

Identifikasi Bidang Kepakaran Program Studi Informatika Berdasarkan Rekam Jejak Publikasi Di Indonesia Menggunakan Rake

Achmad Naufal Fadhilah, Imam Much Ibnu Subroto, Sam Farisa Chaerul Haviana
Teknik Informatika, Fakultas Teknologi Industri, Universitas Islam Sultan Agung

Correspondence Author: imam@unissula.ac.id

Abstract

Program studi pada universitas di Indonesia baik universitas swasta maupun universitas negeri terkadang memiliki bidang kepakaran yang berbeda meskipun nama program studi tersebut terlihat sama, hal ini karena program studi diberbagai universitas sangat tergantung dengan kepakaran dosen pada universitas tersebut. Untuk itu, diperlukan sebuah sistem yang memuat informasi tentang bidang kepakaran program studi khususnya Informatika untuk membantu para calon mahasiswa untuk lebih mudah memperoleh informasi program studi informatika dan universitas mana yang sesuai dengan minat para calon mahasiswa tersebut. Dengan memanfaatkan metode RAKE dapat mewujudkan kemudahan memperoleh informasi. Metode ini dipilih karena sangat cocok untuk digunakan meranking seberapa tinggi bidang kepakaran program studi informatika di Indonesia. Pembuatan sistem diawali dengan mengumpulkan data dari beberapa website seperti SINTA Kemendikbud dan Scopus, implementasi metode dan pembuatan sistem dengan bahasa pemrograman python dan MySQL, dan pengujian sistem. Hasil penelitian menunjukkan bahwa sistem informasi dapat berjalan dengan baik setelah pengujian dan menunjukkan hasil berupa informasi kepakaran program studi ketika user memasukkan keyword ataupun nama program studi yaitu informatika. sistem berhasil menampilkan sistem informasi kepakaran program studi berdasarkan universitas yang dipilih dengan menampilkan kata kunci yang berhasil dihitung berjumlah 2716 dari 5 universitas di Indonesia dan variasi K yang dipakai yaitu K=20 dengan skor 0,53 sesuai perhitungan mean average precision@k
Keyword: rake, program studi informatika, kepakaran.

1. PENDAHULUAN

Menurut kamus *online Webster*, pakar adalah seseorang yang memiliki, melibatkan, atau menampilkan keterampilan atau pengetahuan khusus yang diperoleh dari pelatihan atau pengalaman. Istilah keahlian sering digunakan dalam arti relatif dalam penelitian pendidikan. Keahlian dapat didefinisikan dalam hal kinerja luar biasa dalam suatu domain; penelitian kinerja ahli menyelidiki kinerja unggul secara konsisten dari individu yang unggul dalam tugas perwakilan dalam domain itu[1].

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

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

Selama ini, belum pernah diketahui rule pemilihan program studi yang dilakukan oleh calon mahasiswa. Sehingga belum dapat diketahui strategi khusus yang akan dilakukan pada kegiatan promosi. Metode RAKE sudah digunakan di beberapa penelitian sebelumnya, namun belum ada yang menggunakannya untuk penentuan kepakaran program studi pada universitas di Indonesia.

Pada *website SINTA (Science and Technology Index)* sudah mencantumkan kinerja dosen, peneliti, perguruan tinggi, program studi, dan jurnal Indonesia. Karena SINTA merupakan *framework data website* yang menawarkan akses cepat, sederhana dan menyeluruh [2]. Namun SINTA masih belum ada sistem informasi untuk bidang kepakaran suatu program studi universitas sehingga para calon mahasiswa tidak cukup informasi.

Dengan alasan itulah, penulis mengusulkan membuat sistem yang bisa membantu para calon mahasiswa untuk lebih mudah mengetahui bidang kepakaran dari suatu program studi khususnya informatika pada universitas di

Indonesia, dengan dibuatnya sistem ini diharapkan bisa membantu para calon mahasiswa lebih mudah memperoleh informasi tentang bidang kepakaran suatu program studi informatika dan dapat lebih mudah menentukan program studi informatika dan universitas mana yang sesuai dengan minat para calon mahasiswa tersebut.

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

Data Mining

Data mining adalah strategi yang dapat membantu memprediksi kekhasan suatu *item* yang dikonsentrasikan dengan terlebih dahulu mengetahui atau tidak mengetahui kelas *item* tertentu melalui masukan dari data besar. Suatu masyarakat dapat meningkatkan pengelolaan lingkungannya dengan menggunakan data mining[3].

