

Penerapan Algoritma Pencocokan String pada Pencarian di Aplikasi LinkedIn

Albert Sahala Theodore – 13516022

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

13516022@std.stei.itb.ac.id

Abstrak — LinkedIn merupakan jejaring sosial yang basis penggunaannya merupakan orang-orang profesional dan mempunyai latar belakang bisnis. LinkedIn telah digunakan lebih dari empat ratus juta orang untuk menghubungkan penggunaannya satu sama lain. Melalui aplikasi LinkedIn, penggunaannya dapat membuat profil, menemukan lowongan pekerjaan, menemukan orang-orang profesional lain dengan latar belakang yang sama, tergabung ke dalam suatu grup, dan mendapat wawasan tentang industri pekerjaan, dan menampilkan *curriculum vitae* secara online. Melalui LinkedIn, kebanyakan pelamar kerja mendapatkan kesan pertama dari Salah satu fitur yang sangat berguna di LinkedIn adalah fitur pencariannya, dimana pengguna bisa mencari profil seseorang dengan nama, kata kunci, perusahaan tempat orang itu bekerja, dan *skill* yang dimiliki oleh orang tersebut. Setelah itu, hasil pencarian akan diurutkan berdasarkan relevansi dan derajat koneksi secara terurut membesar. Jika kita sudah terhubung ke orang tersebut, maka hubungan orang tersebut ke pengguna adalah derajat pertama. Jika antar pengguna memiliki koneksi yang sama, maka orang tersebut merupakan derajat koneksi kedua.

Kata kunci — LinkedIn, Pencocokan String, Jejaring Sosial, Metode, Query Tagging

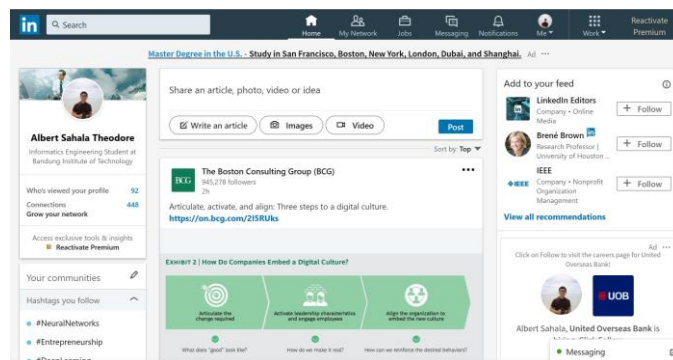
I. PENDAHULUAN

Jejaring atau media sosial saat ini sudah menjadi kebutuhan sehari-hari bagi banyak orang. Jejaring sosial digunakan agar seseorang dapat terhubung dengan orang lain tanpa dibatasi oleh jarak. Saat ini banyak jejaring sosial yang digunakan dan masing-masing memiliki manfaat dan basis penggunaannya sendiri. Seperti contohnya, Facebook digunakan untuk berbagi foto dan status serta penggunaannya lebih banyak orang berumur dua puluh tahun keatas, sedangkan Instagram digunakan untuk berbagi video dan penggunaannya didominasi oleh anak remaja hingga pemuda berumur sekitar dua puluh tahun.

Terdapat juga jejaring sosial lain yaitu LinkedIn yang memiliki basis pengguna didominasi oleh profesional dengan latar belakang bisnis dan penggunaannya terdiri dari segala umur. LinkedIn banyak digunakan untuk memperluas koneksi profesional dan untuk melamar pekerjaan. Melalui LinkedIn, banyak perusahaan yang menyebarluaskan lowongan pekerjaan sehingga para pencari kerja dapat mendaftar dengan mudah. Setiap jejaring atau media sosial, termasuk LinkedIn memiliki kesamaan, yaitu fitur pencarian profil berdasarkan nama

pengguna dan deskripsinya. Fitur ini merupakan salah satu fitur utama yang meningkatkan *discoverability* antar penggunaannya.

