20 kata kunci.

## BAB V

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Dari penelitian ini dapat disimpulkan bahwa sistem kepakaran dosen di lingkungan Universitas Islam Sultan Agung Semarang dapat berjalan dengan baik tanpa adanya kesalahan dan dapat menghasilkan informasi mengenai kepakaran dosen secara spesifik sesuai dengan publikasi dosen yang terindeks di Scopus melalui SINTA. Dalam pengembangan sistem ini, metode yang digunakan adalah *Yet Another Keyword Extractor* (YAKE). Setiap dosen akan menghasilkan kata kunci dengan variasi 1 hingga 3 frasa kata kunci yang dihasilkan, dan jumlah kata kunci maksimal dalam penelitian ini adalah 20 kata kunci setiap dosen berdasarkan perhitungan metode *mean average precision@k* (MAP@K). Sehingga, setiap dosen akan memiliki jumlah kata kunci yang berbeda-beda sesuai dengan isi publikasi dan hasil ekstraksi metode *Yet Another Keyword Extractor* (YAKE) pada publikasi dosen tersebut. Total hasil kata kunci pada penelitian ini adalah 3264 kata kunci yang diidentifikasi sebagai kepakaran dosen secara spesifik.

#### 5.2 Saran

Berdasarkan penelitian yang sudah ada, penulis menyarankan untuk penelitian yang akan datang :

1. Untuk penelitian berikutnya, disarankan untuk mengklasifikasikan hasil identifikasi kepakaran yang spesifik menjadi kategori yang lebih luas berdasarkan bidang ilmu masing-masing. Hal ini akan memungkinkan untuk membagi kepakaran tersebut sesuai dengan aspek umum dari bidang ilmu yang relevan.
2. Pada penelitian ini hanya fokus pada identifikasi kepakaran dosen melalui judul publikasi. Penelitian selanjutnya dapat memperluas cakupan dengan mempertimbangkan sumber data lain yang relevan, seperti pengalaman mengajar, proyek riset, atau keterlibatan dalam proyek-proyek eksternal.

## DAFTAR PUSTAKA

- Campos, R. dan Mangaravite, V. (2018) “YAKE! collection-independent automatic keyword extractor,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag, hal. 806–810. Tersedia pada: [https://doi.org/10.1007/978-3-319-76941-7\\_80](https://doi.org/10.1007/978-3-319-76941-7_80).
- Firdaus, A.F.F. (2020) “Pemetaan Kepakaran Melalui Jurnal yang Telah Dipublikasikan Dengan Metode Association Rules Dan Pendekatan Algoritma Brute-Force,” *JUPITER (Jurnal Penelitian Ilmu dan Teknologi Komputer)*, 12(2).
- Ibrahim, C. (2019) “TREN KEPENULISAN PUBLIKASI PENELITIAN INDONESIA TEINDEKS PADA JURNAL SCOPUS,” *Journal of Documentation and Information Science*, 3(1).
- Kohli, M. dan Kar, A.K. (2022) “Machine learning-based ABA treatment recommendation and personalization for autism spectrum disorder: an exploratory study,” *Brain Informatics*, 9(1). Tersedia pada: <https://doi.org/10.1186/s40708-022-00164-6>.
- Nugroho, P.A. (2022) “Perbedaan antara Kata Kunci yang Dibuat oleh Penulis dengan Kecerdasan Buatan pada Website Scopus Menggunakan Bibliometrik,” *UNILIB: Jurnal Perpustakaan*, 13(1). Tersedia pada: <https://doi.org/10.20885/unilib.vol13.iss1.art7>.
- Panggabean, E. dan Silalahi, N. (2020) “Penerapan Data Mining Untuk Memprediksi Pemesanan Bibit Pohon Dengan Regresi Linear Berganda,” *JURIKOM (Jurnal Riset Komputer)*, 7(1), hal. 56. Tersedia pada: <https://doi.org/10.30865/jurikom.v7i1.1947>.
- Paputungan, F. (2022) “Frezy Paputungan Content list available at: <http://journal.ubmg.ac.id/index.php/JHSS> JOURNAL OF HULONTHALO SERVICE SOCIETY ( JHSS ) Vol . 2 Nomor 1 , September 2022 | ISSN : 2964-9250 ( Media Online ) WORKSHOP UPGRADE SCIENCE AND TECHNOLOGY INDEX ( SINTA,” 2(September).
- Saputra, A. (2020) “Pemanfaatan Science and Technology Index (SINTA) untuk



Publikasi Karya Ilmiah dan Pencarian Jurnal Nasional Terakreditasi,” *Media Pustakawan*, 27(1).

