FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

Tomáš Grošup

Methods of Multi-Modal Data Exploration

Department of Software Engineering

Supervisor of the doctoral thesis: prof. RNDr. Tomáš Skopal, Ph.D.
Study programme: Computer Science
Study branch: Software Systems

Prague 2019

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague date .... signature of the author

I would like to thank my supervisor Tomáš Skopal for guidance, advice, ideas and patience during the years of my studies. I would like to thank my consultant Jakub Lokoč for a lot of valuable ideas and frequent collaboration. And I would also like to thank my co-authors, especially Ladislav Peška and Přemysl Čech, for their fruitful cooperation on our joint projects.

This thesis would not exist without the support of my family – my wife Eva and our two kids Ema and Oliver. I thank them for creating a supportive environment and for understanding my needs.

I would also like to thank all the anonymous reviewers for their valuable feedback and hints for improvements.

Last but not least, I would like to thank to the funding sources which supported my research during my doctoral studies – Charles University, the Grant Agency of Charles University and the Grant Agency of the Czech Republic.

Title: Methods of Multi-Modal Data Exploration

Author: Tomáš Grošup

Department: Department of Software Engineering

Supervisor: prof. RNDr. Tomáš Skopal, Ph.D., Department of Software Engineering

Abstract:

Digitalization throughout the industry leads to rapidly increasing amounts of data captured and stored which brings forth challenges for indexing and accessing large digital repositories. Very often, the data takes form of complex multi-part entities, such as images with relational attributes, photos with geographical coordinates or textual posts with multimedia content and implicit social relationships. The complexity of such entities and lack of fixed structure makes it impossible to use classical information retrieval methods based on attribute filtering, ranking or grouping, as it is not easy or sometimes even possible to define an exact query. In this thesis, we target data exploration as an act of exploring an unfamiliar area via a series of intuitive, effective and efficient system-supported steps. We present methodologies, demo applications and evaluation results targeting different data sources of multimedia data. Furthermore, we focus on the ability to utilize multiple modalities within a single session and on integrating the results into widely used software solutions.

Keywords: multimedia exploration schema extension images

Contents

Preface

This thesis presents selected articles by Tomáš Grošup and their results. All articles are from the area of effective data exploration and target several aspects of the problem domain. The work covers research carried out during studies at the Department of Software Engineering, Faculty of Mathematics and Physics, Charles University in Prague. The author is a member of the SIRET (SImilarity RETrieval) Research Group which focuses on similarity search and its applications.

This thesis contains a filtered selection of articles accompanied by a unifying commentary at the beginning. This work covers 15 papers in total – 14 already reviewed and presented at international conferences and available in their published versions, one submitted to an academical journal available in a pre-print version. All of the papers are available via their respective publishers (Springer, ACM, IEEE, OpenProceedings) and have been authored or co-authored by Tomáš Grošup. The conference paper submissions were presented as demo submissions including software demonstrations, posters or regular talks. The conferences cover general database communities (EDBT, ADBIS, CIKM), multimedia communities (ICMR, MMM, CBMI) and the similarity search community (SISAP).

A full list of all publications is available at the end of this thesis. To provide an historical overview, 8 papers are available as Chapters 2–9 in the printed version of this thesis. Each of the Chapters 2–9 carries the name of the respective publication. Those have been selected to demonstrate different problem areas and evolution over time while keeping a small overlap between the works. The electronic version contains cover pages with references to digitally published versions of the articles.

The commentary uses two special symbols to highlight the author's publications:

- A light bulb to highlight a self-citation to a reference which is not embedded in this thesis. 💡

- A light bulb followed by a chapter link for a reference that is directly embedded in this thesis. 💡 👉 2

The list also includes work published during the author's bachelor's (2009–2012) and master's (2012–2014) studies. The table 1 shows author's contribution to each of the papers, both authored and co-authored.

Ref.	Conf.	Year	Evaluation	Implementation	Analysis	Text
[1]	ICMR	2012		▮▮▮▮	▮▮▮▯	▮▯▯▯
[2]	SISAP	2012		▮▮▮▮	▮▮▮▯	▮▯▯▯
[3]	SISAP	2013	▮▮▮▯	▮▮▮▮	▮▮▯▯	▮▮▯▯
[4]	CBMI	2014		▮▮▮▯	▮▮▯▯	▮▮▯▯
[5]	CBMI	2015	▮▮▮▯	▮▮▮▯		▮▮▮▯
[6]	MMM	2015		▮▮▮▮	▮▮▮▮	▮▮▮▮
[7]	CBMI	2015		▮▮▮▯	▮▮▯▯	▮▮▮▯
[8]	ADBIS	2015		▮▯▯▯	▮▮▯▯	▮▯▯▯
[9]	SISAP	2015	▮▮▯▯	▮▮▯▯		▮▮▯▯
[10]	SISAP	2017		▮▮▯▯	▮▮▯▯	▮▮▯▯
[11]	CIKM	2017	▮▮▯▯	▮▮▯▯	▮▮▯▯	▮▮▯▯
[12]	SISAP	2018	▮▮▮▯	▮▮▮▯	▮▮▮▯	▮▮▯▯
[13]	EDBT	2019			▮▮▮▯	▮▮▮▮
[14]	ICMR	2019			▮▮▮▮	▮▮▮▮
[15]	(submitted)	2019	▮▮▮▯	▮▮▮▯	▮▮▮▯	▮▮▮▮

▮▮▮▮	Full or main contribution
▮▮▮▯	Major contribution (around a half)
▮▮▯▯	Partial contribution
▮▯▯▯	Small contribution
-	Aspect not relevant

Table 1: A table of the author's contributions per article.

Chapter 1

Commentary

1.1 Introduction

The wide spread of affordable data-capturing devices has led to a rapid increase of stored data volumes in the last two decades [16]. The majority of the data volumes go to parts of data elements that are unstructured in nature – images, videos or sounds for example. Such data is very rarely unaccompanied – it is typically compounded with text descriptions, timestamps, spatial information or other metadata; forming a multimedia and multi-modal content [17][1]. Each of the multimedia components might influence information retrieval tasks depending on use cases which complicates systematic software support. Established tools for search, exploration and analysis of data (e.g., see [18, 19, 20]) are based on the relational model of structured entities with attributes , leveraging equality and natural ordering of primitive data types. It is not only supported by efficient indexing data structures [21], it is also well understood by end users and application programmers thanks to expressive query languages like SQL [22]. As the typical representation of rich multimedia content contains binary large objects , searching by equality rarely makes sense – one would need to have the entire object as a query upfront in order to use it as a query.

This limitation is addressed by so called content-based retrieval methods [23] in which the content of an entity is used for a search instead of defining queries on top of structured attributes. The dominant model for content-based multimedia retrieval is the similarity search. It utilizes a pair-wise similarity function and finds objects which are most similar to an input query – an example, a sketch or a reference to an already seen item from the dataset. This has the assumption that users' needs are clearly specified and can be expressed in the respective software. Multimedia exploration admits that this is not always possible, and supports users in navigating the data collections even when the search intent is not precise or not existent at all (”I don't know what I'm looking for, but I'll know when I find it” [24]). The support comes from novel user interfaces, exploratory data operations, and rapid query execution based on explicit and implicit user feedback. Internally, this leads to frequent execution of similarity queries with strict latency requirements, making efficiency an even more important criterion. Due to the iterative learning nature of the exploration process, search intentions are

[1] Multimedia is any combination of text, graphic art, sound, animation, and video that is delivered by computer [17]

also subject to change over time and so is the perception of similarity. Therefore, adopting the internal search model in an effective and efficient way according to user feedback is another challenge imposed on multimedia exploration systems.

The main challenge for any content-based retrieval method is the effectiveness of the solution, i.e., how well the results follow expectations of humans. In case of image retrieval, the traditional methods have been based on analytical approaches such as MPEG-7 [25], SIFT [26] or SURF [27]. The biggest problem is the semantic gap – a minor difference in low level (pixel-wise) representation of an image causing a major difference in human understanding and vice versa. It was the deep learning revolution [28] which advanced state-of-the-art results in many retrieval domains including computer vision, beating classical solutions in many tasks and disciplines.

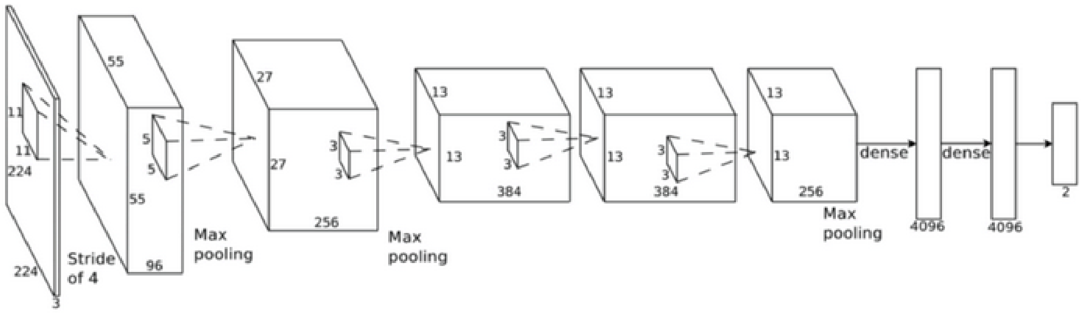


Figure 1.1: Diagram of the architecture of AlexNet, a deep convolutional neural network [29]

Figure 1.1 shows the network architecture of AlexNet [29], a pioneer of modern architectures. The main part of the network consists of five convolutional layers and three fully connected layers. The operation of convolution gradually allows capturing concepts of higher levels of abstraction, and it can be shown what information each neuron detects [30, 31]. With enough training data, this closes the semantic gap between low-level pixel information and human language class identifiers. This field of computer vision is rapidly evolving, and many architectural paradigms have been recently introduced [32, 33]. Although the typical problem is image classification, i.e., assigning a known class to each input image, the inner representation of the network can be used to build valuable feature descriptors. The descriptors can be extracted using neuron activations after a forward-pass of the input [34], and can produce different feature descriptors (high-dimensional vectors) based on the selection of layers.

