Faculty of Informatics
Masaryk University

Data-Driven Exploratory Interactions and Visual Analysis

Habilitation Thesis
(Collection of Articles)

Radek Ošlejšek

August 2018
Brno, Czech Republic

Abstract

Current information technologies enable us to collect a huge amount of data. However, understanding the data, revealing their hidden relationships, mediating them to users meaningfully, and utilizing them for the optimization of related processes still present significant obstacles. One generic approach to solving this challenge is to provide users with methods of guided data exploration. Based on the initial overview of data content and scope, users should be able to continuously formulate hypotheses related to the data and delimit their area of interest. They should be able to explore details on demand so that they can verify their hypotheses or gain insight into the specific pieces of information.

The suggested problem domain is too vast to be covered by a single research activity. Therefore, this work restricts its scope to a specific sub-domain. This thesis does not deal with data collection, storage, size, certainty, accuracy, privacy, or other important aspects of big data processing. Instead, the thesis focuses on issues related to data comprehension and interpretation, semantics, and filtering – aspects that are crucial for user-driven interactive data analysis. The major part of the thesis is devoted to the research of exploratory strategies and techniques for efficient data analysis in two application domains: graphical data and network security data. With graphical data, our research aims at developing methods of dialogue-based interaction with pictures. With network security data, we aim to develop interactive visual analytics tools. Both approaches, dialogues as well as visualizations, share many principles, including semantic modeling of underlying data and tactics for continuous information seeking.

The thesis is structured as a collection of articles accompanied by commentary putting my contributions in the context of the state of the art in the area and linking together the two research areas to an integrated view: a dialogue interaction with graphical data and visual interaction with cybersecurity data.

Keywords: Data semantics, knowledge modeling, exploratory interaction, visual analytics.

Abstrakt

Současné informační technologie umožňují sbírat velké množství dat. S tím souvisí snaha v datech se vyznat, odhalit jejich skryté vazby, zprostředkovat jejich význam uživatelům a využít je pro optimalizaci činností. Jedním z obecných způsobů, jak tuto problematiku řešit, je nabídnout uživatelům metody postupného řízeného prohledávání dat. Cílem je, aby na začátku byli uživatelé schopni získat celkovou představu o obsahu a rozsahu dostupných informací a na základě toho si mohli vytvořit vlastní hypotézy či vyjasnit oblasti svého zájmu. Poté by jim mělo být umožněno postupně se zaměřit na detaily tak, aby byli schopni prověřovat hypotézy nebo detailněji zkoumat specifické oblasti nebo vztahy.

Celá problematika je příliš rozsáhlá na to, aby se dala pokrýt v rámci jednoho výzkumného zaměření. I tato práce je tedy omezena na dílčí oblasti. Práce se nezabývá problematikou získávání a uchovávání dat, jejich velikostí, spolehlivostí, přesností, soukromím, ani dalšími důležitými aspekty, které se zpracováním velkých dat souvisejí. Práce se naopak zaměřuje na problematiku související až s významem dat, jejich interpretací, sémantikou a filtrováním pro potřeby interaktivní datové analýzy prováděné uživateli. Velká část práce se zabývá právě problematikou návrhu explorativních strategií a technik, pro efektivní analýzu dat, a to ve dvou konkrétních aplikačních doménách: obrazová data (například obsah fotografií) a data z oblasti síťové bezpečnosti. V případě obrazových dat bylo našim cílem vyvinout postupy pro interakci s obrázky pomocí dialogů v přirozeném jazyce. V případě bezpečnostních dat usilujeme o vývoj uživatelských analytických vizualizací. Oba tyto přístupy, dialogy a vizualizace, spojují podobné principy, které je nutné uplatňovat. Jedná se například o sémantického modelování dat, nebo o principy podporující postupné vyhledávání informací známé pod pojmem “information seeking strategies”.

Tato práce je souborem publikovaných vědeckých prací doprovozených komentářem, který mé výsledky zasazuje do kontextu aktuálního stavu výzkumu v této oblasti, a propojuje oblast dialogové komunikace nad obrazovými daty s oblastí vizuální interakce nad bezpečnostními daty do integrovaného pohledu.

Klíčová slova: Sémantika dat, znalostní modely, explorativní interakce, vizuální analýza.

Acknowledgements

I would like to express my appreciation to all the mentors I have had along my journey—to Jiří Sochor for guiding me during my Ph.D. studies, to Ivan Kopeček for helping me to dive into the field of semantics modeling and dialogue systems, and to Tomáš Pitner for welcoming me warmly in the LaSArIS lab after I joined FI MU as an assistant professor. I would like to thank all my colleagues, co-authors, LaSArIS members, and KYPO project participants for being such a great team to work with.

Finally and foremost, I wish to thank my family and close friends for their support, understanding, and patience.

Radek Ošlejšek

Contents

Part I

Commentary

Chapter 1

Introduction

Current technologies enable us to collect a huge amount of data. However, the processing, understanding, and analysis of the data still present a significant and challenging problem. Techniques for efficient data exploration have to cope with two primary issues: the amount of data and the complexity of the information. Users dealing with high volumes of data can be easily overloaded by information and thereby prevented from focusing on relevant parts. Similarly, the complexity of the data can make it difficult to understand a problem domain and to uncover hidden relationships. Our approach to dealing with these issues is based on requirements analyses in particular application domains, in-depth understanding of available data, semantic modeling, and the development of efficient techniques and strategies for interactive data exploration enabling us to serve information gradually. We focus on two types of exploratory concepts: dialogues and visual interaction.

Dialogue-based communication has become very popular because it represents a natural method of human communication. Voice control is now widely used in cars, smartphones, electronic voice assistants, etc. For visually impaired people, interaction in natural language has always played an important role in data accessibility. In particular, the accessibility of graphical content is simultaneously a matter of primary importance and a significant obstacle. Therefore, one research branch and part of the thesis is devoted to the dialogue-based exploration of graphical content.

Visualizations represent another widely adopted method of data exploration and analysis. In 1996, Ben Shneiderman [68] addressed the problem of information overload in data visualization, formulating the "information-seeking mantra": overview first, zoom and filter, then details-on-demand. Although this concept was invented primarily for visualizations, its ideas apply to many other interactive techniques that struggle with data complexity and information overload, including dialogue-based communication. Another term widely used in the context of interactive visualizations is Visual Analytics [78]. Approaches to visual analytics cover the complete process of analytical reasoning supported by interactive visual

interfaces. They are applied in various fields, from biology or weather forecasts [43, 42, 18, 19] to education [71, 70]. In our research, we focus on cybersecurity training and education.

1.1 Goals and Structure of the Thesis

This thesis summarizes my contributions to the progress within the field of developing exploratory techniques for gaining insight into complex, unstructured data. The text is structured with regard to my contributions to two mutually connected research areas: interactive exploration of images and exploratory visual analysis in cyber security.

The order in which the information is presented reflects the genesis of my research. The first part of the thesis is devoted to a dialogue-based exploration of images. In Section 2, we focus on formal semantic modeling of graphical data and its use in dialogue-based interaction, non-verbal exploration, and continuous dialogue-driven knowledge extension and management.