Scopus

Database kutipan dan abstrak yang diawasi oleh para profesional industri merupakan pengertian dari Scopus. Berdasarkan data situs Scopus (www.scopus.com) yang diperoleh pada 3 Juni 2020, informasi yang terdapat dalam Scopus terdiri dari 39.744 jurnal ilmiah aktif, 29.683 judul, 1.499 judul seri buku dengan 59.698 *volume*, pertukaran distribusi lebih dari 300 judul dan artikel yang didistribusikan oleh distributor di seluruh dunia berjumlah lebih dari 8.000 judul. Terdapat juga 221.169 judul buku monografi atau buku teks di Scopus. Informasi di Scopus berubah secara konsisten untuk menyesuaikan diri dengan informasi tambahan karena distribusi baru atau penurunan informasi karena distribusi yang rusak[4]. Pencarian Scopus juga mencakup pencarian *database* paten. Selain itu, Scopus menawarkan empat metrik kualitas yang berbeda untuk setiap judul: yaitu h-Record, CiteScore, SJR (*ScImago Diary Rank*) dan Clip[5].

RAKE (Rapid Automatic Keyword Extraction)

RAKE (*Rapid Automatic Keyword Extraction*) adalah algoritma yang digunakan untuk mengekstraksi kata kunci atau frasa penting dari sebuah teks secara otomatis. Algoritma ini dirancang untuk mengidentifikasi kata-kata kunci berdasarkan karakteristik linguistik seperti frekuensi kata, kehadiran kata dalam konteks yang relevan, serta pola penempatan kata yang spesifik. metode tanpa pengawasan, domain-independen, dan bahasa-independen untuk mengekstraksi kata kunci dari dokumen individual[6].

Kinerja RAKE lebih baik dari Textrank. Ini karena kemampuan RAKE untuk merangking kata kunci dalam satu langkah sementara Textrank membutuhkan iterasi terus menerus untuk mendapatkan konvergensi pada skor kata. Keunggulan RAKE adalah efisiensi komputasinya, kecepatannya, presisinya meskipun sederhana dan kemampuannya untuk bekerja pada dokumen individual[7].

SINTA (Science and Technology Index)

Sistem identifikasi bidang kepakaran program studi informatika ini akan mengambil data pada SINTA (*Science and Technology Index*) karena SINTA memberikan akses masuk ke referensi dan keahlian di Indonesia. Sebuah *website* untuk mengukur kinerja dosen, analis, perguruan tinggi, program, studi dan jurnal di Indonesia. SINTA memberikan standar dan analisis, membedakan kualitas penelitian setiap perguruan tinggi dan program studi untuk membina kerjasama yang kooperatif, menyelidiki pola penelitian dan direktori ahli. Selain itu, SINTA merupakan portal *indexing* jurnal yang saat ini diawasi oleh Dinas Pendidikan dan Kebudayaan Republik Indonesia. Selain itu, ada juga yang menegaskan bahwa SINTA merupakan *database* jurnal nasional yang diakreditasi Arjuna[2].

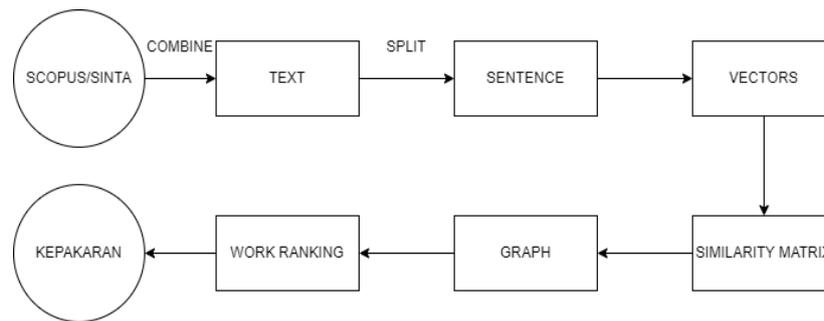
SINTA perlu mengadopsi metode pengukuran yang dinormalisasi. Normalisasi memperhitungkan parameter berikut: jumlah rata-rata kutipan per publikasi (tidak termasuk kutipan sendiri); persentase publikasi tanpa kutipan, jumlah rata-rata kutipan jurnal, dan kinerja unit penelitian di bidang terkait di seluruh dunia. Pengukuran kinerja penelitian yang akan diterapkan dalam SINTA harus mempertimbangkan bidang interdisipliner dan transdisipliner, karena beberapa ukuran mungkin tidak memadai untuk bidang tersebut[8].

2. METODE

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

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

Metode Ekstraksi



Gambar 1 Alur Proses Ekstraksi RAKE

Pada gambar 1 merupakan alur proses dari ekstraksi RAKE diantara lain:

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

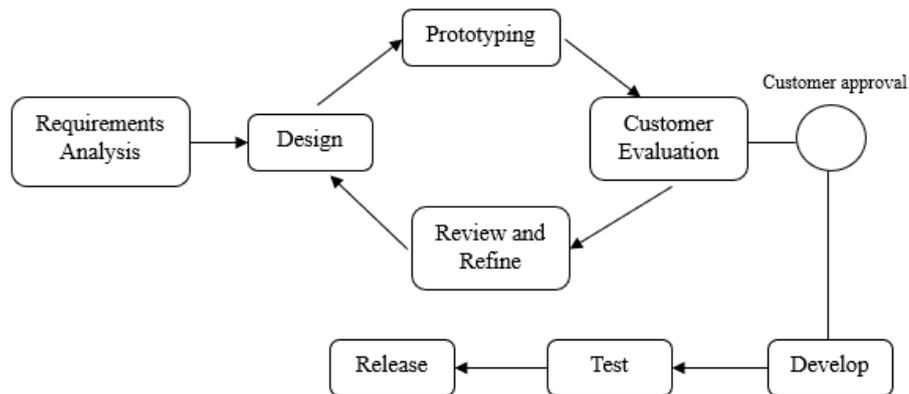
Dalam proses ekstraksi *keywords*, RAKE menggunakan *stoplist* untuk mendapatkan *list* dari *candidate keywords* dari sebuah dokumen. Kemudian skor dari *candidate keywords* dihitung menggunakan perankingan berbasis graf.

Berikut merupakan tahapan-tahapan proses ekstraksi teks menggunakan RAKE:

- Teks Preprocessing*
Didalam teks *preprocessing*, teks diolah dengan cara tokenizing dan case folding.
 - Tokenizing* atau tokenisasi adalah pemisahan urutan teks (kalimat) menjadi potongan-potongan, yang disebut token. Token dapat berupa kata, karakter, atau bagian teks lainnya tergantung pada kebutuhan. Misalnya, dalam tokenisasi kata, teks dibagi menjadi kata-kata individu.
 - Case folding* merupakan proses dalam *text preprocessing* yang dilakukan untuk menyeragamkan karakter pada data. Proses *case folding* adalah proses mengubah seluruh huruf menjadi huruf kecil.
- Delimiters*
Hasil *preprocessing* diatas dibagi menjadi urutan kata-kata yang dipisahkan dengan *delimiters* (tanda seperti koma (,), titik (.), dsb) dan *stopword* sehingga didapatkan *candidate keywords*. *Stopword* merupakan kata yang dianggap tidak informatif dan tidak mengandung makna jadi tidak diperlukan atau dihapus, contoh (yang, begitu, dari, dll.)
- Menghitung *score keyword*
Perhitungan algoritma RAKE mengambil rasio derajat frekuensi dari kata-kata. Tahap pertama untuk menghitung derajat frekuensi adalah dengan membuat *co-occurrence graph* (graf kemunculan bersama). Tahap kedua menghitung score dapat menghitung dua hal terlebih dahulu yaitu:
 - Word Frequency* ($freq(w)$)
Perhitungan *word frequency* merupakan perhitungan dari berapa kali kata tertentu muncul di antara semua *keywords* yang ada.
 - Word Degree* ($deg(w)$)
Perhitungan *word degree* merupakan jumlah nilai dari baris suatu kata tertentu pada *co-occurrence graph*.
 Kemudian untuk menghitung *keywords score*, dapat digunakan rumus berikut ini:

$$Keywords\ score = \left(\frac{deg(w)}{freq(w)} \right) \quad (1)$$

Metode Pengembangan Sistem



Gambar 2 Model *Prototype*

Pada gambar 2 merupakan model *prototype* yang digunakan untuk metode pengembangan sistem. *Prototype* dipilih karena dengan menggunakan *prototype*, pengguna dapat ikut andil dalam proses pengembangan dengan cara mengevaluasi dan memberikan umpan balik. Model *prototype* yang dipakai yaitu *extreme prototyping*, model ini merupakan salah satu pendekatan dalam pengembangan perangkat lunak yang fokus pada pengembangan *prototype* yang dapat digunakan untuk mendemonstrasikan dan menguji fungsi-fungsi utama dari sebuah aplikasi atau sistem secara sangat cepat. Pendekatan ini sering digunakan dalam pengembangan perangkat lunak berbasis *web*, tetapi juga dapat diterapkan pada proyek-proyek perangkat lunak lainnya. *Extreme prototype* dibagi menjadi tiga tahap yaitu

1. Membuat *prototype* dasar yang berisi semua halaman berformat HTML.
 2. Lapisan *prototype service* digunakan untuk mensimulasikan pemrosesan data.
 3. Layanan diimplementasikan dan diintegrasikan ke dalam *prototype* akhir.
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. Tujuan komunikasi ini adalah untuk mendiskusikan dan menggambarkan analisa kebutuhan sistem dan rencana pengembangan sistem secara lebih rinci dan menyeluruh. Hal ini bertujuan untuk memastikan bahwa persepsi dan pemahaman mengenai analisa kebutuhan dan pengembangan sistem ini dapat diselaraskan di antara semua pihak yang terlibat. Hasil dari tahapan ini peneliti mendapatkan data dari tim pengembang SINTA berupa data judul publikasi dosen yang terindeks scopus. Pada tahap ini juga peneliti mendapatkan gambaran mengenai system yang akan dikembangkan yaitu sistem kepakaran program studi informatika di Indonesia berdasarkan publikasi pada SINTA terindeks scopus.

