



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk mengubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Dalam buku yang berjudul *Faster* (1999), James Gleick menjelaskan bahwa manusia pada umumnya menghabiskan 16 menit setiap harinya untuk mencari barang yang salah ditaruh, yang dapat diartikan hampir setahun dari seumur hidup manusia. Lalu pada tahun 2016, sebuah survey yang dilakukan oleh Pixie Technology memperlihatkan bahwa manusia masih salah menaruh barangnya, memakan waktu sekitar 2.5 hari setiap orangnya dalam setahun dengan rata-rata menghabiskan waktu 5 menit 20 detik setiap kali mencari barang. Hal tersebut merugikan sekitar 2.7 miliar dollar Amerika tiap tahunnya untuk menggantikan barang yang hilang.

Dengan adanya perkembangan zaman, tidak hanya barang yang berbentuk fisik yang dapat salah ditaruh, berkas digital juga sering disimpan di tempat yang salah. Aplikasi pencarian berkas sangatlah membantu pengguna dalam mencari berkas yang ada di dalam perangkat pengguna. Dalam artikel yang ditulis oleh Batchelor, dkk. (2018) Windows telah menyediakan aplikasi pencarian berkas dalam Windows Explorer yang bernama Windows Desktop Search (WDS), tetapi fitur pencarian berkas WDS tidak memiliki fitur untuk mendeteksi *typo* atau mencari berkas dengan arti yang mirip dengan *keyword* yang diinput oleh *user*.

Salah penamaan dapat menjadi masalah yang cukup besar dalam pencarian berkas di komputer, misalnya berkas yang bernama “Me” dicari dengan keyword “I”. Walaupun kata “Me” dan “I” memiliki arti yang mirip, tetapi jelas kedua kata

tersebut berbeda. Penelitian yang dilakukan oleh Derry Jatnika, Moch Arif Bijaksana, dan Arie Ardiyanti Suryani (2019) menjelaskan bahwa masalah tersebut dapat ditangani dengan algoritma Word2Vec dikarenakan Word2Vec memiliki fitur *semantic similarity*.

Library fastText adalah salah satu library Word2Vec yang memiliki *model* yang berisikan kata-kata berbahasa Inggris yang sudah dilatih agar kata-kata yang memiliki arti mirip dapat dideteksi. Dalam *model* yang disediakan oleh fastText, berisikan kata-kata yang sudah diproyeksikan ke dalam bentuk vektor. Hal tersebut memungkinkan aplikasi untuk mencari *semantic similarity* dari kata yang dimasukkan oleh pengguna. (Mikolov, 2013)

Untuk mendeteksi typo, algoritma yang dibuat oleh direktur penelitian Google, Peter Norvig di tahun 2007 yang menggunakan metode Damerau-Levenshtein Distance dapat dimodifikasi agar dapat menggunakan model dari fastText untuk menjadi kamus spell checking agar aplikasi dapat membetulkan jika ada typo yang dilakukan pada aplikasi.

Kelebihan dari FastText yang ditambah dengan algoritma dari Peter Norvig dapat digunakan untuk membangun aplikasi pencarian berkas berbasis algoritma Word2Vec yang dapat mencari berkas dengan keyword yang memiliki arti yang mirip dengan nama berkas dan melakukan spell checking, menyelesaikan masalah yang tidak dapat dilakukan oleh pencarian berkas pada Windows Explorer.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan di atas, dapat diambil rumusan masalah, yaitu:

1. Bagaimana cara membangun aplikasi pencarian berkas menggunakan algoritma Word2Vec?
2. Apakah algoritma Word2Vec dapat melakukan pencarian berkas yang memiliki arti mirip dengan *input*?
3. Apakah algoritma oleh Peter Norvig dapat melakukan *spell checking* berdasarkan *model* dari fastText?

### 1.3 Batasan Masalah

Agar penelitian ini tidak melebar dan melenceng dari tujuan awal yang diharapkan, maka ditentukan batasan masalah dari penelitian ini:

1. Aplikasi hanya dibuat untuk mencari nama berkas yang artinya mirip dengan *input* yang dimasukkan oleh pengguna.
2. Aplikasi hanya dapat membetulkan *minor typo* pada *keyword* yang dimasukkan oleh pengguna.
3. Aplikasi hanya dapat memroses *keyword* dan nama berkas yang berbahasa Inggris.
4. Aplikasi dibuat hanya untuk *platform* desktop dengan sistem operasi Windows.

### 1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk merancang aplikasi pencarian berkas menggunakan algoritma Word2Vec.
2. Mencari tahu keakurasian algoritma Word2Vec dalam mencari sinonim dari *keyword* yang dimasukkan oleh pengguna.

3. Menggunakan *model* dari fastText sebagai kamus *spell checking* algoritma Peter Norvig.

### **1.5 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Dapat memudahkan pengguna untuk menemukan berkas yang ada di dalam komputer.
2. Mengetahui apakah algoritma Word2Vec dapat digunakan untuk aplikasi pencarian berkas.
3. Dapat berguna untuk penelitian selanjutnya, baik pada permasalahan yang sama, atau berbeda.

### **1.6 Sistematika Penulisan**

Sistematika penulisan skripsi ini terdiri dari lima bab, yaitu sebagai berikut.

1. **BAB I PENDAHULUAN**

Bab ini terdiri dari Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, dan Sistematika Penulisan.

2. **BAB II LANDASAN TEORI**

Bab ini menjelaskan landasan teori dari penelitian yang dilakukan, antara lain *semantic similarity*, *natural language processing*, Word2Vec, Damerau-Levenshtein distance, dan fastText.

3. **BAB III METODOLOGI DAN PERANCANGAN APLIKASI**

Bab ini berisikan metodologi penelitian yang digunakan dan perancangan aplikasi.

4. **BAB IV UJI COBA DAN EVALUASI**

Bab ini berisi hasil uji coba dan evaluasi aplikasi yang telah dibangun.

## 5. BAB V KESIMPULAN DAN SARAN

Bab ini terdiri atas simpulan dari hasil penelitian dan saran untuk pengembangan aplikasi pada penelitian selanjutnya.