Berdasarkan latar belakang tersebut, makalah ini membahas algoritma dan kemampuan fitur pencarian di LinkedIn melalui pendekatan pencocokan string. Algoritma pencocokan string yang digunakan merupakan kombinasi dari beberapa metode yang dikembangkan sendiri oleh LinkedIn. Algoritma pencocokan string ini dipilih karena dapat menangani berbagai kondisi pencarian dan menangani kasus alfabet dengan berbagai variasi. Pencarian pada LinkedIn tidak hanya mencocokkan teks input dengan nama saja, melainkan juga dengan perusahaan tempat bekerja dan *skill* yang dimiliki oleh orang yang dicari.



Gambar 1.1 Tampilan Antarmuka LinkedIn

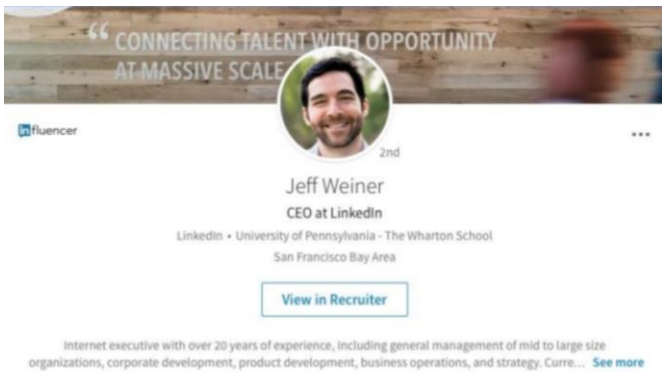
Sumber: <https://www.linkedin.com/feed/> (Diakses 11 Mei 2018 Pukul 13.40)

II. DASAR TEORI

A. Jejaring Sosial

Jejaring sosial atau yang biasa disebut media sosial merupakan aplikasi yang digunakan untuk bersosialisasi dengan orang lain secara *online* tanpa terbatas ruang dan waktu. Untuk berinteraksi melalui jejaring sosial, hanya dibutuhkan perangkat pendukung seperti telepon genggam atau komputer yang tersambung ke jaringan internet. Jejaring sosial dibagi menjadi beberapa kategori, diantaranya:

1. *Social Network* (Facebook, MySpace, LinkedIn)
2. *Publishing* (Wordpress, Wikipedia, Blogspot)
3. *Microblogging* (Twitter, Plurk, Pownce)



Gambar 2.1 Tampilan Profil LinkedIn

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 19.40)

B. Konsep String

String merupakan sekuens dari karakter atau biasa disebut *array of character* yang memiliki makna ataupun tidak memiliki makna. Dalam konsep string, terdapat istilah prefiks dan sufiks. Prefiks merupakan bagian dari string yang dimulai dari karakter pertama hingga satu karakter sebelum karakter terakhir. Sufiks merupakan bagian dari string yang dimulai dari satu karakter setelah karakter pertama sampai karakter terakhir pada string tersebut. Prefiks dan sufiks merupakan *subset* dari string tersebut.

- Misalkan sebuah string S dengan ukuran n seperti $S = X_0X_1X_2 \dots X_n$
- Prefiks yang dimulai dari karakter pertama hingga karakter kedua terakhir adalah $S[0 \dots k]$
- Sufiks yang dimulai dari karakter kedua awal hingga karakter terakhir adalah $S[1 \dots n]$

String seperti “MAKAN” terdiri dari lima karakter dengan prefiks yang memenuhi syarat berupa “M”, “MA”, “MAK”, “MAKA” dan sufiks yang memenuhi syarat berupa “A”, “AK”, “AKA”, “AKAN”.

C. Pencocokan String

Algoritma pencocokan string merupakan suatu algoritma yang sangat umum dalam dunia ilmu komputer, dimana diberikan suatu input atau *pattern* dan teks yang harus dicocokkan untuk mendapatkan suatu hasil tertentu.