Over the years in which I studied the field of dialogue-based exploration of images, I gained experience in using formal ontologies for semantic modeling and information filtering as well as in the design of exploratory strategies for user interaction. The results obtained during this research inspired me to adapt dialogue-based approaches for visual-based explorations and visual analytics in cyber security. Therefore, the second part of the thesis presents our achievements in this research. In Section 3, I present our cyber security testbed and my contributions to the visual analytics in cyber security training.

Each section starts with the presentation of the state of the art within the particular domain, followed by my contributions to its progress and a list of selected articles I have co-authored and that are attached to this text to exemplify my contributions.

The collection of articles is listed in Part II of this thesis.

Chapter 2

Interactive Exploration of Images

Pictures represent windows to the world behind them. This world consists of visible elements and potentially interesting but invisible data. For instance, in a scene with a castle, the castle's history, opening hours, and other pieces of hidden information could be important or interesting for viewers. Within our research, we aimed at providing insight into both visible and invisible information using non-visual interaction techniques that would help visually impaired people to access graphical content. The results of our research have been published primarily at conferences related to assistive technologies.

To cope with this task, we combine dialogue-based exploration with non-verbal communication and navigation. Our approach is based on the research in several fields, as discussed in what follows:

Semantic modeling. To handle the semantic information of depicted objects, it was necessary to propose semantic models that can encode invisible pieces of information. In dialogue-based image exploration, these semantic models have to be well structured according to specific requirements of dialogue strategies. They also have to be formally defined so that it is possible to generate interaction strategies from the semantic model regardless of particular graphical content.

Knowledge management. Gathering the semantic data and assigning them to concrete pictures are exhaustive long-term processes. Moreover, as multiple images depicting similar contents represent multiple windows to the same word, it is natural to create a single shared knowledge base. Techniques for sharing, extending, and managing the knowledge base had to be developed.

Image annotations. While a knowledge base represents a shared dataset capturing the meaning of generic terms, annotations are used to assign meaning to particular objects

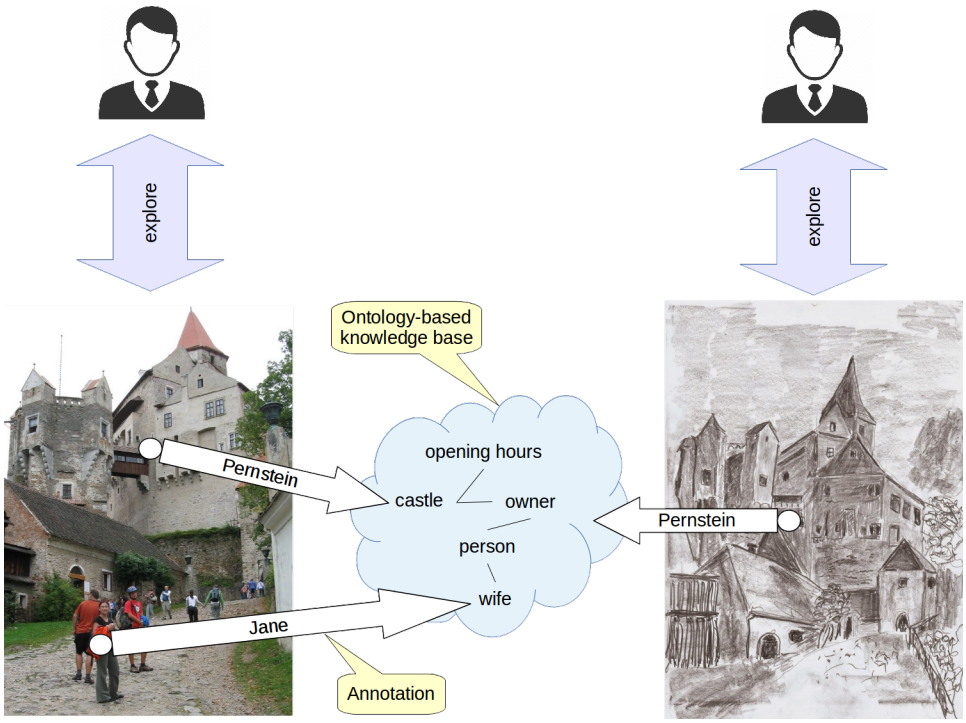

Figure 2.1: Scheme of interactive exploration of formally annotated images.

depicted in a picture. Encoding the annotation data into existing images poses a rather technical obstacle that had to be overcome. The process of annotation represents a more serious conceptual problem. Manual annotation of each picture is time-consuming and does not fit the idea of making a lot of images that represent multiple windows to a single world that is well-described by a shared knowledge base. To fulfill our expectations, techniques of automated image classification, object recognition, and semantics assignment had to be incorporated to achieve a (semi)automated annotation process.

Information filtering. A big knowledge base provides many possibilities for information retrieval. It is necessary to avoid overloading users with information. Suitable approaches that filter out non-relevant data had to be employed so that users could focus their insight. Filtering tactics are highly related to dialogue strategies that would provide users with an overview first and then enable them to explore details on demand iteratively.

The design of a unified dialogue system. A dialogue system represents an exploratory user interface. It has to be unified in the sense that nothing should be assumed about the picture content. On the contrary, the system should deal with arbitrary picture domains and should be driven by only the structure of the provided knowledge base.

The rest of this chapter discusses the aforementioned cross-cutting issues from three different perspectives. Section 2.1 focuses on the dialogue-based exploration of manually

annotated images. Ontology-based semantic models are used for knowledge management. Section 2.2 extends this approach with non-verbal techniques of interaction and navigation, and Section 2.3 discusses the concept of automated annotations and continuous knowledge development based on learning from dialogues.

2.1 Dialogue-Based Exploration

Unlike approaches using tactile feedback, which have been of primary interest to researchers in the past [21, 44, 65, 35], a dialogue-based exploration of graphics does not need any special devices. To mediate graphical content using natural language, we had to complete two primary tasks.

First, it was necessary to design a suitable semantic model that would be able to store semantics values in a well-structured automatically processed form. Our approach is based on formal OWL/RDF ontologies. Ontology-based annotation of photo collections was discussed in [66]. The authors of this paper used ontologies to classify, store, and search relevant photos rather than to structurally annotate its content. The utilization of ontologies for annotation and for providing access to structured vector graphics in various contexts has been explored, for instance, in [27, 46]. However, at that time, the approaches were suitable only for well-designed vector graphics such as graphs or diagrams integrating annotation data inside the scene graph structure. Our goal was more ambitious: to support common raster images such as photos, which represent the majority of visual content. Therefore, we came up with a technical solution that makes it possible to store structured annotation data in original raster images transparently without the need to modify them. We also had to solve the problem with ontology abstraction. In general, ontologies do not restrict the abstraction used to describe the real world. The same domain can be described from various perspectives. A dialogue-based exploration requires a specific semantic structure and abstraction. Therefore, we defined rules and constraints for ontologies that would be utilized by ontology-based knowledge models and annotations, and we proposed a domain-independent upper ontology [76] that was designed for the dialogue-based exploration of pictures. This ontology is used as a primary semantic structure that joins domain-specific ontologies handling picture content into a single uniform knowledge base. Current works following the issues of ontology-based semantic models include [52, 53].

