# Retrieval of multimedia and social media data

Group Memer: Ma Daci, Chen Zhaoheng, Wang Zijie, Huang Lishan, Yuan Fangxu.

For the group project of CSIT 6000P, we are choosing ‘retrieval of multimedia and social media data‘ as our topic to work on.

Retrieval of multimedia information has become an important topic since the development of the Internet has greatly expand the types of information that we transmit over the web. Users send numerous images, audios and videos on social media to share about their life, and researchers work to find a way to index and organize these data, and eventually extract semantic meaning and information from the data collected. Methods for multimedia information retrieval can be generally categorized into three types[1]: a) methods for multimedia information classification. b) methods for filtering multimedia information. c) methods for feature extraction and content description for multimedia information.

For our project, we have chosen four papers in related area. We plan to carry out intensive reading on the papers, and figure out the mechanism introduced respectively. We hope to obtain a general view of multimedia and social media data retrieval, as well as an inside view of the ideas from the selected papers. A formal project report will be formed introducing the latest and representative ideas on the topic as the out come of the project.

## Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis

This paper proposed an open source database management system for multimedia retrieval and following analysis. It is particularly designed to provide scalar and vector attributes support, forming a unified data and query model that is able to perform Boolean retrieval and nearest neighbor search.

Cottontail DB will organize data of the same type in columns, and it put columns of the same semantical meaning into a same entity, which is equivalent to a table in traditional DBMS. The three types supported in Cottontail DB are float vectors, complex numbers and vectors of complex numbers.

For the overall architecture of Cottontail DB, the query requests from applications will be redirected to the DDL(definition of entities) , DML(management of data) or DQL(queries) module due to the types of the requests. An execution plan will be generated by the Query Planner and will be conducted by the graph execution engine. Various indexing methods are supported by the storage engine of the Cottontail DB to speed up the query speed.

The Cottontail DB is implemented in Kotlin and it runs on the JVM. The DB utilizes gRPC to communicate with applications, allowing the system to be deployed across platforms and to be used by numerous applications easily.

Paper 2: vitrivr – A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections

This paper introduces a content-based multimedia retrieval system named vitrivr. In addition to metadata searching, the vitrivr system also provides content-based search that allows users to draw a sketch of the query image and motion paths. Instead of focusing on tag-based or keyword-based search, content-based search is more efficient than the user’s intention. Such an advanced retrieval technique in multimedia becomes the main reason for selecting this paper. Moreover, the database system of vitrivr, which is called ADAM, also is a main factor that attracts us. The ADAM combines both relational database and the vector space database, searching multimedia objects like searching vectors in a high-dimensional feature space and calculating the similarity between objects in the space. Thus, ADAM can retrieve the target content based on the similarity degree of the database. To provide a high-performance video retrieval, vitrivr is also equipped with a content-based search engine named Cineast. This search engine can retrieval video objects and provide easy extensions of other types of multimedia like images and audio. To achieve the searching ability, two different components were utilized by Cineast; the first part is extracting features from multimedia documents and metadata systems, and the second part is interpreting the input queries and retrieving related documents of the input. Combining these two components, Cineast can employ multitude of independent feature modules in parallel to improve the searching speed. Based on the advanced performance of ADAM and Cineast, this paper become one of the best choices for use to study multimedia retrieval.

Paper 3: Deep Learning for Content-Based Image Retrieval:A Comprehensive Study

This paper investigates a framework of deep learning for content-based image retrieval(CBIR) by applying deep learning method CNN for learning feature representations from image data, and conduct an extensive set of empirical studies for a variety of CBIR tasks. The reason why we choose this paper is that it gives us an understanding of how to retrieve multimedia and social media data in the field of deep learning. This paper details how to build a deep learning model and apply it to the new CBIR task, which caught our attention. The deep convolutional network contains two parts, the convolution layers and the fully connection layers. The researchers train this model on the ImageNet’s ILSVRC-2012 training set, which contains about 1.2 million images. After pre-training, the researchers discuss how to apply the trained CNNs from classification to CBIR tasks in ImageNet, they introduce three kinds of feature generalization schemes : direct representation, refining by similarity learning and refining by model retraining. In experiment, they evaluate the CBIR performance using different schemes and on diverse CBIR tasks. At last, they proved that with proper feature refining schemes, the deep learning feature representations consistently outperform conventional hand-crafted features on all datasets. We think this paper inspires us on how to apply deep learning to multimedia data retrieval, and gives a very detailed process, so we should study it.