The evolution of deep learning approaches in computer vision and the ability to extract feature descriptors has also improved content-based retrieval methods for other, secondary tasks. And despite being trained for a different task and even on a different image domain, such descriptors outperform analytical approaches like SIFT [35]. The advances together allow for querying via semantic visual attributes [36, 37] based on neural activations coming from different layers. What if we could close the loop by using such attributes for the extension of relational database schema? That way, we could implement a multi-modal search in a familiar relational environment, and leverage the power of existing software solutions. After identification and extraction of visual attributes, there would be one step left for humans – putting a label on it.

1.1.1 Chapter Organization

In the rest of this chapter, we briefly go through related areas which were involved in our research. The different areas are positioned chronologically with respect to research conducted by the author, and loosely follow the outline given by the introduction. Contributions which were published are marked with the symbol of a light bulband include a reference to the original publication. Selected works are 💡 also directly embedded in the printed version of this thesis, and are in addition referenced using the number of the respective chapter. The chapter 2 is a short summary about all topics touched on by the author, and was presented as a talk at the ICMR 2019 conference [14]. The current chapter puts more focus on the 💡 ☞ 2 background, motivation and related work of different problem domains, as the inner details of individual contributions are described within chapters 3 - 9.

1.2 Similarity Search

The similarity search is the most prominent approach to content-based searching. In general, it is applicable in scenarios where standard exact-match or range queries cannot be used, due to the complexity of the entities and human expectations of the search result. For complex entities, offering an exact search rarely brings the expected outcome, as one would need to have the whole entity as a query parameter upfront in order to find it. Areas with wide usage include images, videos, sounds, DNA sequences, protein structures or 3D shapes, for example. The principle of similarity search is a pair-wise ranking function which assigns a score to each pair of objects. This can be either a similarity score (higher means more similar) or a dissimilarity / distance score (lower value means more similar). The requirement then is to search a database and return objects most similar to an input query.

In distance-based solutions, there are two common queries for search and exploration activities – the range query and the k-nearest neighbours query. A range query returns all objects within a distance r_q from the query object q, and the possible number of returned objects is dynamics. A k-nearest neighbours query (kNN) returns exactly k objects with the lowest distance to query object q. Figure 1.2 provides an illustration of the difference.

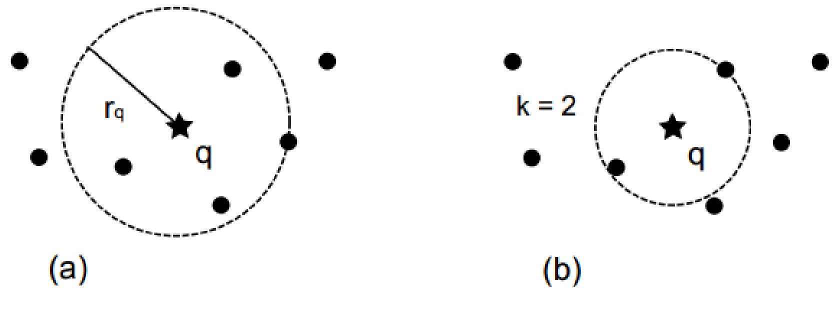


Figure 1.2: a) Range query b)k-nearest neighbors query with k=2

Besides these two basic queries, more complex operations exist as well. One of them is the similarity join. Similar to a relational database join, it takes two datasets as input, and returns all pairs of objects whose distance is lower (or greater) than a parameter r. This makes it useful for tasks like duplicate detection, outlier detection, or identification of common patterns in the data.

We have employed similarity search methods across all the research we have done over the years. It is the most common approach for content-based retrieval and exploration, and has been researched in depth for decades. The quality of developed search solutions has two main evaluation dimensions:

1. Effectiveness, a quality of the results when compared with human expectations. This is typically measured using metrics comparing the results against a ground-truth, such as precision or recall.

2. Efficiency, a measure of speed and resource utilization to deliver a response. This is measured in wall-clock time, but also in terms of memory usage or operations on persistent media.

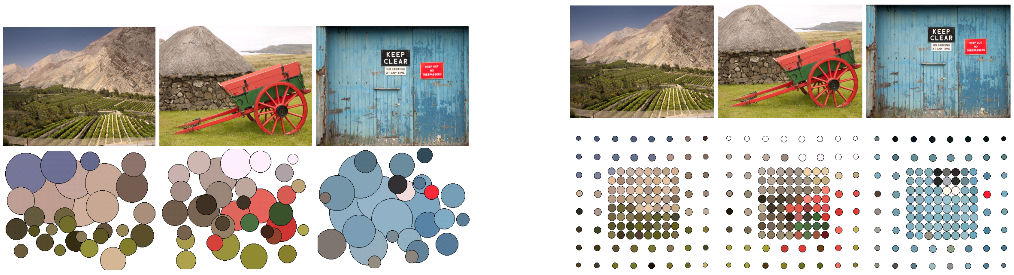
Effectiveness is mainly influenced by the design of the pair-wise similarity function. This is a field that is constantly evolving, and in the exemplary domain of images has been around for decades with active research contributions [38]. In practical implementations, the similarity function does not operate on original images (raw pixels) directly, but rather on extracted feature descriptors which are more compact and more robust to common changes. The robustness usually comes at a cost of increased processing times. Robust features, like SURF [27] or PCT (Position-Color-Texture) signatures [39], are typically meant for scenarios with an offline prepossessing phase where slower extraction times are not a big issue.

Within our early work [2], we have introduced the concept of real-time feature extraction and re-ranking. In order to support the low-latency requirement of the online feature extraction, we have designed a slim feature extractor as an alternative to PCT signatures called *PixelGrid*. Both are designed to be used together with the Signature Quadratic Form Distance (SQFD) [39]. The slim version eliminates expensive operations, especially k-means clustering, and reduces computational complexity to the down-scaling of an image thumbnail in two different sizes – one for the center of the image with higher density, and the other for the edges. Figure 1.3 shows a comparison between the outcomes for the same three images, Figure 1.3a for the original k-means based algorithm, and Figure 1.3b for our fast extractor.

The efficiency of similarity search approaches can be improved not just by using faster extractors or similarity functions, but also by using database indexing techniques. For vector spaces, spatial database indexes can be used, e.g., the R-Tree. However, the search for more effective descriptors and distance functions leads to the requirement for a more generic solution, not restricted to vector spaces and Euclidean distance. One such approach is based on the properties of metric spaces, a mathematical construct enforcing certain properties on the chosen distance function.

1.2.1 Metric Space Approach

Metric space is a set together with a metric on that set. The metric is a function which defines a real-valued distance for any pair of elements from that set. On



(a) PCT extraction based on k-means clustering of pixels.

(b) Slim extractor based on fast down-scaling of an image thumbnail.

Figure 1.3: Visual demonstration of the difference in outcomes for PCT signatures and our slim extractor used for fast re-ranking.

top of that, it enforces certain properties on the metric. This in turn allows for the design of generic solutions for any similarity search problem that can be represented using a metric space, and is not restricted to specialized domains.

Definition 1. *Let $\mathbb{F}$ be a feature space, δ a distance function measured on $\mathbb{F}$. $\mathcal{M} = (\mathbb{F}, \delta)$ is called a metric space if distance function δ: $\mathbb{F} \times \mathbb{F} \mapsto \mathbb{R}$ fulfills the following postulates:*

1. $\forall x, y \in \mathbb{F},\ \delta(x, y) = \delta(y, x)$ *symmetry*
2. $\forall x \in \mathbb{F},\ \delta(x, x) = 0$ *reflexivity*
3. $\forall x, y \in \mathbb{F},\ x \neq y \Rightarrow \delta(x, y) > 0$ *positiveness*
4. $\forall x, y, z \in \mathbb{F},\ \delta(x, z) \leq \delta(x, y) + \delta(y, z)$ *triangle inequality*

It is the triangle inequality property which allows us to create efficient data structures that do not need to calculate distances to all objects in the database in order to answer queries. The inequality properties between all triples of objects makes it possible to estimate lower and upper bounds of a distance between two objects, provided that the other two sides of the "triangle" are already known. Figure 1.4 provides a geometrical illustration of the property.