The second task of the dialogue-based exploration of annotated images was the design of a dialogue system. Because we do not restrict the content of explored pictures, the system has to be generic in the sense that it has to be able to understand domain-specific questions and deal with content-specific dialogues. We aimed to leverage formal ontologies to help the dialogue system to understand domain-specific questions and to drive the dialogue strategies. Some approaches that enhance the efficiency of the dialogue manager exploiting the knowledge bases have been explored and published [56, 58, 49, 5]. It was argued in [24]

that the separation of dialogue management from knowledge management reduces the complexity of the systems and enhances further extensions. To achieve the required generality of the dialogue system, we used the principles of pattern extraction [31, 62] and frame prototyping [50]. These methods enabled us to adapt the system to particular picture content quickly. We also designed a simple yet generic querying language connected to our upper graphical ontology that enables users to retrieve domain-independent graphical information, such as the position of an object in the picture, its relative size, etc.

Contributions

During the early stages of the research, we proposed a novel theoretical concept for dialogue-based image generation and exploration. This concept was based on the idea that well-structured semantic data driven by ontologies can be used for efficient dialogue-based information retrieval using a formal model of Pawlak information systems. The meaningfulness of this approach was discussed and tested by blind students, for whom the dialogue-based image generation and exploration would be the most beneficial. The testing was performed using Wizard of Oz simulations. This proof of concept was published in [37].

Over the next few years, we focused primarily on dialogue-based exploration and elaborated the preliminary idea into necessary details. We proposed taxonomies and ontologies that would be suitable for the semantic description and exploration of graphical content via natural language. We created a generic graphical ontology dealing with visual aspects of graphical content and classifying depicted objects according to their relative size, shape, position, etc. We proposed what we called the *What-Where Language*. This fragment of natural language, with relatively simple grammar, was designed according to the graphical ontology and other specific requirements put on the exploration of graphical content. We also proposed a technical solution enabling us to integrate ontology-based annotations into existing raster and vector images. Our results were summarized and published in [38].

In 2014, we implemented an experimental system called *GATE — Graphics Accessible To Everyone*. This component-based web application provided several services. A semantic module dealt with shared semantic knowledge related to the possible content of uploaded images and provided services for semantic inspection. Users were able to upload an annotated image to the system and then explore it interactively by writing questions in What-Where Language. The underlying dialogue engine leveraged the picture annotation and the knowledge dataset to provide a smooth dialogue, including standard techniques for addressing misunderstandings. The results of the system evaluation were published in [30].

Papers in Collection

[37] I. Kopeček and R. Ošlejšek. Creating pictures by dialogue. In *Computers Helping People with Special Needs: 10th International Conference, ICCHP 2006*, pages 61–68,

Berlin, 2006. Springer-Verlag.

I participated in the elaboration of the basic theoretical concept and in the paper writing. Contribution 50%.

[38] I. Kopeček and R. Ošlejšek. Gate to accessibility of computer graphics. In *Computers Helping People with Special Needs: 11th International Conference, ICCHP 2008*, pages 295–302, Berlin, 2008. Springer-Verlag.

I proposed the semantic structure and its integration to images. I participated in the design of communication strategies and architecture of the GATE system. I wrote corresponding parts of the paper. Contribution 60%.

[30] P. Hamřík, I. Kopeček, R. Ošlejšek, and J. Plhák. Dialogue-based information retrieval from images. In *Computers Helping People with Special Needs:14th International Conference, ICCHP 2014*, pages 85–92, Switzerland, 2014. Springer International Publishing.

I was responsible for design, implementation, and experimental evaluation of the system. I participated in the paper writing. Contribution 25%.

2.2 Non-Verbal Interaction

Communication with images via dialogues in natural language is limited to annotations that have to be available in the images. Searching for another approach to annotation-independent image navigation and exploration was, therefore, also within the interest of our research.

In [36], Kamel et al. proposed grid-based navigation, where an image is divided into nine labeled areas. The uniform lattice that is created can be used for efficient reference to approximate location, for instance via keyboard commands. Moreover, dividing image space recursively makes it possible to focus more precisely.

A lot of effort has been made to transform color information into sounds [15, 47]. For a gray bitmap image, this method was used by Jones in the vOICe system [34] to transform the vertical and horizontal position of pixels into pitch and corresponding stereo panning, respectively; brightness was represented by loudness. Some other approaches were directed to more specialized graphics, such as pie charts [25] and line graphs [8]. Franklin and Roberts presented a general path-based model for these approaches [26]; Brown et al. [10] discussed the role of annotation for the sonification of graph-based diagrams.

In our work, we integrated a navigation grid and a sonification approach to dialogue-based image exploration. A recursive navigation grid with labeled areas is used for rapid and straightforward reference to image locations in dialogues. Our sonification algorithm enhances dialogues with the possibility to explore unannotated parts of the image.

Contributions

We proposed and implemented a novel method of sonification of complex graphical objects, such as color photographs, based on a hybrid approach combining sound and speech communication. The transformation of colors into sounds is supported by a special color model called the *semantic color model*. The integration of a recursive navigation grid into dialogue strategies simplified the navigational tasks. Our results were published in [39]

Papers in Collection

[39] I. Kopeček and R. Ošlejšek. Hybrid approach to sonification of color images. In *The 2008 International Conference on Convergence and Hybrid Information Technologies*, volume 2, pages 722–727. IEEE Computer Society, 2008.

I participated in the design of the algorithm and was responsible for valuation. I was responsible for paper writing. Contribution 50%.

2.3 Dialogue-Driven Knowledge Management

Dialogue-based image exploration requires well-structured annotations to be prepared for images. We were aware that building such annotations and knowledge base is laborious, and we strove to develop techniques for their automatic creation.

Image recognition and auto-detection algorithms present suitable methods of automatically recognizing graphical content and providing basic semantic information and classification. At the time that we published our research, particularly the techniques of face recognition [29, 6] and similarity search algorithms in large image collections [33, 4] were well-developed and applicable for the initial image content retrieval. At that time, the utilization of EXIF and GPS metadata included in photographs for image classification and geolocation were at the forefront of interest of researchers [64, 9, 79].

Our research, however, aimed at leveraging natural features of dialogue interactions for information retrieval. We focused on crowd-sourcing techniques where the knowledge is built continuously and shared across users (across images) during the communication. Our approach is based on the idea that if the dialogue system can take the initiative during the communication and actively request missing pieces of information from the user, then the knowledge can be built continuously and shared across users. In this way, the dialogue system is able to learn from the interaction.

Contributions

In [40], we addressed the issues of suitable ontology structures usable for the automatic classification of annotated objects within an ontology. Based on the discussion of the optimality of ontologies in relation with the annotation process, we proposed optimality measures enabling us to solve this problem algorithmically.

In [41], we elaborated the idea of self-learning images and classified dialogue strategies into three categories. The *image information supplementing mode* discussed in the paper represents a strategy where the dialogue system is proactive and is able to supply the image with missing pieces of information. In the paper, we also proposed a concrete grammar structure for queries.

Papers in Collection

[40] I. Kopeček and R. Ošlejšek. Annotating and describing pictures – applications in e-learning and accessibility of graphics. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010*, pages 124–130, Berlin, 2010. Springer-Verlag.

I proposed and developed a graphical ontology and contributed to the concepts of automated picture annotation. I wrote corresponding parts of the paper. Contribution 50%.

[41] I. Kopeček, R. Ošlejšek, and J. Plhák. Integrating dialogue systems with images. In *Text, Speech and Dialogue. 15th International Conference, TSD 2012*, pages 632–639, Berlin Heidelber, 2012. Springer-Verlag.

