

BAB I

PENDAHULUAN

1.1. Latar Belakang

Seiring bertambahnya umur seseorang, maka jumlah relasi orang tersebut akan semakin meningkat. Pada umumnya, keterbatasan ingatan manusia dapat menyebabkan hilangnya sebagian informasi relasi. Hal ini tentu sangat menyulitkan apabila ingin mengetahui semua informasi relasi. Karena itu, beberapa upaya diusahakan manusia untuk melengkapi informasi relasinya. Upaya-upaya tersebut adalah mencatat informasi relasi ke dalam bentuk *hardcopy* maupun *softcopy*. Pencatatan informasi relasi yang diupayakan manusia dapat berupa pencatatan informasi seangkatan di buku tahunan SMA X tahun akademik 2012 atau *database* mahasiswa IT 2015. Namun masalah lain ditemukan yakni informasi individu yang telah diarsip belum tentu relevan sesuai dengan perkembangan zaman. Hal ini bisa saja terjadi karena pindahan tempat tinggal, ganti nomor *handphone*, atau perubahan lainnya. Akibatnya informasi yang telah disimpan menjadi tidak *valid*. Akhirnya buku tahunan dan *database* tidak relevan dengan informasi individu saat ini. Dengan begitu, manusia tidak akan dapat menemukan informasi relasi yang relevan.

Maka manusia berusaha mencari informasi individu dengan mengontak almamater atau kerabat untuk mendapatkan informasi yang *valid*. Adapun informasi yang didapatkan dari almamater dan kerabat belum tentu *valid* karena perubahan informasi bisa jadi terjadi kapan saja.

Masalah ini akan semakin serius ketika diadakan acara reuni alumni suatu kelompok. Informasi relasi alumni suatu kelompok tertentu sulit ditemukan. Hal ini disebabkan oleh rentang reuni dengan tahun lulus tergolong cukup lama, yakni ± 10 tahun. Akhirnya tidak semua alumni dapat mengikuti reuni karena pesan yang gagal tersampaikan.

Ketersediaan mesin temu balik, seperti: Google, Yahoo! ataupun Bing, memberikan alternatif untuk menelusuri informasi tentang seseorang. Namun, cara ini adalah dirasakan kurang efektif. Belum tentu semua alumni menaruh kontak informasi di *internet*. Selain itu perkembangan informasi yang pesat

membuat banyak *web page* di dunia maya. Keanekaragaman *web page* yang tersebar di *internet* membuat pencarian menjadi lebih sulit. Salah satu penyebabnya adalah pesatnya perkembangan dunia, hingga saat ini populasi dunia mencapai 7.3 miliar jiwa¹.

Permasalahan yang lain yang dapat muncul ialah mesin temu balik informasi tidak memberikan solusi optimal. Misalnya kita ingin mengetahui nomor kontak pribadi dari Evelyn tapi hasil pencarian dari mesin temu balik adalah alamat kuburan dari Evelyn Larissa. Hal inilah yang membuat pencarian dengan mesin temu balik bukan merupakan solusi optimal. Hasil pencarian mesin temu balik akan menjadi optimal jika subjek yang dicari adalah seorang yang terkenal di dunia maya atau *public figure*. Jika subjek adalah orang yang cukup populer di *internet*, maka *top-10 ranking search result* akan mengarah pada informasi individu yang *valid*. Hal ini tidak menutup kemungkinan bahwa akan terdapat hasil pencarian yang tidak diharapkan (tidak mengandung kontak pribadi subjek) masuk dalam *top-10*. Kemungkinan hasil pencarian akan lebih meleset untuk subjek yang kurang atau tidak terkenal (*unknown person*). Misalnya saat mencari kontak Ibu Atun, pembantu Agnes Monica. Hasil pencarian mesin temu balik untuk Ibu Atun yang dihasilkan adalah informasi tentang Agnes Monica, bukan informasi Ibu Atun. Hal ini dikarenakan entitas subjek terkait dengan *popular person* dan mesin temu balik akan cenderung menampilkan informasi individu yang lebih terkenal. Tentunya akan sulit menemukan informasi individu menggunakan mesin temu balik terutama dalam kasus pencarian *unknown person*. Adapun permasalahan sulitnya ditemukan informasi *unknown person* bila nama tersebut di-*share* oleh selebriti atau figur historis, karena hasil pencarian mesin temu balik akan didominasi oleh individu tersebut, hal ini membuat pencarian informasi individu menjadi sulit (Delgado, 2014).