Paper 4: Content Based Image Retrieval with LIRe

In this paper, an open-source library for content-based image retrieval called LIRe (Lucene Image Retrieval) is introduced. Based on the Lucene which is a text search engine using inverted indexing, LIRe extracts a number of low level features from the images for document indexing. More specifically, on the one hand, LIRe has provided lots of state of the art global features, such as *Color and edge directivity descriptor* [1], *The Tamura texture features* [2], etc. On the other hand, LIRe has offered numerous local features for extraction, like *Scale-Invariant Feature Transform* [3], *Maximally Stable Extremal Regions* [4], etc. Furthermore, all the extracted features are stored within the Lucene index and the well-known *bag of visual words* [5] approach, creating a visual sentence from a set of visual words, is applied based on the extracted local features, which allows to utilize the fast text-based retrieval mechanisms of Lucene for search. Moreover, there are two mechanisms in LIRe for faster search in feature spaces, which are the *Fastmap* [6], a fast method for dimensionality reduction, and the *Metric Spaces* [7], a method utilizing inverted lists to characterize data points in feature spaces. In summary, the reason why this paper is selected is that LIRe can be regarded as a great base line and an easy-to-use library in content-based image retrieval. Lots of techniques and methods in both multimedia indexing and searching are mentioned in this paper and are involved in this application. Therefore, it is an inspiring paper for learning the multimedia retrieval.

The papers we have selected cover all aspects of Retrieve of multimedia and social media data with corresponding advanced techniques and perspectives. The scope of our research does not only include the creation of image based retrieval databases on social media. We also cover innovative approaches to information retrieval based on natural language processing in social media and the creation of corresponding new databases. We believe that by understanding these papers we can gain a comprehensive knowledge of multimedia and social media information retrieval and its databases and deepen our understanding of new spatial databases

Responsibility:

Ma Daci: Responsible for the research and development of technologies related to paper 1 and 2, and the preparation of related reports.

Chen Zhaoheng: Responsible for the research and development of technologies related to paper 2 and 3, and the preparation of related reports.

Wang Zijie: Responsible for the research and development of technologies related to paper 3 and 4, and the preparation of related reports.

Huang Lishan: Responsible for the research and development of technologies related to paper 1 and 3, and the preparation of related reports.

Yuan Fangxu: Responsible for the research and development of technologies related to paper 2 and 4, and the preparation of related reports.

Ref

[1] <https://en.wikipedia.org/wiki/Multimedia_information_retrieval>

[2] Ralph Gasser, Luca Rossetto, Silvan Heller, and Heiko Schuldt. 2020. Cottontail DB: An Open Source Database System for Multimedia Retrieval

and Analysis. In Proceedings of the 28th ACM International Conference on

Multimedia (MM ’20), October 12–16, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3394171.3414538

[1] S. A. Chatzichristofis and Y. S. Boutalis. CEDD: Color and Edge Directivity Descriptor. A Compact Descriptor for Image Indexing and Retrieval. In A. Gasteratos, M. Vincze, and J. Tsotsos, editors, Proceedings of the 6th International Conference on Computer Vision Systems, ICVS 2008, volume 5008 of LNCS, pages 312–322, Santorini, Greece, May 2008. Springer.

[2] H. Tamura, S. Mori, and T. Yamawaki. Textural features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics, 8(6):460–472, June 1978.

[3] D. Lowe. Object recognition from local scale-invariant features. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pages 1150 –1157 vol.2, 1999.

[4] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing, 22(10):761 – 767, 2004. British Machine Vision Computing 2002.

[5] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. Computer Vision, IEEE International Conference on, 2:1470, 2003.

[6] C. Faloutsos and K.-I. Lin. Fastmap: A fast algorithm for indexing, data miningand visualization of traditional and multimedia datasets. In ACM SIGMOG, pages 163–174, CA, USA, 1995.

[7] G. Amato and P. Savino. Approximate similarity search in metric spaces using inverted files. In Proceedings of the 3rd international conference on Scalable information systems, InfoScale ’08, pages 28:1–28:10, ICST, Brussels, Belgium, Belgium, 2008. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).