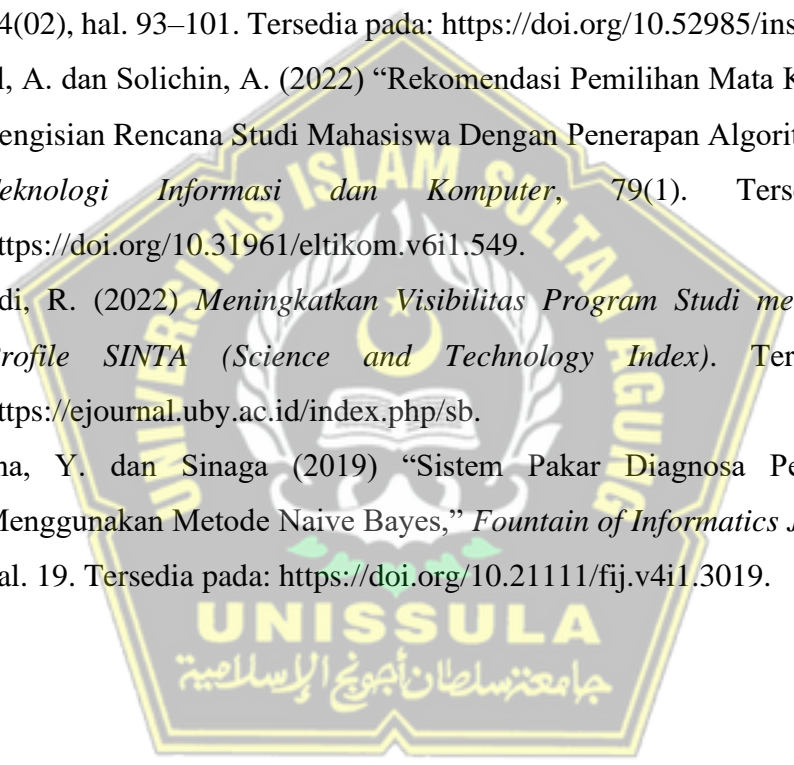
Subroto, I.M.I. dan Haviana, S.F.C. (2018) *Proposal of the S-score for measuring the performance of researchers, institutions, and journals in Indonesia, Chemical and Engineering News*. Tersedia pada: <https://doi.org/10.1021/cen-v073n013.p042>.

Susanto, E.K. (2022) “Metode Pembobotan Hibrida untuk Ekstraksi Frasa Kunci Bahasa Arab,” *INSYST: Journal of Intelligent System and Computation*, 04(02), hal. 93–101. Tersedia pada: <https://doi.org/10.52985/insyst.v4i2.255>.

Syahrul, A. dan Solichin, A. (2022) “Rekomendasi Pemilihan Mata Kuliah Dalam Pengisian Rencana Studi Mahasiswa Dengan Penerapan Algoritma Apriori,” *Teknologi Informasi dan Komputer*, 79(1). Tersedia pada: <https://doi.org/10.31961/eltikom.v6i1.549>.

Wahyudi, R. (2022) *Meningkatkan Visibilitas Program Studi melalui Update Profile SINTA (Science and Technology Index)*. Tersedia pada: <https://ejournal.uby.ac.id/index.php/sb>.

Yuliyana, Y. dan Sinaga (2019) “Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Naive Bayes,” *Fountain of Informatics Journal*, 4(1), hal. 19. Tersedia pada: <https://doi.org/10.21111/fij.v4i1.3019>.



LAMPIRAN