Berdasarkan fakta di lapangan, muncul pula permasalahan ambiguitas nama. Pencarian informasi individu akan semakin sulit mengingat banyaknya populasi manusia di dunia dan sulit menghasilkan hasil pencarian yang relevan, yakni mengarah kepada subjek yang benar-benar diinginkan. Perkembangan dunia yang teramat pesat dan akan bertambah pesat terus-menerus. Perkiraan populasi

¹ Menurut survei *United Nation* (Desember 2015)

dunia pada tahun 2100 adalah 11.2 miliar jiwa². Bertambah maraknya penggunaan media sosial, seperti: Facebook, Twitter, Google+, LinkedIn, mengakibatkan mesin temu balik memberikan hasil pencarian dari beberapa *profile* yang dimiliki oleh individu berbeda namun memiliki kesamaan nama, dan berpotensi untuk dimanfaatkan dalam proses penyaringan nama yang lebih relevan.

Menindaklanjuti persoalan-persoalan yang ditemukan dan potensi yang disediakan oleh berbagai media sosial, maka diusulkanlah sebuah perangkat lunak untuk mempermudah pencarian individu. Perangkat lunak ini akan mengkombinasikan hasil dari mesin temu balik, analisis data, *filtering*, dan penentuan probabilitas tingkat kepuasan pengguna terhadap berbagai metode ekspansi kueri. Selain itu, diterapkan algoritma klusterisasi *Unsupervised Person Name Disambiguator (UPND)* guna membentuk kluster-kluster orang yang berbeda dari satu nama yang ingin dicari.

Penelitian ini juga dibuat dengan latar belakang bahwa *database* alumni Universitas Kristen Maranatha yang kurang lengkap. Akibatnya, banyak data alumni yang belum diketahui pekerjaannya, nomor telepon valid, masa tunggu sebelum bekerja setelah lulus, dan lain-lain. Karena ketidaklengkapan *database* alumni ini, maka akan sulit terjalin hubungan timbal balik antar universitas dengan alumni.

Permasalahan yang lain yang dapat muncul ialah *database* alumni mungkin lengkap, tapi informasinya tidak valid. Akibatnya ketika disebar survei, tidak akan didapatkan hasilnya. Hasil survei bagi suatu institusi adalah hal yang sangat penting. Survei tersebut dianalisis kemudian dapat dilakukan evaluasi terhadap suatu universitas atau program studi. Berdasarkan hasil dari *tracer study* antara tahun 2009-2013, respons balik yang didapatkan oleh UKM berada pada kisaran 4% dari total mahasiswa yang dikirimkan survei (diuraikan dalam Pertemuan PPE 29 Januari 2015). Salah satu permasalahan utama yang teramati adalah banyaknya surel yang *bouncing* (tidak terkirim karena sudah tidak aktif) dan walaupun survei diterima ada semacam keenganan untuk mengirimkan kembali kepada UKM. Padahal kegiatan *tracer study* merupakan salah satu butir penting yang dinilai dalam proses akreditasi institusi ataupun program studi.

² www.un.org/apps/news/story.asp?NewsID=51526#.VmocL4n-lm9

Terlebih informasi mengenai pekerjaan merupakan informasi yang bersifat pribadi. Beberapa orang tidak ingin memberitahukan kepada UKM mengenai riwayat pekerjaannya. Oleh karena itu, perlu dipikirkan cara bagaimana mendapatkan informasi alumni, terutama pekerjaan dan kapan bekerja tanpa menunggu hasil survei dari alumni.

Sepengetahuan peneliti, jurnal-jurnal yang membahas mengenai pelacakan data orang dan penanganan disambiguasi nama, yaitu sebagai berikut.