- Pola dengan panjang string n karakter
- Teks dengan panjang string m karakter
- Panjang *pattern* harus lebih kecil dari panjang teks yaitu $n < m$
- Algoritma pencocokan string akan mencari pola pertama yang ditemukan dalam teks

Pencarian dilakukan seperti pada pola dan teks sebagai berikut:

1. Pola: “pergi”
2. Teks: “James tidak pergi ke kampus setiap hari libur”

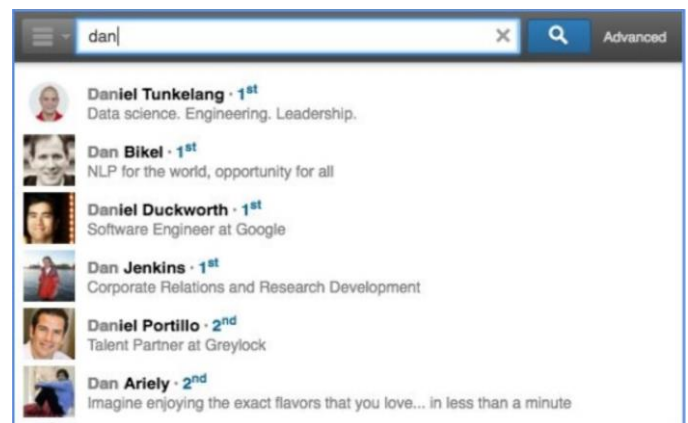
Hasil pencarian tersebut menghasilkan *true* bahwa pola tersebut cocok dengan bagian dari teks yang merupakan karakter ketigabelas.

D. Fitur Pencocokan String LinkedIn

Pada LinkedIn terdapat beberapa fitur pencarian yang menggunakan fitur pencocokan string yang terdiri dari:

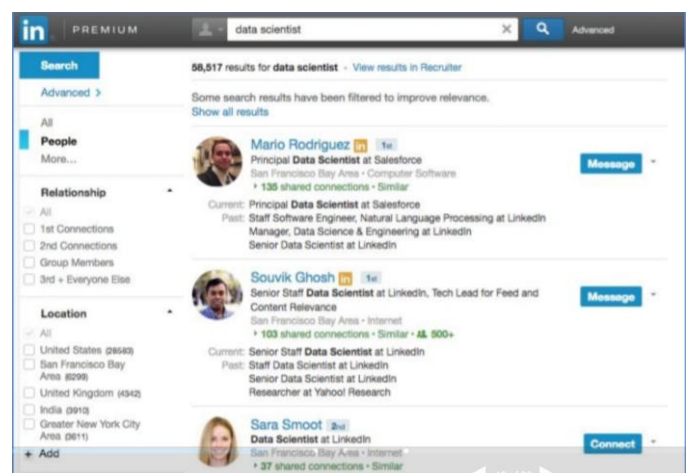
1. *Navigational Search*
2. *Exploratory Search*
3. *Instant Search*

Navigational Search merupakan fitur yang berfungsi untuk mencari seseorang dengan nama yang spesifik dan pola yang dicari akan menghasilkan keluaran yang mutlak sama. *Exploratory Search* merupakan fitur yang berfungsi untuk mencari orang yang cocok dengan kriteria yang diberikan pada pola dan menghasilkan banyak keluaran. *Instant Search* merupakan fitur yang berfungsi untuk menampilkan hasil pencarian secara instan dan melengkapi pola yang dimasukkan pengguna dengan beberapa sugesti pencarian.



Gambar 2.2 Fitur *Navigational Search*

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 20.00)



Gambar 2.3 Fitur *Exploratory Search*

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 20.05)



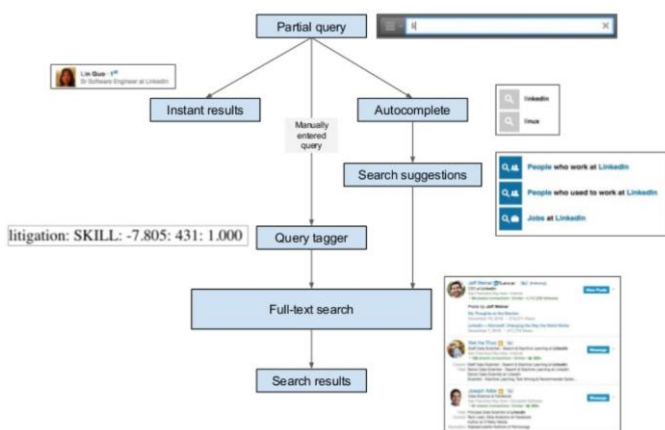
Gambar 2.4 Fitur *Instant Search*

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 20.10)

III. CARA KERJA PENCOCOKAN STRING LINKEDIN

Untuk pencarian pada aplikasi LinkedIn terdapat sebuah kolom pada bagian kiri atas untuk memasukkan query yang hasilnya akan ditampilkan secara *real time*. LinkedIn akan menampilkan hasil pencarian setiap karakter dari sebuah string ditambahkan pada kolom tersebut. Pencarian akan menghasilkan keluaran dengan pengguna lain dengan derajat koneksi yang paling dekat dengan orang tersebut.