I participated in the design of question analysis methods and was responsible for paper writing. Contribution 33%.

Chapter 3

Visual Exploration and Analysis in Cybersecurity

Information and communication systems are exposed to an increasing number of attacks. Apart from simple attacks conducted by hackers and inexperienced individuals who can be tracked down [60], there are professional teams backed by organized crime groups or even governments [45] that carefully hide their activities. A shortage in cybersecurity skills and cybersecurity professionals is a critical vulnerability for companies and nations [11, 13].

Techniques of exploratory visual analysis can significantly help to cope with these threats. In our research, we aim at using a proper visual representation of complex cybersecurity data so that security experts can reveal attack paths, analyze the behavior of network users facing cyber attacks, or efficiently learn from cyber defense training programs and simulations.

The approaches used to reach this goal are in many aspects similar to the approaches used for the interactive exploration of images. Providing insight into cybersecurity processes is also based on the thorough classification of data and its semantics and on the design of analytical exploratory interactions. The differences are in the application domain (graphical content vs. cybersecurity processes) and interaction techniques (dialogue-based vs. visual-based interaction). However, the primary difference is in the scale of analytical tasks. While the goal of the dialogue-based interaction with images was to describe the content of a picture interactively, the goals of the interaction over cybersecurity data significantly differ depending on the analytical tasks. For example, the same network data gathered over a certain period has to be mediated in a different way to an analyst who aims to find out if an attack happened and in another way to an analyst seeking detailed evidence of a particular attack during a forensic investigation. This variability of interaction goals leads to the variability of exploratory visualizations that have to be precisely adapted to individual tasks, processes, and expectations.

The rest of this chapter is divided into two sub-section. Section 3.1 discusses data and processes in the cybersecurity domain and puts them in a broader context of cyber exercises and research platforms that provide frameworks for cyber experiments and visual analysis. Our KYPO Cyber Range and its technical solution to data storage and retrieval are discussed in detail. Section 3.2 presents our results in the field of visualizations and visual analysis. At the current stage of our research, we focus primarily on cyber defense training. Therefore, visualizations gaining insight into cyber exercises and helping to improve their impact are presented in this section.

3.1 Cyber Exercise and Research Platforms

Operational networks are not suitable for building and studying knowledge of cyber threats and to train responses to them. To do that, cyber ranges or testbeds are used. These are usually built to provide secure virtual environments where cybersecurity process can be monitored, studied, and analyzed without the risk of threatening operational infrastructure or where users can learn how to defend their systems against threats and attacks.

Cybersecurity platforms can be divided into three basic categories, each reflecting specific purposes of the cybersecurity domain: generic testbeds, lightweight platforms for cybersecurity training, and cyber ranges.

Generic testbeds provide a basic functionality for the emulation of computer networks. *Emulab/Netbed* [77] has been developed since 2000 and can be considered as a prototype of an emulation testbed for research into networking and distributed systems. It allocates computing resources for a specified network and instantiates the network at a dedicated hardware infrastructure. It provides accurate, repeatable results in experiments with moderate network loads [69, 59]. Another representative of this category, *CyberVAN* [2], is an experimentation testbed with hybrid emulation providing the ability to dynamically reconfigure the simulated network and the host nodes. It is able to simulate large strategic networks approximating a large ISP networks and employs Big Data Analytics engines and techniques for post-mortem analysis.

Lightweight platforms have been developed primarily for cybersecurity training. While some of them evolved from generic testbeds, others were designed from scratch with different needs in mind. *Avatao* [12, 1] is a web-based online e-learning platform offering IT security challenges (hands-on exercises), which can be organized to a path which leads to fulfilling an ultimate learning objective. In the *Hacking-Lab* [67] online platform, teams of participants have to perform several tasks simultaneously. Many lightweight platforms [14, 72, 61] focus on capture-the-flag games, which are similar to multi-level computer games where participants perform cyber-security tasks prescribed by individual levels, e. g. scan the network, find a vulnerable server, overtake the server.

Cyber ranges are complex virtual environments that are used not only for cyber warfare

training but also for cybertechnology development, forensic analysis, and other cyber-related issues. One very popular cyber range is *DETER/DeterLab* [51, 7], which is based on Emulab and was started with the goal of advancing cybersecurity research and education in 2004. There are currently many other cyber ranges, e. g., *National Cyber Range (NCR)* [23, 55], *Michigan Cyber Range (MCR)* [48], *SimSpace Cyber Range* [63], and *EDURange* [3].

A comprehensive survey of state-of-the-art cyber ranges and testbeds [17] published by the Australian Department of Defence in 2013 shows that our research started at the time when the concept of generic cyber ranges was in its initial stages. That was the reason we decided to develop our research platform, which we named *KYPO Cyber Range*. This cyber range is based on several key principles that affected its software architecture and design decisions, as briefly discussed in what follows.

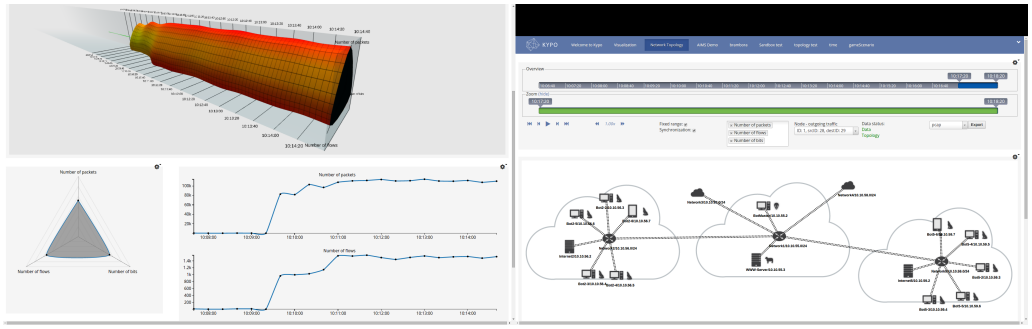

Figure 3.1: Dashboard of the KYPO Cyber Range.

Flexibility of Network Management – computer networks are fully virtualized in a cloud. For the topology nodes, a wide range of operating systems is supported (including arbitrary software packages). Network connections are emulated. Cloud-based virtualization brings the possibility to instantiate networks on demand, clone them, align their parameters and other dynamic aspects of networks management.

Isolation – network topologies and platform users can be isolated from the outside world and each other so that experiments and cyber-related activities cannot threaten other users or infrastructures.

Interoperability – in contrast to isolation, integration with (or connection to) external systems is also achievable with reasonable effort. For example, it is possible to connect an existing physical computer to the virtualized computer network.

Build-In Monitoring and Data Gathering – the platform natively provides both real-time and post-mortem access to detailed monitoring data. These data are related to individual topologies, including flow data and captured packets from the network links, as well as node metrics and logs. The monitoring subsystem is flexible, enabling us to gather heterogeneous data and adapt the monitoring to specific requirements of cyber scenarios or analytical tasks.

Easy Access – users with a wide range of experience should be able to use the platform. For less experienced users, web-based access to its core functions is available. Expert users, on the other hand, can interact with the platform via advanced means, e. g., using remote SSH access.