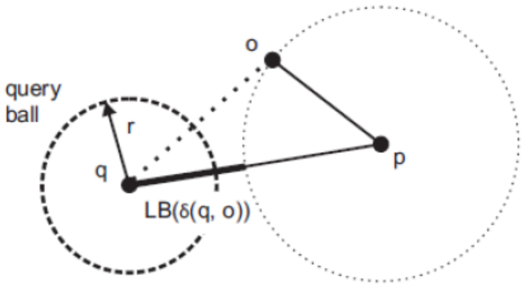


Figure 1.4: Illustration of triangle inequality as a means to estimate lower bound LB for the distance $d(q, o)$ provided that distances $d(q, p)$ and $d(o, p)$ are known.

The family of solutions leveraging the metric space properties are called Metric Access Methods, or MAMs. Zezula et al. [40] provide an extensive summary about the problematic and different solutions. The triangle inequality is used for pruning the search space and avoiding possibly expensive distance calculations in various forms – recursive tree structures, hashing methods, tables with pre-calculated distances to global pivots, or hybrid combinations of multiple approaches.

The obvious downside is that not every distance function is a metric. The search in non-metric spaces has been typically addressed by brute force sequential scanning, e.g., by aggressive optimization of calculations [41] or by executing search queries on GPUs [42]. The *TriGen* algorithm provides a unifying approach for turning non-metric distance functions into metrics via modifying functions that change distance values, but preserve ordering [43, 44]. That way, MAMs can be used even for distance functions which are not originally designed with metric properties in mind.

1.2.2 Synergistic Modelling

The efficiency of MAMs depends on the power of lower bound values provided by triangle inequality. It can be observed that the same distance function can have dramatic differences in the utilization of a MAM depending on the data distribution. In a metric space that is almost equidistant (all triangles are almost equilateral), the lower bound condition cannot prune any objects during the search. In vector spaces, the curse of dimensionality [45] is a known phenomenon decreasing speed and quality of information retrieval and data analytics in high dimensional spaces. In metric spaces, a similar phenomenon is observed and is referred to as intrinsic dimensionality, an experimentally confirmed approximation of indexability when using MAMs [46]. Also referred to as *iDim*, it is defined as $\rho = \frac{\mu^2}{2\sigma^2}$, where μ and σ^2 are the mean and the variance of the distance distribution for the entire collection. The lower this value is, the better the indexability options for the collection. As the exact computation requires evaluation of all possible pair-wise distance values, a selection of techniques exists to estimate *iDim* [47].

In a traditional setup, it is the domain (domains such as computer vision, biology, acoustics) expert defining appropriate descriptors and distance functions with the goal of maximal effectiveness. However, if the distance function is already given, it does not leave much space for speed improvement if iDim turns out to be high. The idea of synergistic modelling is to involve database engineers early in the process in order to identify and measure possible compromises – having a synergy between effectiveness and efficiency. As we have shown in [3], there are options to sacrifice a little bit of precision for rapid increases in efficiency. This especially makes sense in domains where even an exact search cannot provide perfect responses (such as a search in general imagery). On the other hand, it should not be used as a technique in situations that require maximum precision (bio-metrics).

Synergistic modelling provides a solution for balancing precision/speed trade-offs in similarity search, orthogonal to pure database solutions that allow the discovery of the balance between exact and approximate search algorithmically based on the desired approximation factor [44].

1.3 Multimedia Exploration

The basic queries of similarity search, the range query and k-nearest-neighbours query, fall into the category $query-by-example$. The user provides an example object, and receives the most similar objects. This can be difficult in practical applications – if the user already has an example, why would she search for another? More often, the user has just an idea, an image of the result in her mind. One way to address this limitation are $query-by-sketch$ solutions where the user provides a sketch of the desired outcome. Multimedia exploration takes a different approach, admitting that providing a query is not always possible [48] and instead provides an interactive experience. The user takes an active part in the search and her discriminatory power is used to navigate through potentially large collections of objects.

Classical data exploration for relational data is based on multi-dimensional hyper-cubes, drill down operations and explicit filters. We have learned that complex entities cannot use these techniques due to the lack of a standardized structure, and impose their own challenges on exploratory systems. The typical set of requirements for smooth human-computer interaction includes at least the following parts:

1. **Page zero**. A starting point of the exploration process. At this point, the user's intention is not clear and no query has been provided yet. It should provide a visual summary of the data collection and allow further navigation. At the same time, navigation options should maximize reachability, i.e., ensure that every item in the data collection can be eventually reached from the starting page zero.

2. **Results visualization**. In order to utilize the discriminatory power of the human brain, results should be arranged in a way that allows for the quick identification of new areas of interest, without investigating each displayed object in detail. A common approach are similarity-based layouts [49, 50], which put similar objects close to each other on the display (e.g., force-directed layouts, PCA,MDS,t-SNE or IsoMap).

3. **Exploratory operations**. To allow the navigation through the collections, each system provides its own set of operations. These might be based on scrolling, touch inputs, explicit selection and many other methods. Other options might include the look and feel of a geographical map, with the possibility to zoom or move to the sides, giving the impression of data collections being mapped to 2D space [51].

Besides the functional requirements, there are also implicit non-functional requirements raising the challenge for large-scale exploration – efficiency, scalability, multi-user environments or automatic adaptivity to implicit feedback [52].

1.3.1 Reranking

In the first work by the author [1], we have introduced a small scale exploration system built around the idea of re-ranking. Instead of immediately exploring a possibly large collection, a service provider (Bing images) was used to execute a 💡☞ 3 text-based query. The system then offered a similarity-based visualization of parts of the results using force-directed layout [53], and offered zoom-in implemented as kNN queries. In this form, reranking was used as a mechanism to offer cross-modal search capabilities – the user starts the process with text modality, and receives image results arranged based on their visual appearance.

The biggest challenge in this case is the speed of online feature extraction and instant execution of similarity queries which were based on our slim feature extractor and MAM indexing support.

The speed of the process was improved in our further work [2] via the immediate rendering of results after enough objects were downloaded to build a page zero. While the user was inspecting this initial view, the system was downloading the rest of the results, extracting their descriptors, and indexing them. This improved the user-perceived time to first render, while allowing operation on a larger collection for further navigation. 💡

1.3.2 Continuous Exploration

Exploration, as an interactive process, imposes special needs for fast response times and smooth execution. While efficiency aspects of the execution are addressed by index support and server-side optimizations, smoothness of the process from a user perspective also depends on the selected client-side display methods. In our early work, we employed a force-directed layout which was always re-initiated after each operation. That led to confusing movement on the screen, and did not offer continuity for objects that remained displayed – such as in the case of the zoom-in/out operation, where there was an overlap between objects visible before and after the operation.

To provide continuity, we changed the system [7] so that preserved objects maintained their position and new objects were attached based on their proximity to the old ones. Force-directed layout then smooths out possible overlaps or dense areas, but at a slower rate when compared to an initial layout creation. Figure 1.5 provides a screenshot of the exploration system with the available operations and the similarity-based layout. In this example, the underlying dataset is a collection of food items, with similarity defined based on their nutritional attributes. 💡 👉 4

1.3.3 Efficiency of Multimedia Exploration

Metric access methods are an efficient tool to execute basic similarity queries, and can be used as such to enable multimedia exploration. However, MAMs also have a valuable internal structure that already groups data into sections based on distance values between objects. Instead of executing standard range or kNN queries, this internal structure can be natively traversed to support different exploratory operations.

In our work on exploration using the metric space approach [4], we have established the terms 'iterative querying' and 'iterative browsing' as two options for explorations using MAMs. Iterative querying as a generic solution to issue standard queries provided by MAMs to explore a collection, and iterative browsing as a native way to explore a collection using a specialized index structure. Figure 1.6 illustrates the basic idea behind the two approaches. 💡

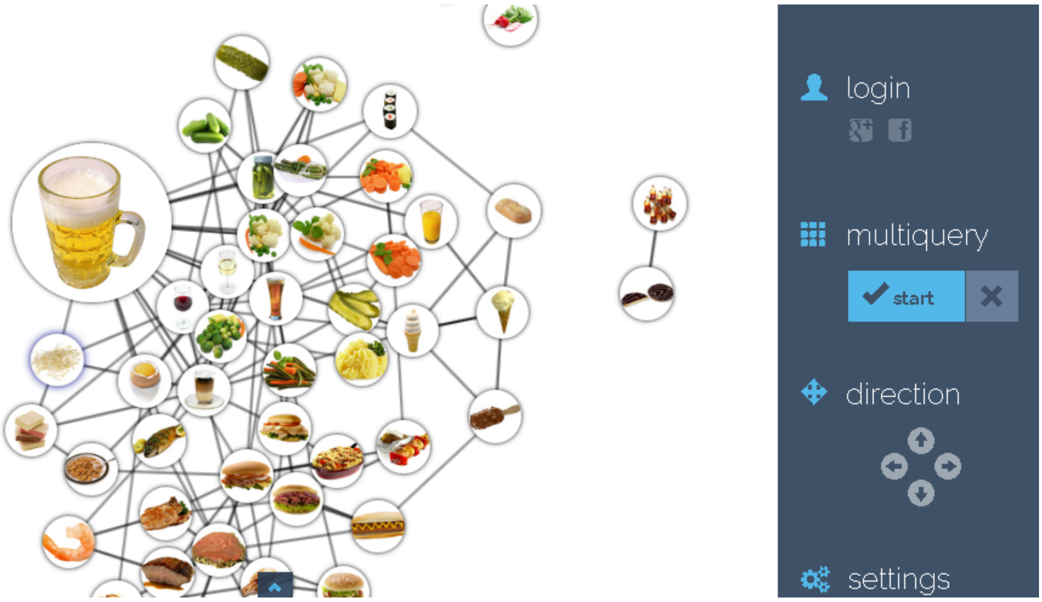


Figure 1.5: Screenshot of our exploration system operating on a food dataset. Similarity is defined by multi-dimensional information about nutrients, such as calories, proteins or fats included.

