

Proposal Proyek : Information Retrieval

Proyek ini adalah sebuah sistem information retrieval untuk mengelola dan menampilkan data publikasi ilmiah. Kami akan mengimplementasikan metode text mining dengan menggunakan TF-IDF dan cosine similarity untuk mengevaluasi relevansi dokumen dengan query yang diberikan. Tujuan akhir dari proyek ini adalah untuk menyediakan layanan pencarian yang mudah digunakan dan efektif bagi pengguna yang membutuhkan informasi tentang publikasi ilmiah.

Tipe Proyek

Tipe proyek yang kami kerjakan adalah proyek pengembangan sistem information retrieval. Kami akan menggunakan dataset dokumen dalam bahasa Indonesia dan mengembangkan model yang dapat melakukan pencarian dokumen berdasarkan relevansi dengan kata kunci yang diberikan oleh pengguna. Proyek ini melibatkan implementasi metode text mining seperti preprocessing data, term weighting, dan algoritma cosine similarity. Selain itu, kami juga akan melakukan evaluasi performa model yang dikembangkan dengan menggunakan metrik precision, recall, dan F1 score.

Harapan Kami

kami berharap bahwa proyek kami akan dapat menghasilkan sebuah sistem information retrieval yang dapat memberikan hasil pencarian yang akurat dan relevan bagi penggunanya. Kami juga berharap dapat mempelajari lebih lanjut tentang teknik-teknik text mining dan information retrieval, serta mengembangkan keterampilan kami dalam melakukan analisis data dan implementasi algoritma.

Finding Existing Research and Dataset

Kami menggunakan penelitian dan dataset yang sudah ada untuk membantu dalam pengembangan sistem information retrieval kami. Penelitian mengenai penggunaan metode TF-IDF dan LSA memberikan pemahaman tentang konsep dasar dalam text mining dan indexing dokumen. Dataset yang digunakan juga dapat membantu dalam memperluas kategori dan jumlah dokumen yang dapat diakses oleh sistem kami. Dengan memanfaatkan sumber daya yang sudah ada, kami dapat mengoptimalkan pengembangan sistem information retrieval kami.

Tujuan proyek

- Definisi tugas kami adalah mengembangkan sistem pencarian informasi yang efektif dan akurat. Kami akan menggunakan teknik Information Retrieval untuk membangun sistem pencarian ini. Tujuan akhir kami adalah untuk memungkinkan pengguna kami menemukan informasi yang relevan dengan cepat dan mudah.
- Kami akan menggunakan dataset dokumen dalam bahasa Inggris yang terdiri dari berbagai jenis sumber daya, seperti artikel jurnal, korpus teks, dan dokumen terindeks lainnya. Kami akan memastikan bahwa dataset yang kami gunakan sesuai dengan kebutuhan tugas kami, termasuk jumlah dokumen yang sesuai dan kualitasnya.
- Metric evaluasi kami akan berfokus pada akurasi dan kecepatan sistem kami dalam menemukan informasi yang relevan. Kami akan menggunakan F1 score dan Mean Average Precision (MAP) sebagai evaluasi utama kami, karena kedua metrik tersebut dapat memberikan informasi yang berguna tentang kinerja sistem kami.
- Keberhasilan proyek kami akan diukur oleh seberapa efektif sistem kami dalam menemukan informasi yang relevan dan memberikan hasil pencarian yang sesuai dengan permintaan pengguna. Kami berharap untuk mencapai F1 score dan MAP yang tinggi, yang menunjukkan bahwa sistem kami dapat memberikan hasil pencarian yang akurat dan tepat waktu. Selain itu, kami juga akan memperhatikan feedback dari pengguna dan meningkatkan sistem kami berdasarkan masukan mereka.

Kesimpulan

Sebagai kesimpulan, Kami ingin mengatakan bahwa proyek ini adalah tentang Information Retrieval dengan menggunakan pendekatan NLP. Kami telah mendefinisikan tujuan dan tugas yang jelas, yaitu untuk membangun sistem pencarian yang dapat mengambil dan menyajikan informasi yang relevan dari dataset yang telah kami tentukan. Kami juga telah menentukan metrik evaluasi kami, yaitu precision dan recall, serta mengidentifikasi kesuksesan proyek sebagai mencapai nilai-nilai tertentu untuk metrik tersebut. Kami berharap bahwa dengan memenuhi semua elemen ini dan bekerja sama sebagai tim, kami akan berhasil dalam membangun sistem pencarian yang efektif dan bermanfaat.

Referensi

<https://ir-datasets.com/vaswani.html>

<https://github.com/Mingmong123/NLP-UAS>

<https://colab.research.google.com/drive/1AizUfUtzFvk0tgKrrBAT8iKhZ5oD8RzH?usp=sharing>

http://ir.dcs.gla.ac.uk/resources/test_collections/npl/npl.tar.gz

http://ir.dcs.gla.ac.uk/resources/test_collections/npl/