Providing Insight and Analytical Tools – the primary goal of cyber ranges is to support users in gaining insight into complex cybersecurity processes. Therefore, the KYPO Cyber Range puts great emphasis on providing exploratory visualizations and user interfaces that would be able to mediate the semantics of cybersecurity data to users, support situational awareness of developments in the cyber range, and support analytical tasks.

Contributions

My colleagues and I designed and implemented a generic cyber range with unique features that enable the use of the infrastructure for a wide variety of cybersecurity tasks, including training and forensic analysis. The KYPO Cyber Range is currently used as a training environment for a regular hands-on cybersecurity course at the Faculty of Informatics MU. It is also used for Cyber Czech, a big two-day event of cyber defense exercises where security professionals can improve their skills, and for many specialized demo presentations and exercises, including a "junior university" intended for children. In 2015, the KYPO Cyber Range received an award for exceptional results in the field of security research from the Ministry of Interior of the Czech Republic.

The architecture of KYPO and design decisions made during its development are summarized in [74]. Non-trivial engineering work resulted in a highly modular platform composed of five primary components that enable us to (a) support multiple cloud providers, (b) monitor and gather scenario-specific heterogeneous data, (c) design user interfaces that are adaptable to the specific requirements of particular scenarios, and (d) organize a wide variety of cost-effective, yet complex cybersecurity events that can vary in requirements on access restrictions, teamwork collaboration, and network properties.

While developing KYPO, we paid great attention to flexible, user-friendly interfaces providing visual interaction with the system. In [22], we described the requirements, principles, and design of mutually connected visualizations supporting several collaboration modes and providing users with exploratory visual feedback.

Papers in Collection

[22] Z. Eichler, R. Ošlejšek, and D. Toth. Kypo: A tool for collaborative study of cyberattacks in safe cloud environment. In *HCI International 2015: Human Aspects of Information Security, Privacy, and Trust*, pages 190–199, Los Angeles, 2015. Springer International Publishing.

I was responsible for the coordination of design and implementation activities of the interactive visual platform. I supervised a team of developers and participated in writing the paper. Contribution 35%.

[74] J. Vykopal, R. Ošlejšek, P. Čeleda, M. Vizváry, and D. Tovarňák. Kypo cyber range: Design and use cases. In *Proceedings of the 12th International Conference on Software Technologies - Volume 1: ICSOFT*, pages 310–321, Madrid, Spain, 2017. SciTePress.

I coordinated the design of the system architecture. I contributed to the data monitoring and management components and was responsible for the design and development of user interfaces and interactions. I wrote corresponding parts of the paper. Contribution 20%.

3.2 Analytical Visualizations for Cyber Exercises

In recent years, there has been a significant increase in hands-on competitions, challenges, and exercises [16, 54] in the cybersecurity domain. It is believed that they enable participants to effectively gain or practice diverse security skills in a fun way. Although the KYPO Cyber Range supports a wide variety of cybersecurity tasks, its current primary utilization is for cybersecurity education and training.

The most popular types of educational training are Capture The Flag (CTF) games [16] and Cyber Defence eXercises (CDX) [54]. While CTF games represent structured, step-by-step, hands-on training guided by an instructor and focusing on attacking, defending, or both, CDXs are usually intensive events lasting several days that solely train in defense and that attempt to mimic sophisticated attacks under real conditions.

To organize high-quality exercises of any type, it is necessary to provide meaningful insight into cyber processes during the training and to thoroughly analyze the exercises after their completion so that it is possible to assess the impact on participants and potentially improve future runs.

Our goal is to utilize exploratory visualization techniques to gain insight and provide exercise analyses. To do that, it is necessary to overcome several obstacles:

- *Formalization of organizational processes.* Complex exercises like CDXs are based on complex long-term processes involving a lot of user roles that have different (sometimes contradictory) goals and requirements. Knowing their goals, requirements, and expectations are necessary for the successful design of supporting visualization tools.
- *Clarification of learning objectives.* Learning objectives form expectations that organizers put on exercises and influence hypotheses that are to be evaluated by analytical tools.
- *Data classification.* Only data that are measurable in the cyber range can be used for runtime situational awareness and ex-post analysis.

- *Design of exploratory visualizations and analytical tools.* When data, processes, and objectives are clarified, then it is possible to design interactive visualizations providing insight into these aspects.

Few public research papers have dealt with the design of an exercise in a cyber range. Granåsen and Andersson conducted a case study on measuring team effectiveness in Baltic Cyber Shield 2010, a multi-national civil-military CDX [28]. The Spanish National Cybersecurity Institute proposed a taxonomy of cyber exercises [20] that recognizes operations-based exercises focused on incident response by participants in technical and management roles. The ISO/TC 223 effort resulted into ISO 22398, which describes general guidelines for exercises, including basic terms and definitions [32]. Unfortunately, the implementation details of an exercise in a cyber range are beyond the scope of this standard.

Contributions

Our largest contribution to cybersecurity education and training is the regular organization of CTF games and CDXs. To organize these events in the KYPO Cyber Range, it was necessary to analyze and formalize organizational processes, clarify learning objectives, classify data, and design user interfaces. Lessons learned from the organization of Cyber Czech, the biggest hands-on exercise in the Czech Republic, which has been organized eight times so far, were summarized and published in [75].

In spite of the progress in the formalization and classification of cybersecurity data and processes, big events like Cyber Czech are still organized with a lot of manual effort and with no support or only ad-hoc support for analytical tasks. There is no current systematic support for gaining insight and performing analyses directly in the cyber range. In [57], we discuss using a formal visual analytics model in the organization of CDXs. Using our approach, individual users participating in CDX organization could be systematically supported in their analytical and surveillance activities and, moreover, they could continuously build a knowledge base that could be shared across organizers in time.

In [73], we investigated how to provide valuable feedback to learners right after a CDX. Based on a scoring system integrated into the cyber range, we have developed a new feedback tool that presents an interactive, personalized timeline of exercise events and helps participants to learn from their experience gained during the exercise. To the best of our knowledge, this was the first paper attempting to study the means of providing visual feedback to learners participating in cyber defense exercises.

Exploratory visualizations for CTF games are also a primary interest for us. We have designed and developed several monitoring views gaining insight into the state of individual players, feedback visualizations for players, and analytical visualizations for game designers enabling them to evaluate the game difficulty and adjust its parameters to target types of

participants. These visualizations are currently under evaluation and prepared for publication.

Papers in Collection

[75] J. Vykopal, M. Vizváry, R. Ošlejšek, P. Čeleda, and D. Tovarňák. Lessons learned from complex hands-on defence exercises in a cyber range. In *2017 IEEE Frontiers in Education Conference*, pages 1–8, Indianapolis, IN, USA, 2017. IEEE.

I was responsible for the design of visualizations for cyber defense exercises and wrote several parts of the paper. Contribution 20%.

[57] R. Ošlejšek, J. Vykopal, K. Burská, and V. Rusňák. Evaluation of cyber defense exercises using visual analytics process. In *2018 IEEE Frontiers in Education Conference*. IEEE, 2018. To appear.

I was the author of the idea, I wrote several sections of the paper and supervised the preparation. Contribution 55%.

[73] J. Vykopal, R. Ošlejšek, K. Burská, and K. Zákopčanová. Timely feedback in unstructured cybersecurity exercises. In *Proceedings of Special Interest Group on Computer Science Education, Baltimore, Maryland, USA, February 21–24, 2018(SIGCSE'18)*, pages 173–178, Baltimore, Maryland, USA, 2018. ACM.