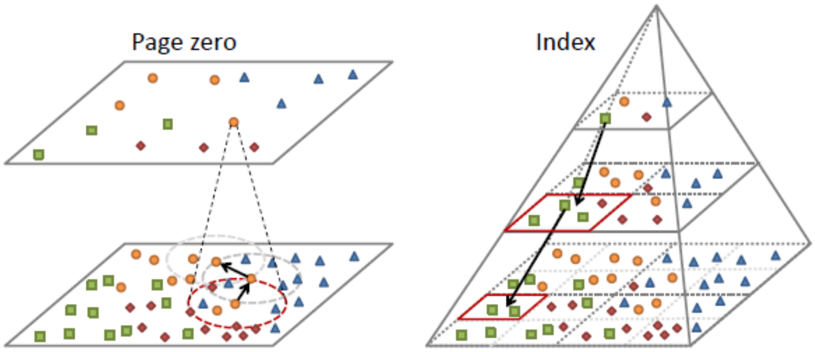
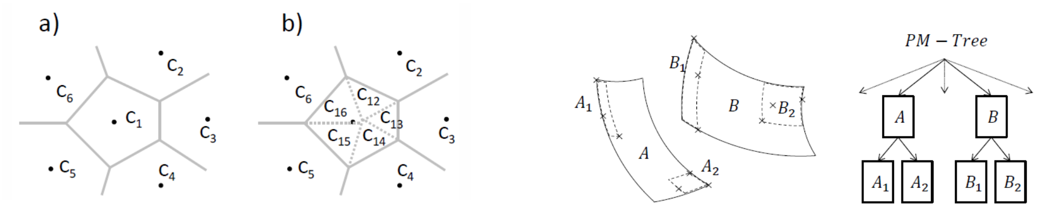


Figure 1.6: Illustration of iterative querying as a sequence of basic queries (left) and iterative browsing as a traversal of an index structure.

We have implemented native traversal into M-Index [54] and PM-tree [55] as two representatives of different MAMs. M-Index is built around Voronoi clustering of a metric space based on ordering (permutations) of distances to a global set of pivots. PM-Tree as a successor of M-Tree uses a recursive hierarchy of ball regions to maintain data, and each ball is further trimmed into cut-regions [56] using distances to a global set of pivots. Figure 1.7 illustrates the main construction ideas behind M-Index (Figure 1.7a) and PM-Tree (1.7b).

The downside of native MAM browsing is that their inner structure is optimized for database performance and not for exploration effectiveness. We have introduced *MLES* [8] as a generic exploration meta-structure, which can plug-in different MAMs internally. It defines and implements exploratory operations on a set of N-maintained inner indexes, each corresponding to a certain level and covering increasingly larger sections of the database. The first layer, L_0, covers just enough objects to contain the initial view, the page zero. Each layer then contains all objects from the previous layer and additional objects from the



(a) Recursive Voronoi-clustering of a metric space based on distances to global pivots, illustrated for the first and second level.

(b) PM-Tree defined as a tree structure of cut-regions based on ball regions and distances to global pivots.

Figure 1.7: Illustration of the basic building blocks used by M-Index (left) and PM-Tree (right).

database. The last layer, L_n contains all of the objects. As we can see in Figure 1.8, the operations defined on $MLES$ can navigate within a single layer (panning), and go a layer up (zoom-in) or down (zoom-out). Beyond performance, we have also researched the theoretical properties of exploration reachability and possible ”exploration dead-ends” when using iterative querying.

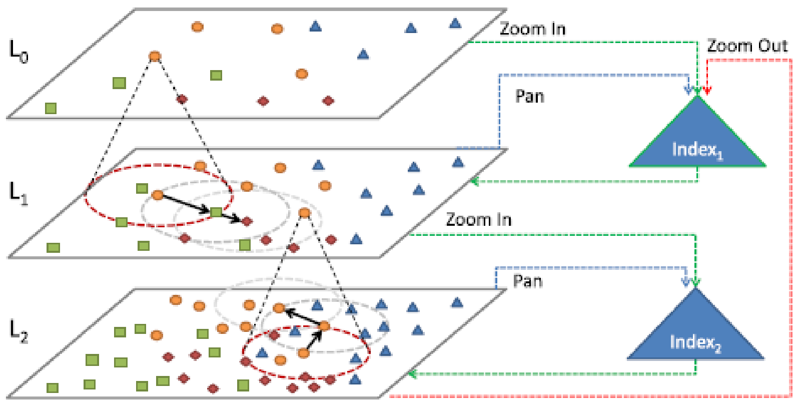


Figure 1.8: Illustration of an exploration session, in this case on the $MLES$ data structure – an initial view, zoom-in/out and panning.

1.3.4 Evaluation

We have proposed different methods to multimedia exploration in the previous paragraphs. Since exploration is by definition a user-centric and interactive process, it is very difficult to estimate the quality of a solution algorithmically. Although certain properties like reachability or efficiency can be measured automatically, the goal of exploration is to satisfy a search intention and that is difficult to automate into a standard ground-truth.

To allow a fair comparison between our methods, we designed a user study. We prepared a set of artificial tasks on a medium-size collection of images consisting of 21 993 annotated images associated in 100 different classes [5]. The user study is called *Find the Image* and the task is to find images of a certain class in a limited number of steps. The user is presented with one example of a class, e.g. a pyramid, and is tasked to search for images of the same class. The user study then measures the time it takes to find the first instance, the number of objects found over time, wall-clock time and the number of executed distance computations among others.

For the purpose of $MLES$, the user study was repeated in a similar setting to measure the differences between $MLES$ and standard iterative querying. With 94 participants, the outcome is that $MLES$ can consistently decrease the number of steps needed to find first representative of a desired class [9].

1.3.5 Domain Applicability

Although our techniques are built around general concepts (similarity search, metric access methods, similarity-based layouts or exploratory operations), all our demonstrations so far have been targeting images. We have applied [10] the ideas of multimedia exploration to the domain of cybersecurity and malware discovery. The form was a tool to enable expert support for exploration of descriptor spaces and validate their functioning. The tool provided different views, allowing the inspection of:

- Overview and clustering of machines.
- Similarity of client machines.
- Client-server similarities based on patterns of network behaviour.
- Drill-down to multiple inflected clients and their bins in a descriptor (codebook of fingerprints clustered into a fixed number of codewords/bins). An example can be seen in Figure 1.9.

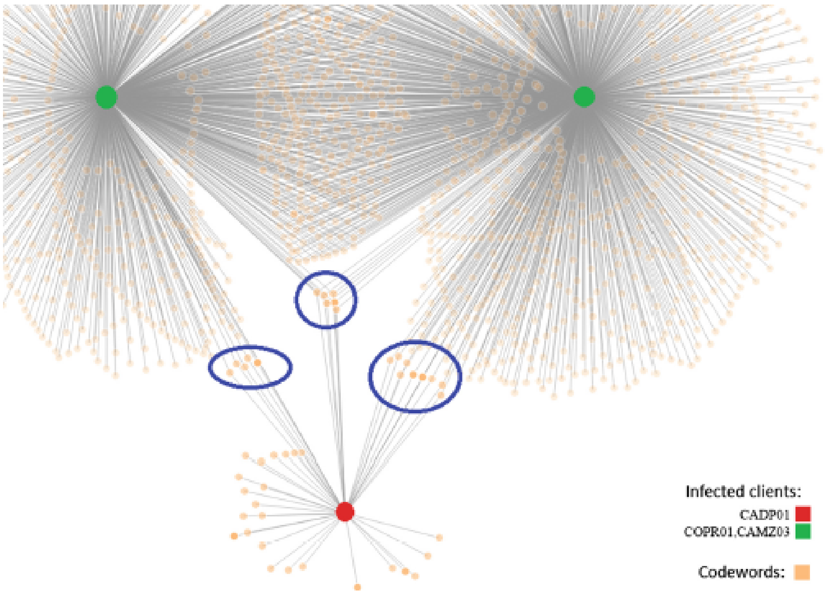


Figure 1.9: A part of the screen for the visual inspection of codewords/bins of a descriptor created to detect malware. In this picture, infected client computers are sharing multiple codewords in the representation of the descriptor.

1.4 Multi-Modal Search

As per the Oxford dictionary[2], modality is defined as a particular mode in which something exists or is experienced or expressed. Within software, example modalities (both input and output experience) include, for example, vision (images), text, sounds, touch input, mouse input, and non-text computer visualizations.

When it comes to information retrieval, we can achieve multi-modality in several ways:

- Presenting an object via multiple modalities
- Offering user-interaction via multiple modalities
- Enabling search criteria via multiple modalities
- A special form of multi-modality is the cross-modal search, where one modality is the input (e.g., a text-based query) and other is the output (e.g., videos).