- 1) Jurnal “*A Data Driven Approach for Person Name Disambiguation in Web Search Results*” yang ditulis Delgado, Martinez, Fresno, dan Montalvo membahas tentang disambiguasi nama di hasil pencarian.
- 2) Jurnal “*Result Disambiguation in Web People Search*” yang ditulis oleh Berendsen, Kovachec, Nastou, Rijke, dan Weerkamp membahas tentang hasil disambiguasi nama.
- 3) Jurnal “*On The Modeling of Entities for Ad-Hoc Entity Search in The Web of Data*” yang ditulis Neumayer, Balog, dan Nørvåg membahas tentang model entitas hasil pencarian.
- 4) Jurnal “*A Review of The Technologies and Methods in Profiling and Profile Classification*” yang ditulis Pampapathi, Mirkin, dan Levene membahas tentang metode klasifikasi profil seseorang.

Diharapkan penelitian dan pembahasan “Prediksi Bidang Kerja dan Asal Fakultas untuk Alumni Melalui Klusterisasi dan Klasifikasi Kemunculan Kata” dapat menghasilkan konsep, model, alur proses, prinsip, atau skema yang dapat diaplikasikan oleh pengguna penelitian. Ditinjau dari segi praktikal diharapkan penelitian ini dapat membantu dalam proses *tracer study* agar dapat respons timbal balik dari pihak universitas maupun alumni. Dalam konteks penelitian ini, *user* yang dapat memanfaatkannya adalah semua orang yang ingin mencari data alumni maupun pihak perusahaan yang sedang mencari pegawai. Ditinjau dari segi teoretis, diharapkan hasil penelitian ini memiliki bobot dan relevansi untuk digunakan sebagai acuan, literatur, atau referensi oleh kalangan akademis, khususnya pemerhati dan pembelajar ilmu teknik informatika.

Sesuai dengan jurnal dan tulisan ilmiah yang telah ditinjau di atas penelitian dan pembahasan tentang “Prediksi Bidang Kerja dan Asal Fakultas

untuk Alumni Melalui Klusterisasi dan Klasifikasi Kemunculan Kata” yang dilakukan peneliti memiliki kekhasan, kemuktahiran, dan perspektif yang berbeda dengan penelitian sejenis sebelumnya, yaitu penelitian ini mencoba menebak jenis bidang kerja, dan melacak informasi-informasi dari orang tersebut, seperti nomor telepon, *email*, instansi pekerjaan, domisili, dan lain-lain. Tidak hanya menebak informasi individu, namun penelitian ini juga menyandingkan informasi individu yang didapat dalam media sosial. Ditinjau dari data yang digunakan sumber-sumber di atas, peneliti mempergunakan data yang khas, yaitu data-data alumni Universitas Kristen Maranatha.

Jika permasalahan cacatnya basis data alumni tidak dicarikan solusinya dan dibiarkan berlarut-larut, apakah kita rela menyaksikan kegiatan *tracer study* tidak berjalan dengan baik? Bagaimana tanggung jawab moral kita sebagai kalangan akademis yang memiliki kompetensi di bidang teknologi informasi jika permasalahan survei yang sedikit direspons tidak kunjung terselesaikan? Apa yang akan kita lakukan jika akreditasi universitas maupun program studi menurun hanya karena cacatnya basis data alumni? Dengan demikian, penelitian dan pembahasan “Prediksi Bidang Kerja dan Asal Fakultas untuk Alumni Melalui Klusterisasi dan Klasifikasi Kemunculan Kata” menjadi penting dan mutlak dilakukan.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas, diketahui bahwa pencarian informasi individu memerlukan pendekatan yang lebih dalam dari sekedar merangkumkan hasil dari mesin temu balik. Selain itu Oleh karena itu, akan dibatasi dan dirumuskan pokok-pokok persoalan yang akan dibahas, dianalisis, diuji, dan dijawab dalam penelitian, yaitu sebagai berikut.

1. Bagaimana menentukan relevansi sebuah tautan terkait dengan nama individu?
2. Bagaimana mengatasi ambiguitas nama dari individu yang berbeda?
3. Bagaimana menentukan metode terbaik dalam prediksi bidang kerja dan asal fakultas individu?