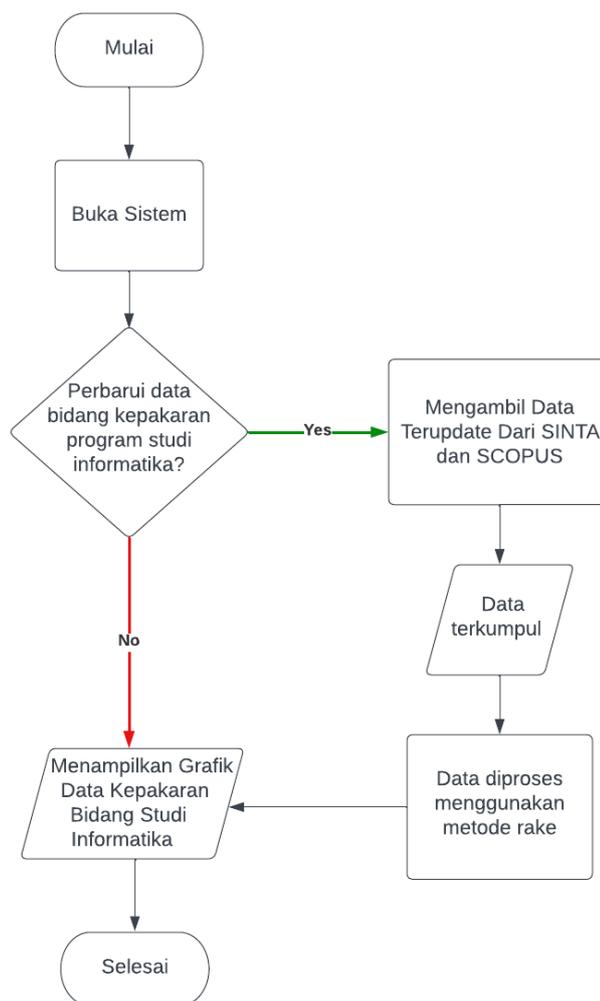
Pada tahap ini juga peneliti menganalisa kebutuhan sistem yaitu apa saja tools yang dibutuhkan untuk mengembangkan sistem ini sehingga proses *input* sampai dengan sistem berjalan sesuai dengan yang diinginkan. Adapun *tools* yang dibutuhkan antara lain meliputi *python* sebagai bahasa pemrograman untuk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