When it comes to human computer interaction (HCI), [57] provides an extensive survey of results in the HCI community. In this work, we focus mainly on the multi-modal search, for example, allowing multiple modalities for information retrieval.

The most common approach to multi-modal searches is textual annotation of non-textual data. The best representative is Google Images, which automatically annotates images using the surrounding HTML page information, and subsequently allows the search for images using text queries. This fact was also leveraged in our approach to online reranking [1], where a text query is used as 💡☞ 3 a starting point continued with multimedia exploration based on content-based similarity.

Although the technique of automatic annotation works very well for the retrieval of images coming from web pages, it is not a generally applicable solution. The amount of multimedia data collected every day exceeds manual annotation possibilities, and is subject to human interpretation of the content and a personal view on the expected granularity of the provided annotations. This can lead to missing or misguiding annotations, opening the door to false positives and false negatives during a search. There are attempts at automatic annotation leveraging linguistic methods and public ontologies to improve the quality of annotations, as well as approaches for content-based keyword assignment. However, these solutions are far from a perfect and general applicability, and a content-based search often remains the only option to investigate a specific modality.

Our work on visual reranking used content-based searches to further explore objects with equal textual annotation. Imagine searching for a particular image of a Jaguar (the animal). By typing in 'Jaguar', the results will include the animal, as well as the brand of car called Jaguar. Visually, they are very different, but the text annotation is the same. One could extend the query to 'Jaguar animal', however this would sacrifice recall as not all images of animals are explicitly labelled as 'animal.' A content-based reranking is a means to visually inspect the result-set using image modality and to complete the search.

[2] `https://www.lexico.com/en/definition/modality`

For complex modalities, there are multiple ways how to design and implement content-based retrieval, and they serve a different purpose. At the same time, it is often not clear what aspect is most relevant to the user if her intention is not clear. As an example: For image analysis, there are different descriptors and distance functions for color, edges, texture, global information, local image information and different levels of the semantic scale. Each of the aspects might be relevant in certain search contexts. In our work on multi-model[3] exploration [6], we have 💡☞ 6 proposed two mechanisms to combine multiple similarity and retrieval models in a single exploration session. The contribution are two learning functions which interpret implicit feedback coming from executed operations and use it to guess the balance between implemented similarity models. This has been evaluated to automatically adapt a search between local and global image descriptors, depending on what the user's search intention is. To support indexability of dynamically adapting distances, a modified version of MAMs [58] exists that can supported unbounded combinations of balance weights with minimal space overhead.

1.5 Visual Attributes

Visual attributes are a special form of content-based annotation driven by the simplicity and wide usage of relational models. By converting a particular visual aspect from the content of the data into a clearly named and isolated attribute, existing tools for searching, analytics or recommendation could utilize it in the familiar model of relations and attributes.

The state-of-the-art tools for image classification have already outperformed humans in a supervised scenario with fixed and known classes. This means that given a catalogue of desired attributes and a sufficient number of annotated examples, the tools (typically based on deep learning) can learn a function that assigns one of the known classes to each input with certain confidence.

In real world scenarios, both conditions might be difficult to achieve – many domains are too complicated to allow a fixed set of classes upfront, and also do not posses a huge standardized collection of labelled examples.

An interesting approach for discovering visual attribute is the usage of web data - text labels, surrounding HTML pages and hyperlinks between items [59, 60]. When connected with ontologies and natural language processing, this can also connect semantic concepts described by different words, and even show an evolution of a certain linguistic or visual term over time [61].

A specialized area of research is the ability to discover visual concepts in parts of images while only having an annotation for the image as a whole. By discovering additional visual concepts at the local level [62], the quality of the dataset can be improved and further provided to supervised methods that need a lot of training data upfront.

What still remains a challenge, especially in long-tail domains, is the following combination of properties:

- There is no standardized classifier specialized for the domain.

[3] In our work, model refers to a similarity model. The term model is overloaded, and is not to be confused with multi-model databases that are named after multiple storage models (e.g., relational, document and graph-based storage and query models).

- There is no large labelled dataset and it is either impossible or impractical to create one.
- The domain vocabulary is rich, dynamic and subjective, therefore not possible to be defined at design time.

We have selected the domain of fashion items in an e-commerce setup as our example model, as it meets all of the three properties. At the same time, it is visually very rich and the visual information is an essential part of typical retrieval needs.

1.5.1 Latent Visual Attributes

Historically, computer vision models have suffered from the semantic gap phenomena – a mismatch between low-level computer representation and human understanding of image appearance. Certain visual aspects, such as 'floral pattern,' are difficult to describe using low-level constructs, but are still present and clearly understandable by humans. We denote them as latent, and we leverage the power of deep learning architectures [28, 33] to use them.

💡 ☞ 7 We have presented [11, 63] a prototypical e-shop application which on top of classical navigational options (categories, tags) also offers the option to search using visual appearance of a product. It allows an exploration of the product space, and by offering multi-example-query capabilities, can be used to describe visual patterns. Such visual patterns can be named and matched against the entire dataset, which turns them into binary attributes. The demonstration shows how such created latent attributes can be used to recommend individual products and outfits.

The main distinction from other works in the fashion domain working on similar tasks (outfit recommendation [64], product retrieval [65], semantic clothing attributes [66]) is the absence of an initial design-time preparation and no dependency on specialized training data. The visual search in our prototype was using a publicly available model of AlexNet pre-trained on generic imagery, without any special training phase on fashion data – neither supervised nor unsupervised. As part of our demo, we have also experimentally evaluated the quality of the search on prepared search categories while comparing different layers of AlexNet as source for feature descriptors.

1.5.2 Local and Global Attributes

💡 In our next prototype [12], the idea was further improved by considering individual parts of images and allowing users to select one of the image patches as an area of interest. For an empirical evaluation, we have prepared categories and different configurations of search to find them. Mainly, our interest was to find a qualitative difference between a global search, a local search and a weighted combination of the two. It is the weighted combination which turned out to be the best choice, and a very significant improvement over the original global search. In Figure 1.10 we can see that the weighted combination (green line) was the best performing option for a majority of search tasks, sometimes outperforming the original global search by more than twice (in terms of distinctive cumulative

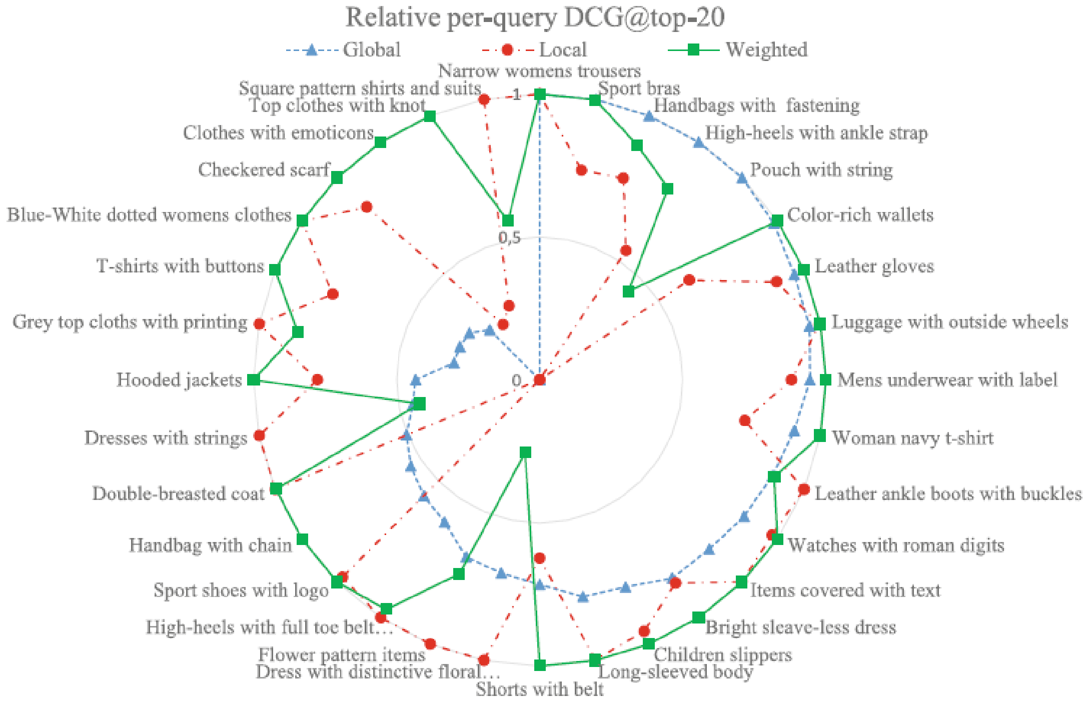


Figure 1.10: A relative comparison of distinctive cumulative gain values for the first returned 20 objects (DCG@20) between global, local and weighted search executions. The circle circumference always contains the best performing option for the given search task; the remaining two are positioned relative to it.

gain for first 20 retrieved items). The chart is sorted clockwise by the descending relative performance of the global search, showing also what search tasks were suitable for a global search (e.g., 'Leather gloves') and where a local or weighted search was significantly better (e.g., 'T-shirts with buttons').