I participated in the design of visualization and was responsible for the evaluation of results. I wrote several sections of the paper. Contribution 30%.

Chapter 4

Conclusion

In this text, I have presented my research contributions to the progress within the area of insight into unstructured data via exploratory techniques. The individual research contributions were accompanied with selected representative articles I co-authored, which are also attached to this text [1].

In the future, we would like to continue elaborating techniques for visual analysis in cyber exercises. To increase the impact of the training courses, providing timely visual feedback and meaningful interpretation of cybersecurity data and processes is crucial. Moreover, cybersecurity experts are looking for new interactive techniques that would enable them to perform forensic analysis efficiently and to protect critical infrastructures against cyber threats. This challenging and still rather unexplored area presents an application domain with high potential, and our generic KYPO Cyber Range provides us with a great opportunity for this kind of research.

[1] The full texts of the articles are excluded from the public version of this text to avoid copyright violation.

Bibliography

[1] Avatao. `https://avatao.com`. Accessed: 2017-05-22.

[2] Cyber virtual ad hoc network (CyberVAN). `http://www.appcomsci.com/research/tools/cybervan`. Accessed: 2017-05-22.

[3] EDURange. `http://www.edurange.org`. Accessed: 2017-05-22.

[4] R. Abbasi, S. Chernov, W. Nejdl, R. Paiu, and S. Staab. Exploiting flickr tags and groups for finding landmark photos. In *European Conference on Information Retrieval*, pages 654–661. Springer, 2009.

[5] M. Araki. Owl-based frame descriptions for spoken dialog systems. In *Proceedings of International Workshop on Semantic Web Foundations and Application Technologies, Nara Prefecture Public Hall, Nara, Japan*. Citeseer, 2003.

[6] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6):1450–1464, 2002.

[7] T. Benzel. The science of cyber security experimentation: The DETER project. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pages 137–148. ACM, 2011.

[8] T. L. Bonebright, M. A. Ness, T. T. Connerley, and G. R. MaCain. Testing the effectiveness of sonified graphs for education: A programmatic research project. In *Proc Int Conf on Auditory Displays*, Espon, Finland, 2001.

[9] M. Boutell and J. Luo. Photo classification by integrating image content and camera metadata. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 901–904. IEEE, 2004.

[10] A. Brown, R. Stevens, and S. Pettifer. Making graph-based diagrams work in sound: the role of annotation. *Human–Computer Interaction*, 28(3):193–221, 2013.

[11] Burning Glass Tech. Job market intelligence: Cybersecurity jobs. `http://burning-glass.com/wp-content/uploads/Cybersecurity_Jobs_Report_2015.pdf`, 2015. Accessed: 2018-08-08.

[12] L. Buttyán, M. Félegyházi, and G. Pék. Mentoring talent in IT security–A case study. In *2016 USENIX Workshop on Advances in Security Education (ASE 16)*, 2016.

[13] Cisco Systems. Cisco 2014 annual security report. `http://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2014_ASR.pdf`, 2014. Accessed: 2018-08-08.

[14] CTF365. Capture the flag 365. `https://ctf365.com`. Accessed: 2017-05-22.

[15] G. Daunys and V. Lauruska. Maps sonification system using digitiser for visually impaired children. In *Int. Conf. ICCHP*, pages 12–15. Berlin : Springer-Verlag, July 2006.

[16] A. Davis, T. Leek, M. Zhivich, K. Gwinnup, and W. Leonard. The fun and future of ctf. In *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, San Diego, CA, 2014. USENIX Association.

[17] J. Davis and S. Magrath. A survey of cyber ranges and testbeds. Technical report, DTIC Document, 2013.

[18] A. Diehl, L. Pelorosso, C. Delrieux, C. Saulo, J. Ruiz, M. E. Gröller, and S. Bruckner. Visual analysis of spatio-temporal data: Applications in weather forecasting. *Computer Graphics Forum*, 34(3):381–390, May 2015.

[19] A. Diehl, L. Pelorosso, K. Matkovic, J. Ruiz, M. E. Gröller, and S. Bruckner. Albero: A visual analytics approach for probabilistic weather forecasting. *Computer Graphics Forum*, 36(7):135–144, Oct. 2017.

[20] E. G. Díez, D. F. Pereira, M. A. L. Merino, H. R. Suárez, and D. B. Juan. Cyber exercises taxonomy. Technical report, INCIBE, 2015.

[21] P. K. Edman. *Tactile Graphics*. American Foundation for the Blind, New York, 1992.

[22] Z. Eichler, R. Ošlejšek, and D. Toth. Kypo: A tool for collaborative study of cyberattacks in safe cloud environment. In *HCI International 2015: Human Aspects of Information Security, Privacy, and Trust*, pages 190–199, Los Angeles, 2015. Springer International Publishing.

[23] B. Ferguson, A. Tall, and D. Olsen. National cyber range overview. In *2014 IEEE Military Communications Conference*, pages 123–128, Oct 2014.

[24] A. Flycht-Eriksson. *Design and use of ontologies in information-providing dialogue systems PDF*. PhD thesis, Linköping University, 2004.

[25] K. M. Franklin and J. C. Roberts. Pie chart sonification. In *Proceedings of Information Visualization (IV03)*, pages 4–9, London, UK, July 2003. IEEE Computer Press.

[26] K. M. Franklin and J. C. Roberts. A path based model for sonification. In *Proceedings of Information Visualization (IV04)*, pages 865–870. IEEE Computer Society, 2004.

[27] Z. B. Fredj and D. Duce. Grassml: Accessible smart schematic diagrams for all. In *Theory and Practice of Computer Graphics*, pages 49–55. IEEE, June 2003.

[28] M. Granåsen and D. Andersson. Measuring team effectiveness in cyber-defense exercises: a cross-disciplinary case study. *Cognition, Technology & Work*, 18(1):121–143, Feb 2016.

[29] J. Haddadnia and M. Ahmadi. N-feature neural network human face recognition. *Image and Vision Computing*, 22(12):1071–1082, 2004.

[30] P. Hamřík, I. Kopeček, R. Ošlejšek, and J. Plhák. Dialogue-based information retrieval from images. In *Computers Helping People with Special Needs:14th International Conference, ICCHP 2014*, pages 85–92, Switzerland, 2014. Springer International Publishing.

[31] E. Hovy, U. Hermjakob, and D. Ravichandran. A question/answer typology with surface text patterns. In *Proceedings of the second international conference on Human Language Technology Research*, pages 247–251. Morgan Kaufmann Publishers Inc., 2002.

[32] Societal security – guidelines for exercises. Standard, International Organization for Standardization, Geneva, CH, Sept. 2013.

[33] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries and visualization for large collections of geo-referenced photographs. In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*, pages 89–98. ACM, 2006.

[34] W. D. Jones. Sight for sore ears. *IEEE Spectrum*, pages 11–12, 2004. http://www.seeingwithsound.com/javoice.htm.

[35] K. Kaczmarek. Electrotactile display for computer graphics to blind. Research Report 5-R01-EY10019-08, University of Wisconsin, 2004.