2. *Design*

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

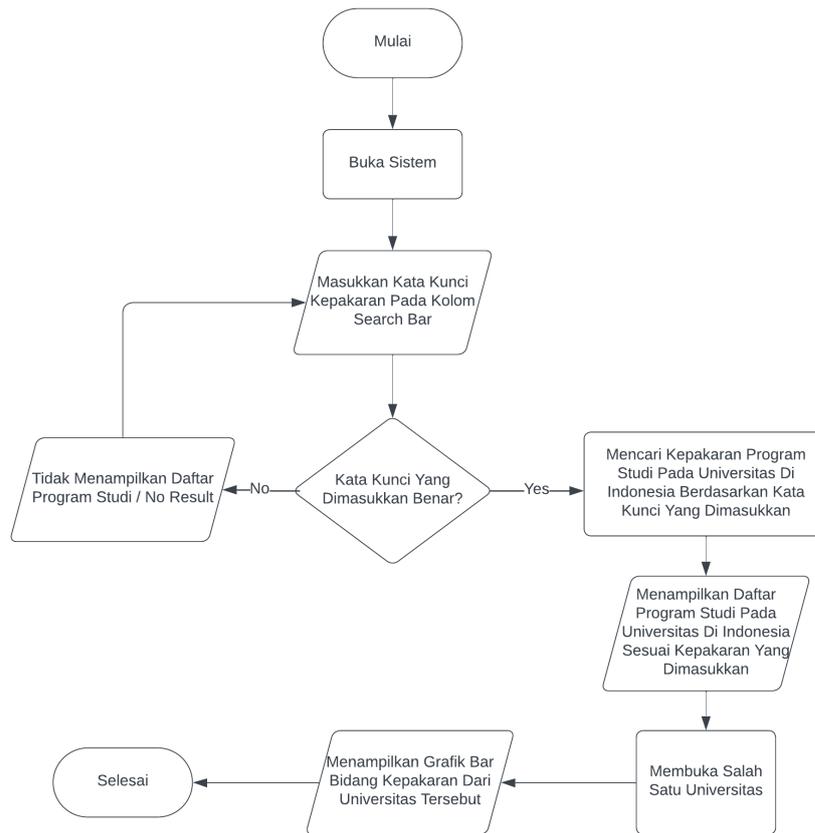
a. *Flowchart* Sistem

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



Gambar 3 *Flowchart* Sistem untuk admin

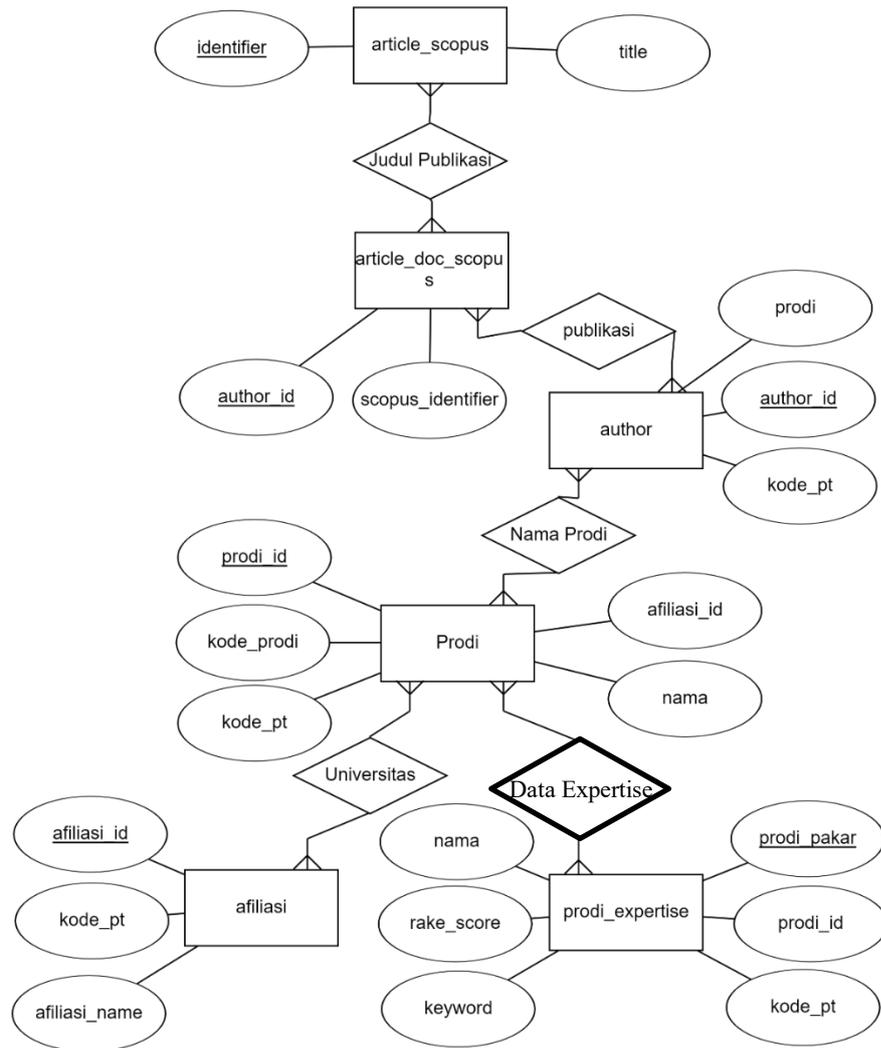
Gambar 3 memperlihatkan *flowchart* sistem untuk *admin* dimana dimulai dari membuka sistem kemudian akan terdapat pilihan apakah ingin memperbarui data dari bidang kepakaran atau tidak. Jika iya, maka sistem akan mengambil data yang *terupdate* dari SINTA/SCOPUS, setelah data terkumpul kemudian diproses menggunakan metode *RAKE* dengan hasil output berupa grafik. Jika tidak, maka hanya menampilkan grafik data kepakaran bidang studi informatika.



Gambar 4 Flowchart Sistem Untuk User

Pada gambar 4 merupakan *flowchart* alur kerja dari sistem *user* yaitu mulai dengan membuka sistem, kemudian *user* memasukkan kata kunci kepakaran pada kolom *search bar*, apabila kata kunci yang dimasukkan tidak benar, maka tidak akan menampilkan daftar program studi atau pencarian *not result* yang kemudian akan diminta untuk memasukkan kata kunci kembali pada *search bar*. Jika kata kunci yang dimasukkan benar, maka sistem akan mencari kepakaran program studi pada universitas di Indonesia berdasarkan kata kunci yang dimasukkan, selanjutnya akan ditampilkan daftar program studi pada universitas sesuai kata kunci kepakaran yang dimasukkan, selanjutnya *user* dapat memilih salah satu universitas yang kemudian akan menampilkan grafik bar bidang kepakaran dari universitas yang dipilih.