Query atau pola yang dimasukkan pada kolom pencarian pertama-tama akan menghasilkan *instant result* yaitu beberapa profil pengguna yang namanya sama persis dengan pola. Setelah itu secara otomatis akan diberikan sugesti yang mungkin cocok dengan hasil yang diharapkan oleh pengguna. Tahap selanjutnya adalah pencocokan query atau pola dengan pencarian seluruh string yang menghasilkan hasil pencarian terakhir. Fitur pencarian pada LinkedIn didesain agar menghasilkan keluaran secara cepat, mengatasi kesalahan input pengguna, akurat, dan sesuai dengan personal pengguna.



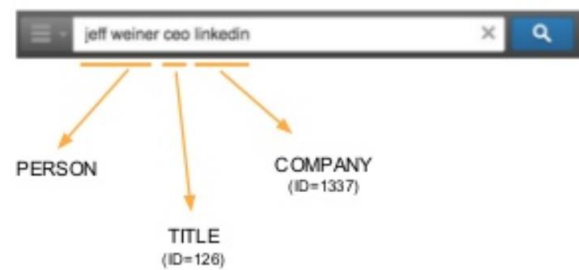
Gambar 3.1 Algoritma Pencocokan String

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 20.30)

A. Query Tagging

Dalam pencocokan string pada fitur pencarian di aplikasi LinkedIn, terdapat algoritma *query tagging* yang dapat membagi suatu input pola dari pengguna menjadi beberapa bagian bagian. Pada kasus input string dari pengguna berupa “jeff weiner ceo linkedin” maka secara otomatis algoritma pencocokan string akan mencocokkan pola tersebut dan membaginya dengan tiga kondisi yang relevan.

Dua kata awal akan dikenali sebagai nama orang yaitu “jeff weiner” selanjutnya kata ketiga akan dikenali sebagai title dari orang tersebut yaitu “ceo” dan kata terakhir akan dikenali sebagai nama perusahaan tempat orang itu bekerja yaitu “linkedin”. Algoritma pencocokan string tidak akan langsung mengenali ketiga jenis pola yang diberikan pengguna, tetapi akan membaginya ke dalam beberapa kemungkinan yaitu nama, pekerjaan, perusahaan, pendidikan, lokasi, dan kemampuan. Setelah itu, algoritma baru bekerja dengan mencocokkan input pengguna dengan kategori masing-masing pada basis data.



Gambar 3.2 *Query Tagging*

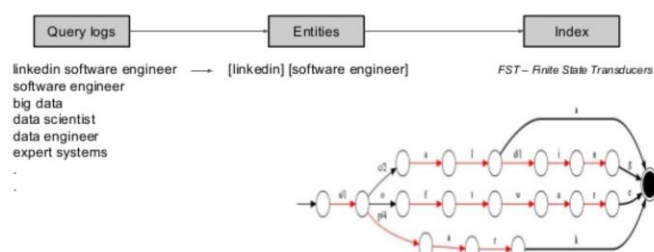
Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 20.50)

B. Query Autocomplete

Pola yang dimasukkan oleh pengguna pada kolom pencarian mungkin saja salah atau masih belum lengkap sehingga butuh dilengkapi untuk menjadi sebuah string yang kontekstual dan relevan sehingga dapat memenuhi ekspektasi pengguna.

Seperti contohnya input pengguna dengan query “linkedin software engineer” akan dibagi menjadi “linkedin” dan “software engineer” yang masing-masing akan menjadi suatu entitas sehingga pencocokan string pada teks yang berada di basis data lebih mudah untuk dilakukan dan terkategoriisasi.