1.5.3 Attribute Discovery

After the successful experimental evaluation of a manual search for visual attributes using system-supported multi queries and image patches, we developed a vision of generating the attributes in the background. That way, the work of identifying promising patterns across the data would be delegated to software, and humans would only execute the last steps – verifying the attribute, and putting a name on it [13]. 💡👉 8

We looked in to the problem, described the outstanding challenges, and proposed a workflow in the form of a data-pipeline (see Figure 1.11). The vision was built around the following conditions and outcomes:

- The domain is visually rich and concepts cannot be separated at design time into a fixed number of concepts.
- There is no pre-trained model or labelled dataset, but generic solutions exist for the same type of data. (specifically for images, this means that a generic classifier exists, but there is no specialized classifier for the same problematic)

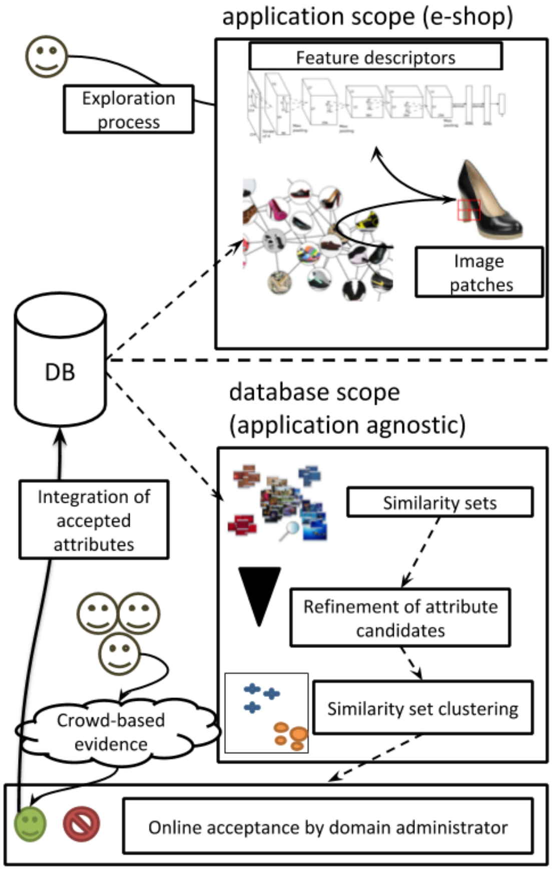

application scope (e-shop)
Feature descriptors
Exploration process
Image patches
DB
database scope (application agnostic)
Similarity sets
Integration of accepted attributes
Refinement of attribute candidates
Similarity set clustering
Crowd-based evidence
Online acceptance by domain administrator

Figure 1.11: Proposed workflow for the discovery of latent visual attributes and for database schema augmentation.

- In order to use the results in widely adopted software, the work happens on the database layer and is not an application-specific task. By integrating the results into the established model of database relations, multiple application-specific use-cases like recommendation, search or analytics can benefit from it at the same time. This is in contrast to solving one application-specific task, e.g. outfit recommendation, which has decreased applicability options beyond the scope of the original task.

In our latest research activities, we continued in the direction outlined by that vision. We have formalized the problematic and the proposed workflow, implemented a software architecture for offline attribute discovery and online attribute acceptance and integration, and provided the results in an SQL-integrated solution. The model was evaluated on the combined domains of shoes, clothes, accessories and household items. The online acceptance phase was tested by eight users in the role of domain administrators, and resulted in 218 new attributes covering 6,901 database items.

The formal methodology, software architecture and evaluation results were submitted to the journal *InformationSystems*, and pre-print version of it is available as Chapter 9 of this thesis. 💡☞ 9

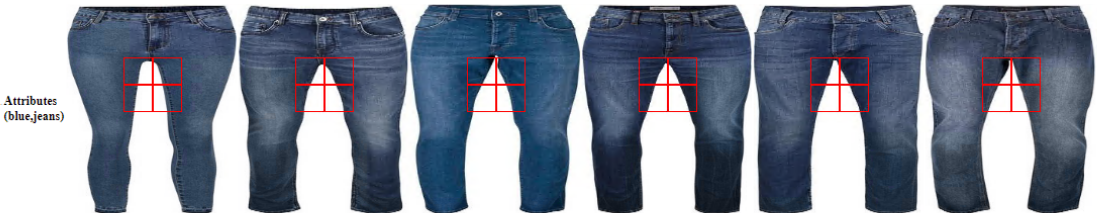


Figure 1.12: Example of a discovered visual attribute and the image patches defining it. In this case, the system also identified correlation of this visual attribute to an intersection of two existing independent attributes 'blue' and 'jeans.'

Chapter 2

Methods of Multi-Modal Data Exploration

- Tomáš Grošup

Methods of Multi-Modal Data Exploration [14]

Paper at International Conference on Multimedia Retrieval (ICMR 2019), Ottawa, Canada, 2019.

Part of Proceedings of the 2019 on International Conference on Multimedia Retrieval, pages 34-37, ISBN 978-1-4503-6765-3, published by ACM New York, NY, USA.

Chapter 3

Image exploration using online feature extraction and reranking

- Jakub Lokoč
- Tomáš Grošup
- Tomáš Skopal

Image exploration using online feature extraction and reranking [1]

Paper at International Conference on Multimedia Retrieval (ICMR 2012), Hong Kong, China, 2012.

Part of Proceedings of the 2nd ACM International Conference on Multimedia Retrieval, ISBN 978-1-4503-1329-2, published by ACM New York, NY, USA.

Chapter 4

Continuous hierarchical exploration of multimedia collections

- Tomáš Grošup
- Juraj Moško
- Přemysl Čech

Continuous hierarchical exploration of multimedia collections [7]

Paper at 13th International Workshop on Content-Based Multimedia Indexing (CBMI 2015), Prague, Czech Republic, 2015.

Part of 13th International Workshop on Content-Based Multimedia Indexing, ISBN 978-1-4673-6870-4 , published by IEEE.

Chapter 5

MLES: Multilayer Exploration Structure for Multimedia Exploration

- Juraj Moško
- Jakub Lokoč
- Tomáš Grošup
- Přemysl Čech
- Tomáš Skopal
- Jan Lánský

MLES: multilayer exploration structure for multimedia exploration [8]

Paper at 19th East-European Conference on Advances in Databases and Information Systems (ADBIS 2015), Poitiers, France, 2015.

Part of the Communications in Computer and Information Science book series (CCIS, volume 539), ISBN 978-3-319-23200-3, published by Springer, Cham.

Chapter 6

A web portal for effective multi-model exploration

- Tomáš Grošup
- Přemysl Čech
- Jakub Lokoč
- Tomáš Skopal

A web portal for effective multi-model exploration [6]

Paper at 21st International Conference on MultiMedia Modeling (MMM 2015), Sydney, Australia, 2015.

Part of the Lecture Notes in Computer Science book series (LNCS, volume 8936), ISBN 978-3-319-14441-2, published by Springer, Cham.

Chapter 7

Product exploration based on latent visual attributes

- Tomáš Skopal
- Ladislav Peška
- Gregor Kovalčík
- Tomáš Grošup
- Jakub Lokoč

Product exploration based on latent visual attributes [11]

Paper at the 26th ACM International Conference on Information and Knowledge Management (CIKM 2017), Singapore, Singapore, 2017.

Part of Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, ISBN 978-1-4503-4918-5, published by ACM New York, NY, USA.

Chapter 8

Towards Augmented Database Schemes by Discovery of Latent Visual Attributes

- Tomáš Grošup
- Ladislav Peška
- Tomáš Skopal

Towards Augmented Database Schemes by Discovery of Latent Visual Attributes [13]

Paper at the 22nd International Conference on Extending Database Technology (EDBT 2019), Lisbon, Portugal, 2019.

Part of Proceedings of EDBT 2019, ISBN 978-3-89318-081-3, published by OpenProceedings.org.

Chapter 9

Augmenting Database Schemes by Latent Visual Attributes

- Tomáš Grošup
- Ladislav Peška
- Tomáš Skopal

Augmenting Database Schemes by Latent Visual Attributes [15]

Submitted to Information Systems, Elsevier – in review as of June 2019.

Conclusion

In this work, we have provided an overview of different options to multi-modal exploration. For demonstration purposes, we typically used complex entities involving images, texts and relational attributes – however, the presented models are general and can be adapted to other forms of data as well.

We have developed several distance-based methodologies for multimedia exploration, considering both the effectiveness and efficiency of delivered solutions. We have conducted large user studies to measure both processual (user time, number of steps) and computational (wall-clock time for computer operations, index utilization) aspects of developed solutions.

Lastly, we utilized recent advances in Deep Learning as a loosely-coupled (replaceable) component to search for latent visual attributes. We have proposed and demonstrated a pipeline for discovering such attributes and turning them into immediate augmentation of relational schemes. That way, a multi-modal search is achieved via standard SQL queries utilizing these visual attributes together with other forms of data such as full-text and non-visual attributes (e.g., price).

Future Work

In the future, we would like to identify new applications for the proposed methods. Firstly, identifying and experimentally evaluating other domains with big potential gains. Secondly, demonstrating the methods on non-visual sources of multimedia such as sound or natural language.