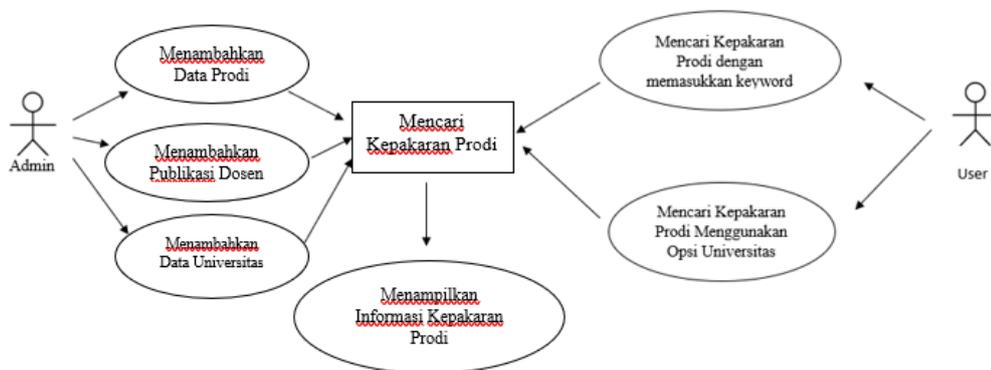
b. ERD



Gambar 5 ERD Model

Pada gambar 5 merupakan ERD dari penelitian yang akan dilakukan, jadi mengambil judul publikasi dari tabel artikel scopus berdasarkan publikasi dari author. Prodi mengambil data dari author dan dibuat tabel *prodi expertise* sebagai tabel penyimpanan data, dan data dari afiliasi digunakan untuk mengambil afiliasi nama agar dapat diinput bersama dengan data pada tabel *prodi expertise*.

c. Use Case



Gambar 6 Use Case sistem

Pada gambar 6 merupakan gambar dari gambar *use case* sistem dimana *admin* dapat menambahkan data prodi, menambahkan publikasi dari dosen, dan juga menambahkan data universitas tetapi penambahan ini dilakukan oleh *admin* SINTA karena peneliti mendapatkan data dari SINTA. Kemudian diproses oleh sistem akan menghasilkan informasi kepakaran prodi tertentu. Sedangkan *user* dapat mencari kepakaran prodi dengan memasukkan *keyword* atau mencari kepakaran menggunakan opsi universitas. Kemudian sistem akan memproses dan menghasilkan data kepakaran prodi yang dicari *user*.

d. Perancangan *User Interface*

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

3. Pembentukan *Prototyping*

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

- a. Pengumpulan Data: setelah melakukan wawancara dengan tim pengembang SINTA, didapatkan data yang bersumber dari SINTA yang terindeks scopus, data yang diperoleh mencakup publikasi *author*, prodi, dan universitas yang ada di Indonesia.
- b. Data yang sudah diperoleh kemudian diolah menggunakan bahasa pemrograman *python*. *Python* dipilih sesuai dengan kemampuan dalam melakukan analisis data.
- c. Data diolah dengan mengimplementasikan metode RAKE (*Rapid Automatic Keyword Extraction*) untuk menentukan kata kunci dan skor dari jurnal publikasi dari *author* pada *jupyter notebook*
- d. Setelah mendapatkan skor kepakaran dari ekstraksi kata kunci, kemudian data terbaru disimpan didalam *database* baru pada *MySQL*.
- e. Data yang tersimpan pada *database* dijadikan dasar utama untuk ditampilkan pada *website* yang akan dibuat pada *visual studio code*
- f. Pengembangan sistem yaitu dengan membuat sebuah sistem dan data kepakaran tersebut sebagai dasar bagi pengembangan sistem. Sistem yang dibuat berbasis *website* agar mudah diakses oleh pengguna. Pengembangan sistem akan menggunakan *platform streamlit*. Penggunaan *platform* ini didasarkan kemampuan dalam membuat *website* dan sangat praktis bagi *developer*.

4. *Customer Evaluation*

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

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*.

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

6. *Develop*

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

7. *Release*

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

3. HASIL DAN PEMBAHASAN

Hasil dari penelitian pengidentifikasi bidang kepakaran program studi informatika di Indonesia menggunakan RAKE telah memenuhi akhir dengan jumlah data yang berhasil dikumpulkan yaitu sebanyak 2716 dengan mengambil 5 universitas di Indonesia.

Berikut merupakan contoh data yang sudah diolah menggunakan RAKE dan python dan hasilnya seperti gambar 7

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

Gambar 7 Data yang sudah diolah dengan RAKE

Pada gambar 7 merupakan data program studi informatika unissula yang sudah diolah dengan menggunakan metode RAKE dan python dan diperoleh skor kepakaran beserta keyword.

Namun kata kunci yang sudah diolah menggunakan RAKE tidak sepenuhnya akurat, karena terdapat *relevant document* dari *retrieved document* yang tidak bisa di *filter* oleh RAKE, untuk meminimalisir hal tersebut, dapat menggunakan metode *Mean Average Precision at K* (MAP@K) dengan menggunakan rumus:

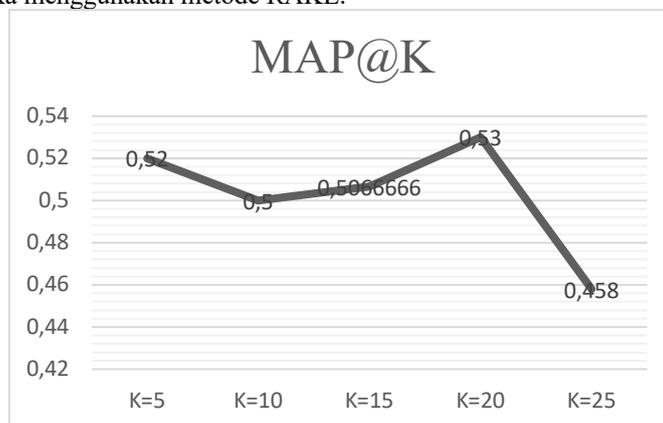
$$Precision = \frac{|{relevant documents} \cap {retrieved documents}|}{|{retrieved documents}|} \tag{1}$$

Dengan menggunakan rumus tersebut, kata kunci kembali diolah untuk mendapatkan dokumen yang relevan pada setiap universitas, dengan hasil seperti pada tabel berikut

Tabel 1 *Mean Average Precision at K*

	Unissula	Uppgris	UKSW	UNIKA	Udinus	Rata-rata
K=5	0,40	0,60	0,2	0,8	0,6	0,52
K=10	0,40	0,60	0,4	0,6	0,5	0,5
K=15	0,33	0,73	0,47	0,6	0,4	0,51
K=20*	0,45	0,70	0,55	0,5	0,45	0,53
K=25	0,40	0,56	0,56	0,52	0,25	0,458

Pada tabel 1 merupakan tabel hasil dari *Mean Average Precision@K* (MAP@K) untuk menentukan kerelevan kata kunci ketika menggunakan metode RAKE.



Gambar 8 MAP@K

Gambar 8 merupakan gambar dari perhitungan *Mean Average Precision@K* (MAP@K) dimana nilai tertinggi yang mendekati satu yaitu K=20 dengan nilai 0,53

Anda dapat memilih opsi universitas yang tersedia:

Universitas Islam Sultan Agung ▼

Cari

Profil Universitas Islam Sultan Agung

Program Studi: Teknik Informatika

Keyword	Diagram Donat
various similarity measures	

Gambar 9 Pengimplementasian hasil ekstraksi ke *website*

Pada gambar 9 merupakan pengimplementasian hasil ekstraksi ke website dalam hal ini saya ambil ekstraksi keyword dari program studi unissula, dimana divisualisasikan dengan donat chart dimana skor menunjukkan 100% menandakan keyword tersebut keparannya sangat tinggi.

Data yang dikumpulkan berupa prodi id, kode pt, nama program studi, keywords yang dihitung, dan skor perhitungan dengan RAKE seperti pada tabel berikut diambil data 3 sampel.

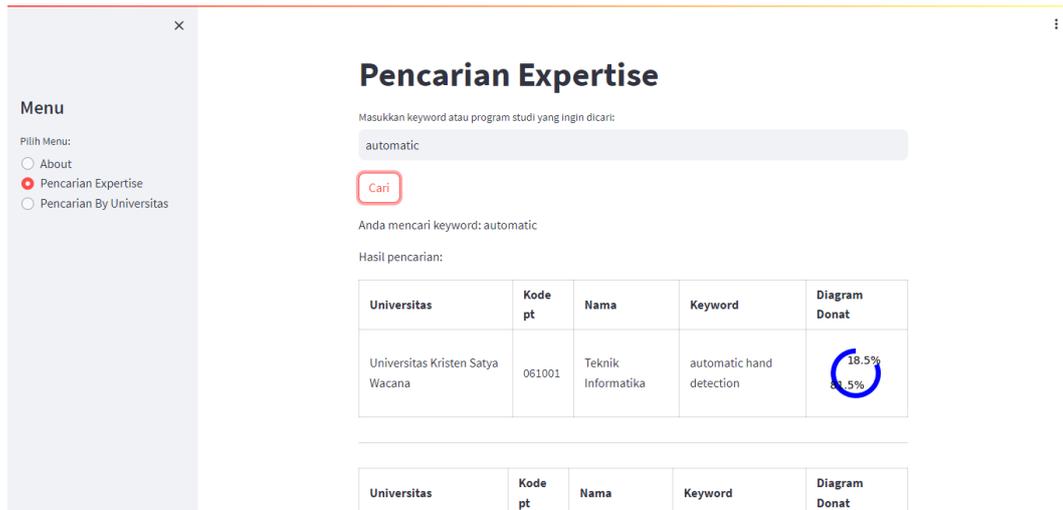
Tabel 2 Data Terkumpul

Prodi id	Kode pt	Nama	Keyword	RAKE Score
55201	061002	Teknik Informatika	<i>Average image hash</i>	8.5
55201	061001	Teknik Informatika	<i>without initial route</i>	9
55201	061001	Teknik Informatika	<i>financial data integration</i>	8.85714

Pada tabel 1 merupakan data yang sudah berhasil dikumpulkan dan dihitung dengan menggunakan metode RAKE. Analisis hasil menunjukkan pada kolom kode pt, nama, *keyword*, dan skor sesuai dengan universitas dengan pengambilan sampel 3 data. Skor keparan terbesar yaitu dengan nilai 9.0 dan skor keparan terkecil dengan nilai 1.0. *keywords* yang diekstraksi diambil satu sampai tiga kata setiap judul publikasi scopus.

