Documentation for P05

How to open:

Compile backend with Lucene, optionally add a command line argument to specify how many documents are being used, default being 1000. It'll run a RESTful server on http://localhost:1234/ with the only route being /search, taking one query parameter, *query*, which is the query the user searches for.

In order to run the Backend.jar the following command can be issued:

```
java -jar Backend.jar [num_docs]
```

Where [num_docs] is an optional argument specifying the number of documents used, defaulting to 1000.

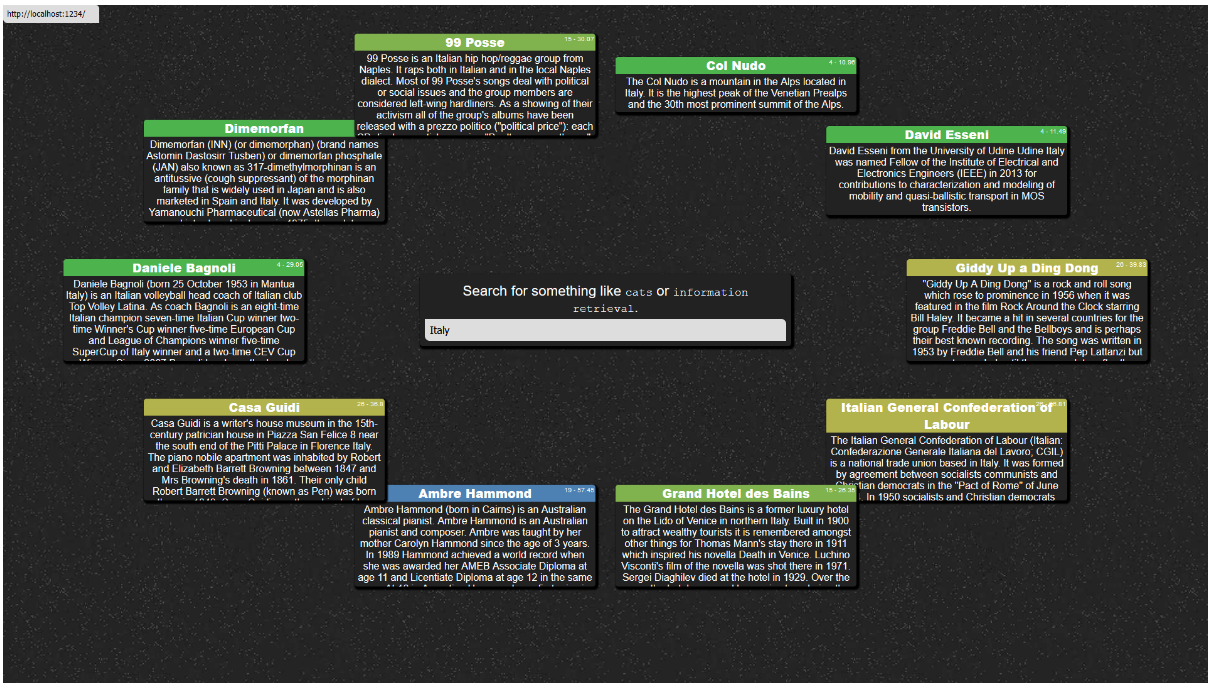
For the frontend install the necessary packages using npm, by running *npm install*, then run using *npm run serve* to start the web server. The website is accessible under http://localhost:8080/

Frontend:

The Frontend is built using *Vue3*. It sends requests to the backend via *Axios.* These requests' answers follow this scheme:

```
{
        query: {
                classification: number,
                title: string,
                vector: number[]
        },
        result: {
                title: string,
                context: string,
                class: number,
                vector: number[]
        }[]
}
```

Using the given vectors of the query and results the distance is calculated using Euclidean distance and displayed in the top right of the cards. Each class has a different color to easily distinguish them if multiple classes are returned. Each card shows the title of the document, a preview and the distance to the query. Clicking on a card will open a new tab with the Wikipedia article for the search result, by constructing the URL of the article using the title of the document. E.g. *Albert Einstein* opens https://en.wikipedia.org/wiki/Albert_Einstein.



http://localhost:1234/

99 Posse 15 - 30.07

99 Posse is an Italian hip hop/reggae group from Naples. It raps both in Italian and in the local Naples dialect. Most of 99 Posse's songs deal with political or social issues and the group members are considered left-wing hardliners. As a showing of their activism all of the group's albums have been released with a prezzo politico ("political price"): each

Col Nudo 4 - 10.96

The Col Nudo is a mountain in the Alps located in Italy. It is the highest peak of the Venetian Prealps and the 30th most prominent summit of the Alps.