Selanjutnya, input pencarian oleh pengguna akan melalui proses *indexing* atau diubah menjadi FST atau *Finite Set Transducers* yang merupakan penghitungan relasi antara dua set. Ketiga tahap ini mulai dari input, kategorisasi per entitas, dan *indexing* sangat dibutuhkan agar hasil pencocokan string tidak hanya tepat, tetapi juga relevan.



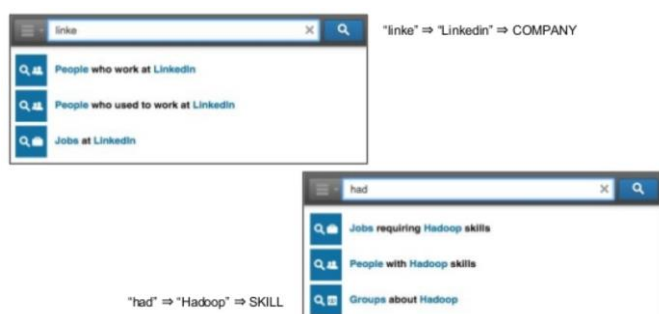
Gambar 3.3 Query Tagging

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 11 Mei 2018 Pukul 20.50)

C. Query Suggestions

Query Suggestions merupakan sugesti string dari algoritma pencocokan string pada LinkedIn, dimana pengguna akan mengisi kolom pencarian dengan pola atau query yang diinginkan dan nantinya secara otomatis akan diberikan sugesti string yang relevan dan kontekstual. Hal ini dimungkinkan karena query suggestions merupakan gabungan dari *query autocomplete* dan *query tagging*.

Seperti contohnya saat pengguna memasukkan string berupa “linke”, maka algoritma pencocokan string akan bekerja dan mencocokkan string yang belum selesai tersebut dengan teks yang terdapat pada basis data yaitu string berupa “LinkedIn” yang terdapat pada entitas nama perusahaan. Contoh lain yaitu saat pengguna memasukkan string berupa “had”, maka algoritma pencocokan string akan mulai mencocokkan dengan berbagai kemungkinan dan didapatkan hasilnya “Hadoop” yang merupakan bagian dari entitas nama *skill* pada basis data.



Gambar 3.4. Query Suggestion

Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 12 Mei 2018 Pukul 15.30)

D. Indexing

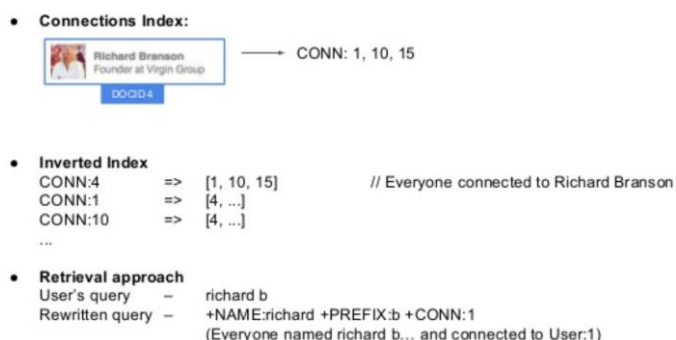
Berdasarkan masukan pencarian pola atau query dari pengguna, sistem akan menghasilkan *instant result* dengan cara membagi pola kepada kategori yang nantinya akan dilakukan pencocokan string dengan teks yang ada pada basis data. Kategori *indexing* terbagi menjadi tiga, yaitu sebagai berikut:

- Prefix-based tokenization
- Inverted index
- Retrieval approach

Prefix-based tokenization merupakan pembagian setiap kata pada input query user menjadi semua prefiks yang memungkinkan. Seperti contohnya saat pengguna memasukkan “Richard Branson” maka akan dilakukan *parsing* setiap kata dan membaginya menjadi dua string yaitu “Richard” dan “Branson”. Untuk “Richard”, prefiks yang memungkinkan berupa “R”, “Ri”, “Ric”, “Rich”, “Richa”, dan “Richar”. Untuk “Branson”, prefiks yang memungkinkan adalah “B”, “Br”, “Bra”, “Bran”, “Brans”, dan “Branso”.