Implementasi Sistem

Implementasi antarmuka merupakan tahap pemaparan tampilan sistem dan kegunaan fungsi dari *form* yang tersedia. Pada tahap ini, implementasi hasil rancangan dan analisa telah menghasilkan sebuah *website* identifikasi bidang keparan program studi informatika dengan bahasa pemrograman python dan mysql.

Gambar 10 Menu *keyword website*

Pada gambar 10 merupakan tampilan menu *keywords* dari *website* identifikasi kepakaran, dimana ada kolom *search bar* untuk memasukkan *keyword* yang ingin dicari, *user* juga dapat memasukkan nama program studi pada *search bar* tersebut. Terdapat tombol cari untuk melakukan pencarian *keyword* yang sudah dimasukkan. Hasilnya berupa tabel informasi mengenai universitas, program studi dan skor kepakaran yang divisualisasikan dengan diagram donat chart yang dimana jika skor kepakaran tinggi, maka akan penuh dan mencapai 100%.

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



Gambar 11 Menu opsi universitas

Pada gambar 11 merupakan menu opsi pencarian kepakaran berdasarkan opsi universitas, jadi terdapat menu memilih universitas dan juga opsi program studi untuk dicari kepakarannya dengan klik tombol cari sebagai bottonnya. Hasil dari pencarian ini juga berupa tabel informasi mengenai skor kepakaran pada program studi tersebut dengan divisualisasikan dengan diagram donat *chart*.

4. KESIMPULAN

Dari penelitian ini dapat disimpulkan bahwa sistem kepakaran program studi di Universitas di Indonesia dapat berjalan dengan baik tanpa adanya kesalahan dan dapat menghasilkan informasi mengenai kepakaran program studi secara spesifik sesuai dengan publikasi dosen yang terindeks di Scopus melalui SINTA. Dalam pengembangan sistem ini, metode yang digunakan adalah dengan mengimplementasikan metode RAKE. Metode RAKE telah berhasil diimplementasikan untuk menghitung kata kunci secara otomatis. Kata kunci yang berhasil dihitung berjumlah 2716 dari 5 universitas di Indonesia. Setiap program studi akan menghasilkan kata kunci dengan variasi 1 hingga 3 kata kunci yang dihasilkan, dan jumlah kata kunci yang revelant yaitu $K=20$ dengan skor 0,53 sesuai dengan perhitungan

menggunakan metode *mean average precision@k* (MAP@K). Sehingga, setiap program studi akan memiliki jumlah kata kunci yang berbeda-beda sesuai dengan isi publikasi dan hasil ekstraksi metode RAKE pada publikasi dosen tersebut.

DAFTAR PUSTAKA

- [1] V. D. Teodoro dan J. L. Schwartz, "Encyclopedia of the Sciences of Learning," *Encycl. Sci. Learn.*, no. March, 2012.
- [2] R. Wahyudi dan Berlilana, "Meningkatkan Visibilitas Program Studi melalui Update Profile SINTA (Science and Technology Index)," *uby.ac.id*, vol. 3, no. 2, hal. 32–38, 2022.
- [3] A. P. Natasuwarna, "Tantangan menghadapi era revolusi 4.0 - Big data and data mining," *Semin. Nas. Has. Pengabd. Kpd. Masy.*, hal. 23–27, 2019.
- [4] L. Hakim, "Analisis Bibliometrik Penelitian Inkubator Bisnis pada Publikasi Ilmiah Terindeks Scopus," *Procur. J. Ilm. Manaj.*, vol. 8, no. 2, hal. 176–189, 2020.
- [5] Rahmadi Islam, "Analisa Integrasi Data SINTA (Science and Technology Index) Menggunakan Website Internasional Dengan Manajemen Sistem Informasi EIS (Executive Information System)," *Anal. Integr. Data SINTA (Science Technol. Index) Menggunakan Website Int. Dengan Manaj. Sist. Inf. EIS (Executive Inf. Syst.*, no. 3, hal. 1–13, 2018.
- [6] S. Rose dan D. dkk Engel, "Automatic Keyword Extraction from Individual Documents," *Text Min. Appl. Theory*, no. October 2017, hal. 1–20, 2010.
- [7] M. G. Thushara, T. Mownika, dan R. Mangamuru, "A comparative study on different keyword extraction algorithms," *Proc. 3rd Int. Conf. Comput. Methodol. Commun. ICCMC 2019*, no. Iccmc, hal. 969–973, 2019.
- [8] I. M. Ibnu Subroto dan S. F. Chaerul Haviana, "Proposal of the S-score for measuring the performance of researchers, institutions, and journals in Indonesia," *Chem. Eng. News*, vol. 73, no. 13, hal. 42–49, 2018.