Dimemorfan

Dimemorfan (INN) (or dimemorphan) (brand names Astomin Dastosirr Tusben) or dimemorfan phosphate (JAN) also known as 317-dimethylmorphinan is an antitussive (cough suppressant) of the morphinan family that is widely used in Japan and is also marketed in Spain and Italy. It was developed by Yamanouchi Pharmaceutical (now Astellas Pharma)

David Esseni 4 - 11.49

David Esseni from the University of Udine Udine Italy was named Fellow of the Institute of Electrical and Electronics Engineers (IEEE) in 2013 for contributions to characterization and modeling of mobility and quasi-ballistic transport in MOS transistors.

Daniele Bagnoli 4 - 29.05

Daniele Bagnoli (born 25 October 1953 in Mantua Italy) is an Italian volleyball head coach of Italian club Top Volley Latina. As coach Bagnoli is an eight-time Italian champion seven-time Italian Cup winner two-time Winner's Cup winner five-time European Cup and League of Champions winner five-time SuperCup of Italy winner and a two-time CEV Cup

Search for something like `cats` or `information retrieval.`

Italy

Giddy Up a Ding Dong 26 - 39.83

"Giddy Up A Ding Dong" is a rock and roll song which rose to prominence in 1956 when it was featured in the film Rock Around the Clock starring Bill Haley. It became a hit in several countries for the group Freddie Bell and the Bellboys and is perhaps their best known recording. The song was written in 1953 by Freddie Bell and his friend Pep Lattanzi but

Casa Guidi 26 - 36.8

Casa Guidi is a writer's house museum in the 15th-century patrician house in Piazza San Felice 8 near the south end of the Pitti Palace in Florence Italy. The piano nobile apartment was inhabited by Robert and Elizabeth Barrett Browning between 1847 and Mrs Browning's death in 1861. Their only child Robert Barrett Browning (known as Pen) was born

Italian General Confederation of Labour 26 - 36.81

The Italian General Confederation of Labour (Italian: Confederazione Generale Italiana del Lavoro; CGIL) is a national trade union based in Italy. It was formed by agreement between socialists communists and Christian democrats in the "Pact of Rome" of June In 1950 socialists and Christian democrats

Ambre Hammond 19 - 57.45

Ambre Hammond (born in Cairns) is an Australian classical pianist. Ambre Hammond is an Australian pianist and composer. Ambre was taught by her mother Carolyn Hammond since the age of 3 years. In 1989 Hammond achieved a world record when she was awarded her AMEB Associate Diploma at age 11 and Licentiate Diploma at age 12 in the same

Grand Hotel des Bains 15 - 26.35

The Grand Hotel des Bains is a former luxury hotel on the Lido of Venice in northern Italy. Built in 1900 to attract wealthy tourists it is remembered amongst other things for Thomas Mann's stay there in 1911 which inspired his novella Death in Venice. Luchino Visconti's film of the novella was shot there in 1971. Sergei Diaghilev died at the hotel in 1929. Over the

Backend:

From the Wikipedia dataset topics are downloaded and composed into a csv file that contains these topics and their relevant body texts. After that tf and idf values are calculated for the documents. K-means-clustering algorithm is used in order to cluster documents. The initial clustering centers are initialized by selecting a random document. After that the euclidean distance between documents and central points are calculated and each document is clustered according to these distances. This process continues until clustering converges or a maximum amount of 100 iterations passed. The Spring Boot framework is used in order to create the web application and JSON serialization. After the “query” parameter is taken, its tf-idf vector and its distances to other documents is calculated. Finally, the JSON response which includes the query string and query vector and up to 10 closest document’s tf-idf vector, title, class and context is returned.