Inverted Index merupakan cara indexing dengan memetakan token ke list teks atau dokumen yang memiliki konten yang stringnya cocok dengan token tersebut. Seperti contohnya input string berupa “Richard”, maka akan dipetakan ke [1, 4, 10, 15, ...] yaitu list dokumen yang persis cocok dengan nama “Richard”. Untuk prefiks “Ri” dari string “Richard” akan dipetakan ke [1, 2, 4, 7, 10, 15, ...] dimana setiap nama yang diawali prefiks “Ri” akan cocok dengan kategori ini.

Retrieval Approach merupakan cara untuk mencocokkan string dengan membagi ke dua jenis query berdasarkan input pengguna yaitu *user’s query* dan *rewritten query*. *User’s query* merupakan pola yang sama persis dengan input string dari pengguna pada kolom pencarian. Sedangkan *rewritten query* merupakan pembagian input pencarian pengguna menjadi kata awal, prefiks, dan keterhubungannya dengan orang lain. Seperti contohnya saat input pencarian berupa “richard b” maka akan dibagi menjadi +NAME:richard +PREFIX:b +CONN:1 yang artinya semua orang yang mempunyai nama awal Richard dengan nama selanjutnya diawali huruf b dan keterhubungan dengan user lain: 1.



Gambar 3.5 Indexing

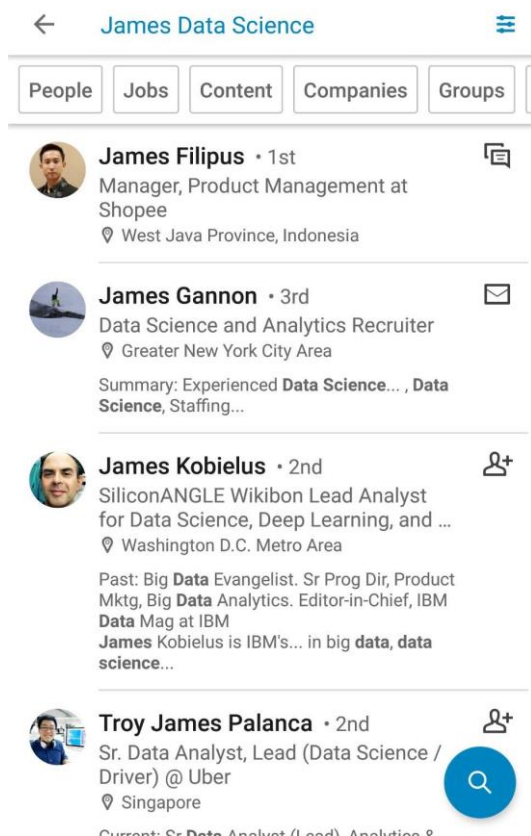
Sumber: <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses 13 Mei 2018 Pukul 06.00)

IV. ANALISIS

Pencarian yang dilakukan pada LinkedIn menggunakan dasar-dasar pencocokan string dan dilengkapi dengan metode atau cara-cara lain agar hasil yang dikeluarkan lebih relevan. Untuk menentukan relevansi dan agar kontekstual dilakukan beberapa cara seperti *query tagging*, *query autocomplete*, *query suggestions*, dan *indexing*.

Kombinasi keempat metode tersebut dalam pencarian pada LinkedIn menghasilkan keluaran yang akurat karena adanya *tagging* yang dapat mengenali makna dari kata berdasarkan input pengguna, *autocomplete* yang melengkapi input pengguna sehingga bisa mengatasi masalah pengejaan atau *typo*, *suggestion* yang dapat memberikan saran input pengguna agar relevan dan sesuai konteks yang diinginkan, dan *indexing* yang dapat membagi input pengguna menjadi beberapa bagian kata dan keterhubungannya dengan koneksi pengguna sehingga hasilnya terpersonalisasi.

Seperti contoh berikut dimana penulis mencari dengan kata kunci “James Data Science”. Berdasarkan empat metode yang sudah dijelaskan sebelumnya, pencocokan string tidak mutlak hanya mencocokkan dengan nama pengguna, tetapi dilakukan parsing dan dilakukan pencarian sesuai dengan konteksnya masing-masing. Hasil pencarian mengeluarkan nama orang lain yang mengandung prefiks “James” dan “Data Science” sebagai pekerjaan atau deskripsi profilnya serta terurut berdasarkan relevansi.