4. Bagaimana menyertakan informasi individu yang terambil dari media sosial yang valid pada sebuah aplikasi *web*?

1.3. Tujuan

Sesuai dengan pokok-pokok permasalahan yang telah dibatasi dan telah dirumuskan dalam rumusan masalah di atas, berikut ini akan dijabarkan garis-garis besar hasil pokok yang ingin diperoleh setelah permasalahan dibahas dan diuji yang dapat terkategori sebagai luaran penelitian yaitu sebagai berikut:

1. Pengembangan perangkat lunak dapat mengidentifikasi suatu tautan yang dianggap relevan dengan suatu nama individu;
2. Pengembangan perangkat lunak yang mampu mengatasi ambiguitas nama dari individu yang berbeda dengan memanfaatkan teknik ekspansi kueri melalui kata-kata yang sudah terdefinisi, seperti: *email*, nomor telepon atau dengan menggunakan kemunculan kata-kata yang dominan dalam kandidat tautan yang berkualitas baik.
3. Pengembangan perangkat lunak yang dapat menebak bidang kerja dan asal fakultas individu dengan mengidentifikasi metode terbaik yang memanfaatkan klasifikasi, pembentukan model, maupun kemunculan kata-kata yang dianggap relevan dengan bidang kerja tertentu.
4. Pengembangan perangkat lunak yang dapat menyertakan informasi individu yang terambil dari kluster media sosial *link* yang valid dan menampilkannya pada sebuah aplikasi *web*.

1.4. Ruang Lingkup

Menimbang bahwa permasalahan ini cukup kompleks dan tergolong ilmu baru, maka terdapat ruang lingkup pengembangan, antara lain sebagai berikut:

1. Perangkat lunak tidak menghasilkan hasil pencarian yang bersumber pada situs media sosial. Hal ini dikarenakan untuk mengambil informasi ini mengharuskan *login* ke media sosial tersebut.
2. Perangkat lunak berbasis *desktop application* dan *web application* terkoneksi dengan *internet*.

3. Dalam laporan ini memuat hasil *data extraction* dari *search engine* Google, yang diakses mulai bulan Juli-Oktober 2015.

1.5. Sistematika Penulisan

Dalam tulisan akademik ini dimuat bab demi bab, subbab demi subbab, dan pasal demi pasal yang memiliki keterkaitan, keterpaduan (kohesi), dan kesinambungan (koherensi).

BAB I PENDAHULUAN

Bab ini menjelaskan latar belakang, rumusan masalah, tujuan, batasan masalah, dan sistematika pembahasan. Dalam bab ini pembaca dapat mengetahui gambaran kasar mengenai isi laporan ini.

BAB II LANDASAN TEORI

Bab ini berisi seluruh teori yang dipakai dan relevan dengan pengembangan aplikasi yang dibahas dalam laporan ini.

BAB III ANALISIS DAN DESAIN

Dalam bab III dijabarkan secara terperinci dan mendetail tentang cara kerja dan alat kerja penelitian serta deskripsi data baik secara kualitas, kuantitas, dan dasar atau prinsip pemilihan data yang digunakan. Bab ini terdiri atas Analisis, Gambaran Keseluruhan dan Perancangan Desain Perangkat Lunak. Bab ini memuat diagram alir sistem dari aplikasi, meliputi *Flowchart*, *Use Case*, *Class Diagram*, *Sequence Diagram*, Rancangan *User Interface*, dan penjelasan aplikasi.

BAB IV PENGEMBANGAN PERANGKAT LUNAK

Dalam bab ini menjelaskan secara lengkap implementasi seluruh hasil rancangan di Bab III, mulai dari implementasi *class*, simpanan data sampai dengan antar muka.

BAB V TESTING DAN EVALUASI SISTEM

Bab ini berisi *testcase* dan pelaksanaan pengujian akurasi penebakan bidang kerja dan asal fakultas.

BAB VI KESIMPULAN DAN SARAN

Bab ini terdiri dari simpulan mengenai aplikasi yang telah dibuat, yakni hasil pokok yang diperoleh setelah masalah dibahas dan diselidiki. Selain itu, bab ini berisi saran-saran untuk perbaikan serta penyempurnaan aplikasi ini dan langkah-langkah apa yang dapat ditempuh pengguna penelitian sebagai konsekuensi atau implikasi dari simpulan.