[36] H. M. Kamel and J. A. Landay. Sketching images eyes-free: a grid-based dynamic drawing tool for the blind. In *Proceedings of the fifth international ACM conference on Assistive technologies*, pages 33–40. ACM Press, 2002.

[37] I. Kopeček and R. Ošlejšek. Creating pictures by dialogue. In *Computers Helping People with Special Needs: 10th International Conference, ICCHP 2006*, pages 61–68, Berlin, 2006. Springer-Verlag.

[38] I. Kopeček and R. Ošlejšek. Gate to accessibility of computer graphics. In *Computers Helping People with Special Needs: 11th International Conference, ICCHP 2008*, pages 295–302, Berlin, 2008. Springer-Verlag.

[39] I. Kopeček and R. Ošlejšek. Hybrid approach to sonification of color images. In *The 2008 International Conference on Convergence and Hybrid Information Technologies*, volume 2, pages 722–727. IEEE Computer Society, 2008.

[40] I. Kopeček and R. Ošlejšek. Annotating and describing pictures – applications in e-learning and accessibility of graphics. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010*, pages 124–130, Berlin, 2010. Springer-Verlag.

[41] I. Kopeček, R. Ošlejšek, and J. Plhák. Integrating dialogue systems with images. In *Text, Speech and Dialogue. 15th International Conference, TSD 2012*, pages 632–639, Berlin Heidelber, 2012. Springer-Verlag.

[42] B. Kozlíková, M. Krone, M. Falk, N. Lindow, M. Baaden, D. Baum, I. Viola, J. Parulek, and H.-C. Hege. Visualization of biomolecular structures: State of the art revisited. *Computer Graphics Forum*, n/a(n/a):n/a–n/a, 2016.

[43] M. Krone, B. Kozlikova, N. Lindow, M. Baaden, D. Baum, J. Parulek, H.-C. Hege, and I. Viola. Visual analysis of biomolecular cavities: State of the art. *Computer Graphics Forum*, 2016.

[44] M. Kurze. Tdraw: a computer-based tactile drawing tool for blind people. In *Proceedings of the second annual ACM conf. on Assistive technologies*, pages 131–138, 1996.

[45] Mandiant Corp. Exposing one of China's cyber espionage units – Mandiant APT1 report. `http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf`, 2013. Accessed: 2018-08-08.

[46] R. M. Mathis. Constraint scalable vector graphics, accessibility and the semantic web. In *SoutheastCon Proceedings*, pages 588–593. IEEE Computer Society, 2005.

[47] S. Matta, H. Rudolph, and D. K. Kumar. Auditory eyes: Representing visual information in sound and tactile cues. In *13th European Signal Processing Conf.* Antalya, 2005.

[48] MCR. The michigan cyber range. `https://www.merit.edu/cyberrange/`. Accessed: 2017-05-22.

[49] D. Milward and M. Beveridge. Ontology-based dialogue systems. In *Proc. 3rd Workshop on Knowledge and reasoning in practical dialogue systems (IJCAI03)*, pages 9–18, 2003.

[50] M. Minsky. A framework for representing knowledge. *The Psychology of Computer Vision*, 1974.

[51] J. Mirkovic, T. V. Benzel, T. Faber, R. Braden, J. T. Wroclawski, and S. Schwab. The DETER Project: Advancing the Science of Cyber Security Experimentation and Test. In *Proceedings of the 2010 IEEE International Conference on Technologies for Homeland Security (HST '10)*, Waltham, Massachusetts, Nov. 2010.

[52] T. Murillo-Morales and K. Miesenberger. Ontology-based semantic support to improve accessibility of graphics. *Studies in health technology and informatics*, 217:255–260, 2015.

[53] T. Murillo-Morales, K. Miesenberger, and J. Plhák. Authoring semantic annotations for non-visual access to graphics. *Journal on Technology & Persons with Disabilities*, 6, 2018.

[54] NATO cooperative cyber defence centre of excellence. Locked shields. `http://ccdcoe.org/event/cyber-defence-exercises.html`. Accessed: 2017-05-22.

[55] NCR. The national cyber range. `http://www.acq.osd.mil/dte-trmc/docs/Docs/NCR/2015_NCR%20Info%20Sheet_Updated.pdf`. Accessed: 2017-05-22.

[56] S. Nyrkko, L. Carlson, M. Keijola, H. Ahonen-Myka, J. Niemi, J. Piitulainen, S. Viitanen, M. Meri, L. Seitsonen, P. Mannonen, et al. Ontology-based knowledge in interactive maintenance guide. In *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on*, pages 47–47. IEEE, 2007.

[57] R. Ošlejšek, J. Vykopal, K. Burská, and V. Rusňák. Evaluation of cyber defense exercises using visual analytics process. In *2018 IEEE Frontiers in Education Conference*. IEEE, 2018. To appear.

[58] G. Pérez, G. Amores, P. Manchón, F. Gómez, and J. González. Integrating owl ontologies with a dialogue manager. *Procesamiento del Lenguaje Natural*, 2006.

[59] A. Perez-Garcia, C. Siaterlis, and M. Masera. Designing repeatable experiments on an emulab testbed. In *International Conference on Broadband Communications, Networks and Systems*, pages 28–39. Springer, 2010.

[60] A. Pras, A. Sperotto, G. Moura, I. Drago, R. Barbosa, R. Sadre, R. Schmidt, and R. Hofstede. Attacks by "anonymous" wikileaks proponents not anonymous. Technical report, University of Twente, Centre for Telematics and Information Technology (CTIT), 2010.

[61] A. S. Raj, B. Alangot, S. Prabhu, and K. Achuthan. Scalable and lightweight ctf infrastructures using application containers. In *2016 USENIX Workshop on Advances in Security Education (ASE 16)*, Austin, TX, Aug. 2016. USENIX Association.

[62] D. Ravichandran and E. Hovy. Learning surface text patterns for a question answering system. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 41–47. Association for Computational Linguistics, 2002.

[63] L. Rossey. SimSpace cyber range. `https://www.acsac.org/2015/program/ACSAC%202015%20CEF%20Panel%20-%20Rossey.pdf`. ACSAC 2015 Panel: Cyber Experimentation of the Future (CEF): Catalyzing a New Generation of Experimental Cybersecurity Research.

[64] F. E. Sandnes. Where was that photo taken? deriving geographical information from image collections based on temporal exposure attributes. *Multimedia Systems*, 16(4-5):309–318, 2010.

[65] I. Satoshi. Computer graphics for the blind. *ACM SIGCAPH Newsletter Page*, 55:16–21, 1996.

[66] A. T. G. Schreiber, B. Dubbeldam, J. Wielemaker, and B. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16(3):66–74, May 2001.

[67] Security Competence. Hacking-lab. `http://www.hacking-lab-ctf.com/technical.html`. Accessed: 2017-05-22.

[68] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[69] C. Siaterlis, A. P. Garcia, and B. Genge. On the Use of Emulab Testbeds for Scientifically Rigorous Experiments. *IEEE Communications Surveys Tutorials*, 15(2):929–942, Second 2013.

[70] C. Vaitsis, G. Nilsson, and N. Zary. Big data in medical informatics: improving education through visual analytics. In *MIE*, pages 1163–1167, 2014.

[71] R. Vatrapu, C. Teplovs, N. Fujita, and S. Bull. Towards visual analytics for teachers' dynamic diagnostic pedagogical decision-making. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, pages 93–98. ACM, 2011.