Gambar 4.1 Hasil Pencarian pada LinkedIn
Sumber: Dokumen Pribadi Penulis

V. KESIMPULAN

Algoritma pencocokan string merupakan hasil pengembangan berkelanjutan agar menghasilkan hasil pencarian yang sesuai. Pengguna mendapatkan hasil pencarian instan yang kontekstual dan relevan, dimana hasil *navigational search* menghasilkan keluaran yang spesifik dengan pola dan teks pencarian yang sama.

Fitur pencarian lain yaitu *exploratory search* dan *instant search* juga membantu pengguna mencari sesuai kategori dan mendapatkan hasil langsung yang sesuai. Pada *exploratory search*, algoritma pencocokan string tidak hanya mencocokkan pada nama pengguna LinkedIn saja, tetapi juga pada deskripsi profil dan daftar pekerjaan yang tercantum pada setiap penggunaannya.

Pengguna juga dapat meminimalisir kesalahan dari input pencarian dengan adanya fitur *query autocomplete* dan *search suggestions*. Fitur tersebut dapat melengkapi pola atau input pencarian pengguna walaupun belum lengkap penulisannya serta memberikan beberapa pilihan yang mungkin sesuai dengan maksudnya pengguna, karena suatu kata bisa memiliki banyak makna dan konteks.

Algoritma pencocokan string pada LinkedIn juga mengkombinasikan beberapa metode seperti *query tagging*, *query completion*, *query suggestions*, dan *indexing*. Melalui kombinasi metode-metode tersebut, pencocokan string pada fitur pencarian LinkedIn dilakukan secara terpisah per kata sehingga kontekstual dan lebih efektif. Suatu string yang dikategorikan sebagai nama orang hanya dicari di basis data pada entitas nama orang.

Metode *indexing* juga membantu pengguna dalam pencarian agar pencocokan string dilakukan *parsing* terlebih dahulu yang membagi nama yang sudah lengkap yaitu yang berakhir sebelum spasi dan prefiks selanjutnya yang belum lengkap setelah spasi. Pencocokkan string dengan metode *indexing* selain menghasilkan banyak kemungkinan yang sesuai, hasil keluarannya juga akan diurutkan sesuai derajat koneksi.

Pada pengembangan pencarian selanjutnya yang menggunakan algoritma pencocokan string, dapat dilakukan *personalized completion* pada pencarian yang didasarkan pada pencarian yang sudah dilakukan pada kesempatan sebelumnya. *Multi-entry suggestions* juga dapat ditambahkan menjadi salah satu fiturnya sehingga sugesti pencarian kepada pengguna tidak didasarkan pada satu kata saja, tetapi bisa dari banyak pilihan kata.

UCAPAN TERIMA KASIH

Puji dan syukur ke hadapan Tuhan Yesus Kristus yang hanya karena berkat dan rahmat-Nya penulis dapat menyelesaikan makalah ini dengan baik dan tepat waktu. Penulis juga mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan serta bantuan dalam menyelesaikan makalah yang berjudul “Penerapan Algoritma Pencocokan String pada Pencarian di Aplikasi LinkedIn”.

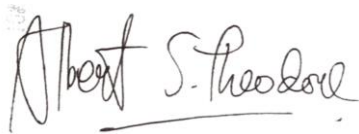
REFERENSI

- [1] Anany Levitin, "Design and Analysis of Algorithm", Pearson Education Inc. 2012.
- [2] <http://www.informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/stima17-18.htm#SlideKuliah> (Diakses pada tanggal 10 Mei 2018 Pukul 19.00)
- [3] <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin>
- [4] <https://www.linkedin.com/help/linkedin/answer/4447/linkedin-search-relevance-people-search?lang=en> (Diakses pada tanggal 10 Mei 2018 Pukul 20.00)
- [5] <https://www.slideshare.net/abhimanyulad/fast-lenient-and-accurate-building-personalized-instant-search-experience-at-linkedin> (Diakses pada tanggal 11 Mei 2018 Pukul 18.00)

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 14 Mei 2018

A handwritten signature in black ink, appearing to read "Albert S. Theodore". The signature is fluid and cursive, with a horizontal line underneath the name.

Albert Sahala Theodore - 13516022