Bibliography

[1] Jakub Lokoč, Tomáš Grošup, and Tomáš Skopal. Image exploration using online feature extraction and reranking. In *Proceedings of the 2Nd ACM International Conference on Multimedia Retrieval*, ICMR '12, pages 66:1–66:2, New York, NY, USA, 2012. ACM.

[2] Jakub Lokoč, Tomáš Grošup, and Tomáš Skopal. Sir: The smart image retrieval engine. In Gonzalo Navarro and Vladimir Pestov, editors, *Similarity Search and Applications*, pages 240–241, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[3] Jakub Lokoč, Tomáš Grošup, and Tomáš Skopal. On scalable approximate search with the signature quadratic form distance. In Nieves Brisaboa, Oscar Pedreira, and Pavel Zezula, editors, *Similarity Search and Applications*, pages 312–318, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[4] J. Lokoč, T. Grošup, P. Čech, and T. Skopal. Towards efficient multimedia exploration using the metric space approach. In *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–4, June 2014.

[5] P. Čech and T. Grošup. Comparison of metric space browsing strategies for efficient image exploration. In *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, June 2015.

[6] Tomáš Grošup, Přemysl Čech, Jakub Lokoč, and Tomáš Skopal. A web portal for effective multi-model exploration. In Xiangjian He, Suhuai Luo, Dacheng Tao, Changsheng Xu, Jie Yang, and Muhammad Abul Hasan, editors, *MultiMedia Modeling*, pages 315–318, Cham, 2015. Springer International Publishing.

[7] T. Grošup, J. Moško, and P. Čech. Continuous hierarchical exploration of multimedia collections. In *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–4, June 2015.

[8] Juraj Moško, Jakub Lokoč, Tomáš Grošup, Přemysl Čech, Tomáš Skopal, and Jan Lánský. Mles: Multilayer exploration structure for multimedia exploration. In Tadeusz Morzy, Patrick Valduriez, and Ladjel Bellatreche, editors, *New Trends in Databases and Information Systems*, pages 135–144, Cham, 2015. Springer International Publishing.

[9] Juraj Moško, Jakub Lokoč, Tomáš Grošup, Přemysl Čech, Tomáš Skopal, and Jan Lánský. Evaluating multilayer multimedia exploration. In Giuseppe

Amato, Richard Connor, Fabrizio Falchi, and Claudio Gennaro, editors, *Similarity Search and Applications*, pages 162–169, Cham, 2015. Springer International Publishing.

[10] Jakub Lokoč, Tomáš Grošup, Přemysl Čech, Tomáš Pevný, and Tomáš Skopal. Malware discovery using behaviour-based exploration of network traffic. In Christian Beecks, Felix Borutta, Peer Kröger, and Thomas Seidl, editors, *Similarity Search and Applications*, pages 315–323, Cham, 2017. Springer International Publishing.

[11] Tomáš Skopal, Ladislav Peška, Gregor Kovalčík, Tomáš Grošup, and Jakub Lokoč. Product exploration based on latent visual attributes. In *Proc. of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 2531–2534. ACM, 2017.

[12] Tomás Skopal, Ladislav Peska, and Tomás Grošup. Interactive product search based on global and local visual-semantic features. In *Similarity Search and Applications - 11th Int. Conference, SISAP 2018, Lima, Peru*, pages 87–95, 2018.

[13] Tomás Grosup, Ladislav Peska, and Tomás Skopal. Towards augmented database schemes by discovery of latent visual attributes. In *Advances in Database Technology - 22nd International Conference on Extending Database Technology, EDBT 2019, Lisbon, Portugal, March 26-29, 2019*, pages 670–673, 2019.

[14] Tomáš Grošup. Methods of multi-modal data exploration. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, ICMR '19, pages 34–37, New York, NY, USA, 2019. ACM.

[15] Tomáš Grošup, Ladislav Peška, and Tomáš Skopal. On augmenting database schemes by latent visual attributes. Submitted to Information Systems journal, 2019.

[16] Nawsher Khan, Ibrar Yaqoob, Ibrahim Abaker, Ibrahim Hashem, Zakira Inayat, Waleed Kamaleldin, Mahmoud Ali, Muhammad Alam, Muhammad Shiraz, and Abdullah Gani. Big data: Survey, technologies, opportunities, and challenges. 06 2013.

[17] Tay Vaughn. *Multimedia Making It Work*. McGraw-Hill, Inc., New York, NY, USA, 1993.

[18] Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, and Jeffrey Heer. Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2917–2926, December 2012.

[19] Stratos Idreos, Olga Papaemmanouil, and Surajit Chaudhuri. Overview of data exploration techniques. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, SIGMOD '15, pages 277–281, New York, NY, USA, 2015. ACM.

[20] I. J. Good. The philosophy of exploratory data analysis. *Philosophy of Science*, 50(2):283–295, 1983.

[21] R. Bayer and E. M. McCreight. Organization and maintenance of large ordered indexes. *Acta Informatica*, 1(3):173–189, Sep 1972.

[22] Jan L. Harrington. 3 - introduction to sql. In Jan L. Harrington, editor, *SQL Clearly Explained (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems, pages 65 – 74. Morgan Kaufmann, Boston, third edition edition, 2010.

[23] T Dharani and Laurence Aroquiaraj. A survey on content based image retrieval. pages 485–490, 02 2013.

[24] A. H. M. ter Hofstede, H. A. Proper, and Th. P. van der Weide. Query Formulation as an Information Retrieval Problem. *The Computer Journal*, 39(4):255–274, 01 1996.

[25] Shih-Fu Chang, T. Sikora, and A. Purl. Overview of the mpeg-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):688–695, Jun 2001.

[26] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov 2004.

[27] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. In Aleš Leonardis, Horst Bischof, and Axel Pinz, editors, *Computer Vision – ECCV 2006*, pages 404–417, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[28] Yann LeCun, Y Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[30] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 818–833. Springer, 2014.

[31] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.

[32] Samira Pouyanfar, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu Ching Chen, and S. S. Iyengar. A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys*, 51(5), 8 2018.

[33] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11 – 26, 2017.

[34] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. *CoRR*, abs/1310.1531, 2013.

[35] Philipp Fischer, Alexey Dosovitskiy, and Thomas Brox. Descriptor matching with convolutional neural networks: a comparison to sift. *CoRR*, abs/1405.5769, 2014.

[36] Vittorio Ferrari and Andrew Zisserman. Learning visual attributes. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 433–440. Curran Associates, Inc., 2008.

[37] Sirion Vittayakorn, Takayuki Umeda, Kazuhiko Murasaki, Kyoko Sudo, Takayuki Okatani, and Kota Yamaguchi. Automatic attribute discovery with neural activations. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 252–268, Cham, 2016. Springer International Publishing.

[38] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, 40, 01 2008.

[39] Jakub Lokoč, David Novák, Michal Batko, and Tomáš Skopal. Visual image search: Feature signatures or/and global descriptors. In Gonzalo Navarro and Vladimir Pestov, editors, *Similarity Search and Applications*, pages 177–191, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[40] Pavel Zezula, Giuseppe Amato, Vlastislav Dohnal, and Michal Batko. *Similarity Search: The Metric Space Approach (Advances in Database Systems)*. Springer-Verlag, Berlin, Heidelberg, 2005.

[41] Leonid Boytsov and Bilegsaikhan Naidan. Engineering efficient and effective non-metric space library. In Nieves Brisaboa, Oscar Pedreira, and Pavel Zezula, editors, *Similarity Search and Applications*, pages 280–293, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[42] Natalia Miranda, Edgar Chávez, María Fabiana Piccoli, and Nora Reyes. (very) fast (all) k-nearest neighbors in metric and non metric spaces without indexing. In Nieves Brisaboa, Oscar Pedreira, and Pavel Zezula, editors, *Similarity Search and Applications*, pages 300–311, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[43] Tomáš Skopal and Jakub Lokoč. Nm-tree: Flexible approximate similarity search in metric and non-metric spaces. In Sourav S. Bhowmick, Josef Küng, and Roland Wagner, editors, *Database and Expert Systems Applications*, pages 312–325, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[44] Tomáš Skopal. Unified framework for fast exact and approximate search in dissimilarity spaces. *ACM Trans. Database Syst.*, 32(4), November 2007.

[45] R. B. MARIMONT and M. B. SHAPIRO. Nearest Neighbour Searches and the Curse of Dimensionality. *IMA Journal of Applied Mathematics*, 24(1):59–70, 08 1979.

[46] Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates, and José Luis Marroquín. Searching in metric spaces. *ACM Comput. Surv.*, 33(3):273–321, September 2001.

[47] Gonzalo Navarro, Rodrigo Paredes, Nora Reyes, and Cristian Bustos. An empirical evaluation of intrinsic dimension estimators. *Information Systems*, 64:206 – 218, 2017.

[48] Arnab Nandi and H V. Jagadish. Guided interaction: Rethinking the query-result paradigm. *PVLDB*, 4:1466–1469, 08 2011.

[49] Chaoli Wang, John P. Reese, Huan Zhang, Jun Tao, Yi Gu, Jun Ma, and Robert J. Nemiroff. Similarity-based visualization of large image collections. *Information Visualization*, 14, 01 2013.

[50] Giang NGuyên and Marcel Worring. Interactive access to large image collections using similarity-based visualization. *Journal of Visual Languages and Computing Journal of Visual Languages and Computing*, 19:203–224, 04 2008.

[51] Kai Uwe Barthel, Nico Hezel, and Radek Mackowiak. Imagemap - visually browsing millions of images. In Xiangjian He, Suhuai Luo, Dacheng Tao, Changsheng Xu, Jie Yang, and Muhammad Abul Hasan, editors, *MultiMedia Modeling*, pages 287–290, Cham, 2015. Springer International Publishing.

[52] Christian Beecks, Tomáš Skopal, Klaus Schoeffmann, and Thomas Seidl. Towards large-scale multimedia exploration. In Gautam Das, Vagelis Hsristidis, and Ihab Ilyas, editors, *Proceedings of the 5th International Workshop on Ranking in Databases (DBRank 2011)*, pages 31–33, Seattle, WA, USA, aug 2011. VLDB.

[53] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Softw., Pract. Exper.*, 21:1129–1164, 1991.

[54] David Novak and Michal Batko. Metric index: An efficient and scalable solution for similarity search. *2009 Second International Workshop on Similarity Search and Applications*, pages 65–73, 2009.

[55] Tomás Skopal, Jaroslav Pokorný, and Václav Snásel. Pm-tree: Pivoting metric tree for similarity search in multimedia databases. In *ADBIS*, 2004.

[56] Jakub Lokoč, Juraj Moško, Přemysl Čech, and Tomáš Skopal. On indexing metric spaces using cut-regions. *Inf. Syst.*, 43(C):1–19, July 2014.

[57] Fakhri Karray, Milad Alemzadeh, Jamil Abou Saleh, and Mo Nours Arab. Human-computer interaction: Overview on state of the art. *International Journal on Smart Sensing and Intelligent Systems*, 1:137–159, 01 2008.

[58] Benjamin Bustos and Tomáš Skopal. Dynamic similarity search in multi-metric spaces. In *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, MIR '06, pages 137–146, New York, NY, USA, 2006. ACM.

[59] Tamara L. Berg, Alexander C. Berg, and Jonathan Shih. Automatic attribute discovery and characterization from noisy web data. In *Proceedings of the 11th European Conference on Computer Vision: Part I*, ECCV'10, pages 663–676, Berlin, Heidelberg, 2010. Springer-Verlag.

[60] X. Chen, A. Shrivastava, and A. Gupta. Neil: Extracting visual knowledge from web data. In *2013 IEEE International Conference on Computer Vision*, pages 1409–1416, Dec 2013.

[61] Santosh Kumar Divvala, Ali Farhadi, and Carlos Guestrin. Learning everything about anything: Webly-supervised visual concept learning. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3270–3277, 2014.

[62] Bolei Zhou, Vignesh Jagadeesh, and Robinson Piramuthu. Conceptlearner: Discovering visual concepts from weakly labeled image collections. *CoRR*, abs/1411.5328, 2014.

[63] Ladislav Peska, Tomas Grosup, Gregor Kovalcik, Jakub Lokoc, and Tomas Skopal. Vadet: Visual attributes exploration and discovery tool, 2017.

[64] Si Liu, Jiashi Feng, Zheng Song, Tianzhu Zhang, Hanqing Lu, Changsheng Xu, and Shuicheng Yan. Hi, magic closet, tell me what to wear! In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 619–628, New York, NY, USA, 2012. ACM.

[65] Si Liu, Zheng Song, Guangcan Liu, Changsheng Xu, Hanqing Lu, and Shuicheng Yan. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. pages 3330–3337, 10 2012.

[66] Huizhong Chen, Andrew Gallagher, and Bernd Girod. Describing clothing by semantic attributes. In Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato, and Cordelia Schmid, editors, *Computer Vision – ECCV 2012*, pages 609–623, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

List of Figures

List of Tables

List of Publications

Following references are all published and in-review articles which were authored or co-authored by Tomáš Grošup. The works are sorted chronologically and also include work published during bachelor (2009-2012) and master (2012-2014) studies of the author.

Jakub Lokoč, Tomáš Grošup, and Tomáš Skopal. Image exploration using online feature extraction and reranking. In *Proceedings of the 2Nd ACM International Conference on Multimedia Retrieval*, ICMR '12, pages 66:1–66:2, New York, NY, USA, 2012. ACM

Jakub Lokoč, Tomáš Grošup, and Tomáš Skopal. Sir: The smart image retrieval engine. In Gonzalo Navarro and Vladimir Pestov, editors, *Similarity Search and Applications*, pages 240–241, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg

Jakub Lokoč, Tomáš Grošup, and Tomáš Skopal. On scalable approximate search with the signature quadratic form distance. In Nieves Brisaboa, Oscar Pedreira, and Pavel Zezula, editors, *Similarity Search and Applications*, pages 312–318, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg

J. Lokoč, T. Grošup, P. Čech, and T. Skopal. Towards efficient multimedia exploration using the metric space approach. In *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–4, June 2014

P. Čech and T. Grošup. Comparison of metric space browsing strategies for efficient image exploration. In *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, June 2015

Tomáš Grošup, Přemysl Čech, Jakub Lokoč, and Tomáš Skopal. A web portal for effective multi-model exploration. In Xiangjian He, Suhuai Luo, Dacheng Tao, Changsheng Xu, Jie Yang, and Muhammad Abul Hasan, editors, *MultiMedia Modeling*, pages 315–318, Cham, 2015. Springer International Publishing

T. Grošup, J. Moško, and P. Čech. Continuous hierarchical exploration of multimedia collections. In *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pages 1–4, June 2015

Juraj Moško, Jakub Lokoč, Tomáš Grošup, Přemysl Čech, Tomáš Skopal, and Jan Lánský. Mles: Multilayer exploration structure for multimedia exploration. In Tadeusz Morzy, Patrick Valduriez, and Ladjel Bellatreche, editors, *New Trends* *in Databases and Information Systems*, pages 135–144, Cham, 2015. Springer International Publishing

---

Juraj Moško, Jakub Lokoč, Tomáš Grošup, Přemysl Čech, Tomáš Skopal, and Jan Lánský. Evaluating multilayer multimedia exploration. In Giuseppe Amato, Richard Connor, Fabrizio Falchi, and Claudio Gennaro, editors, *Similarity Search and Applications*, pages 162–169, Cham, 2015. Springer International Publishing

---

Jakub Lokoč, Tomáš Grošup, Přemysl Čech, Tomáš Pevný, and Tomáš Skopal. Malware discovery using behaviour-based exploration of network traffic. In Christian Beecks, Felix Borutta, Peer Kröger, and Thomas Seidl, editors, *Similarity Search and Applications*, pages 315–323, Cham, 2017. Springer International Publishing

---

Tomáš Skopal, Ladislav Peška, Gregor Kovalčík, Tomáš Grošup, and Jakub Lokoč. Product exploration based on latent visual attributes. In *Proc. of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 2531–2534. ACM, 2017

Tomás Skopal, Ladislav Peska, and Tomás Grošup. Interactive product search based on global and local visual-semantic features. In *Similarity Search and Applications - 11th Int. Conference, SISAP 2018, Lima, Peru*, pages 87–95, 2018

Tomás Grosup, Ladislav Peska, and Tomás Skopal. Towards augmented database schemes by discovery of latent visual attributes. In *Advances in Database Technology - 22nd International Conference on Extending Database Technology, EDBT 2019, Lisbon, Portugal, March 26-29, 2019*, pages 670–673, 2019

Tomáš Grošup. Methods of multi-modal data exploration. In *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, ICMR '19, pages 34–37, New York, NY, USA, 2019. ACM

Tomáš Grošup, Ladislav Peška, and Tomáš Skopal. On augmenting database schemes by latent visual attributes. Submitted to Information Systems journal, 2019

Appendix A

Attachments

All attachments to this thesis are available electronically under publicly available URLs. Table 1 lists URLs for source code and deployed versions of the software.

<table>
<tr><th>Name</th><th>Type</th></tr>
<tr><td>Multimedia Exploration</td><td>SW</td></tr>
<tr><td colspan="2">http://herkules.ms.mff.cuni.cz/</td></tr>
<tr><td>Multimedia Exploration</td><td>Code</td></tr>
<tr><td colspan="2">https://subversion.assembla.com/svn/multimedia-exploration</td></tr>
<tr><td>Find-the-image</td><td>SW</td></tr>
<tr><td colspan="2">http://herkules.ms.mff.cuni.cz/find-the-image</td></tr>
<tr><td>Attribute Discovery</td><td>SW</td></tr>
<tr><td colspan="2">http://herkules.ms.mff.cuni.cz/vadet-admin/</td></tr>
<tr><td>Visual attributes</td><td>SW</td></tr>
<tr><td colspan="2">http://herkules.ms.mff.cuni.cz/vadet-merged</td></tr>
<tr><td>Attribute Discovery UI</td><td>Code</td></tr>
<tr><td colspan="2">https://github.com/T-Gro/VADET-Admin</td></tr>
<tr><td>Attribute Discovery preprocessing</td><td>Code</td></tr>
<tr><td colspan="2">https://github.com/T-Gro/Visual-Attribute-Filtering-Scripts</td></tr>
</table>

Table A.1: Table listing relevant attachments which are freely available online. It distinguishes between source code (Type=Code) and deployed versions of the software (Type=SW)