[72] G. Vigna, K. Borgolte, J. Corbetta, A. Doupe, Y. Fratantonio, L. Invernizzi, D. Kirat, and Y. Shoshitaishvili. Ten years of iCTF: The good, the bad, and the ugly. In *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*, 2014.

[73] J. Vykopal, R. Ošlejšek, K. Burská, and K. Zákopčanová. Timely feedback in unstructured cybersecurity exercises. In *Proceedings of Special Interest Group on Computer Science Education, Baltimore, Maryland, USA, February 21–24, 2018(SIGCSE'18)*, pages 173–178, Baltimore, Maryland, USA, 2018. ACM.

[74] J. Vykopal, R. Ošlejšek, P. Čeleda, M. Vizváry, and D. Tovarňák. Kypo cyber range: Design and use cases. In *Proceedings of the 12th International Conference on Software Technologies - Volume 1: ICSOFT*, pages 310–321, Madrid, Spain, 2017. SciTePress.

[75] J. Vykopal, M. Vizváry, R. Ošlejšek, P. Čeleda, and D. Tovarňák. Lessons learned from complex hands-on defence exercises in a cyber range. In *2017 IEEE Frontiers in Education Conference*, pages 1–8, Indianapolis, IN, USA, 2017. IEEE.

[76] X. H. Wang, D. Q. Zhang, T. Gu, and H. K. Pung. Ontology based context modeling and reasoning using owl. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 18–22. Ieee, 2004.

[77] B. White, J. Lepreau, L. Stoller, R. Ricci, S. Guruprasad, M. Newbold, M. Hibler, C. Barb, and A. Joglekar. An integrated experimental environment for distributed systems and networks. *SIGOPS Oper. Syst. Rev.*, 36(SI):255–270, Dec. 2002.

[78] P. C. Wong and J. Thomas. Guest editors' introduction–visual analytics. *IEEE Computer Graphics and Applications, 24 (5): 20-21*, 24(PNNL-SA-41935), 2004.

[79] J. Yuan, J. Luo, and Y. Wu. Mining compositional features from gps and visual cues for event recognition in photo collections. *IEEE Transactions on Multimedia*, 12(7):705–716, 2010.

Part II

Collection of Selected Publications

Appendix A

List of Publications

This appendix together with Appendix B contains the total of 11 research papers that were selected as the representatives of my contributions within the studied research field. The full texts of the papers are inserted into the corresponding appendixes of the printed version of this thesis[1] and referenced via the paper numbers assigned in the list below (replacing page numbers). The same holds for Appendix B. Papers are marked with CORE ranking[2] and publication type.

Article A (Springer | Book Chapter): I. Kopeček and R. Ošlejšek. Creating pictures by dialogue. In *Computers Helping People with Special Needs: 10th International Conference, ICCHP 2006*, pages 61–68, Berlin, 2006. Springer-Verlag

Article B (Springer | Book Chapter): I. Kopeček and R. Ošlejšek. Gate to accessibility of computer graphics. In *Computers Helping People with Special Needs: 11th International Conference, ICCHP 2008*, pages 295–302, Berlin, 2008. Springer-Verlag

Article C (Springer | Book Chapter): P. Hamřík, I. Kopeček, R. Ošlejšek, and J. Plhák. Dialogue-based information retrieval from images. In *Computers Helping People with Special Needs:14th International Conference, ICCHP 2014*, pages 85–92, Switzerland, 2014. Springer International Publishing

Article D (IEEE | Conference): I. Kopeček and R. Ošlejšek. Hybrid approach to sonification of color images. In *The 2008 International Conference on Convergence and Hybrid Information Technologies*, volume 2, pages 722–727. IEEE Computer Society, 2008

[1] The full texts of the articles are excluded from the publicly available electronic version of this text to avoid copyright violation.

[2] http://portal.core.edu.au/conf-ranks/

Article E (Springer | Book Chapter): I. Kopeček and R. Ošlejšek. Annotating and describing pictures – applications in e-learning and accessibility of graphics. In *Computers Helping People with Special Needs: 12th International Conference, ICCHP 2010*, pages 124–130, Berlin, 2010. Springer-Verlag

Article F (Springer | Book Chapter): I. Kopeček, R. Ošlejšek, and J. Plhák. Integrating dialogue systems with images. In *Text, Speech and Dialogue. 15th International Conference, TSD 2012*, pages 632–639, Berlin Heidelber, 2012. Springer-Verlag

Article G (Springer | Book Chapter): Z. Eichler, R. Ošlejšek, and D. Toth. Kypo: A tool for collaborative study of cyberattacks in safe cloud environment. In *HCI International 2015: Human Aspects of Information Security, Privacy, and Trust*, pages 190–199, Los Angeles, 2015. Springer International Publishing

Article H (CORE B | Conference): J. Vykopal, R. Ošlejšek, P. Čeleda, M. Vizváry, and D. Tovarňák. Kypo cyber range: Design and use cases. In *Proceedings of the 12th International Conference on Software Technologies - Volume 1: ICSOFT*, pages 310–321, Madrid, Spain, 2017. SciTePress

Article I (CORE B | Conference): J. Vykopal, M. Vizváry, R. Ošlejšek, P. Čeleda, and D. Tovarňák. Lessons learned from complex hands-on defence exercises in a cyber range. In *2017 IEEE Frontiers in Education Conference*, pages 1–8, Indianapolis, IN, USA, 2017. IEEE

Article J (CORE A | Conference): J. Vykopal, R. Ošlejšek, K. Burská, and K. Zákopčanová. Timely feedback in unstructured cybersecurity exercises. In *Proceedings of Special Interest Group on Computer Science Education, Baltimore, Maryland, USA, February 21–24, 2018(SIGCSE'18)*, pages 173–178, Baltimore, Maryland, USA, 2018. ACM

Article K (CORE B | Conference): R. Ošlejšek, J. Vykopal, K. Burská, and V. Rusňák. Evaluation of cyber defense exercises using visual analytics process. In *2018 IEEE Frontiers in Education Conference*. IEEE, 2018. To appear

Appendix A. List of Publications

Appendix B

Collection of Articles

Paper A

Creating Pictures by Dialogue

Ivan Kopeček, Radek Ošlejšek

Masaryk University, Faculty of Informatics, Brno, Czech Republic

ICCHP – 10th International Conference on Computers Helping People with Special Needs, Springer LNCS, volume 4061, 2006, p. 61-68, 8 pp. https://doi.org/10.1007/11788713_10

Abstract

This paper deals with the problem of generating a picture of a scene by means of describing it in the terms of natural language ontologies. The corresponding graphical format, being formally represented in the form of the Pawlak information system, involves a full semantic description of the picture. Therefore, the pictures generated in this way "blind-friendly", i.e. they can be automatically fully described. A simple example of such a picture generation is presented here.

Paper B

GATE to Accessibility of Computer Graphics

Ivan Kopeček, Radek Ošlejšek

Masaryk University, Faculty of Informatics, Brno, Czech Republic

ICCHP – 11th International Conference on Computers Helping People with Special Needs, Springer LNCS, volume 5105, 2008, p. 295-302, 8 pp. https://doi.org/10.1007/978-3-540-70540-6_44

Abstract

This paper presents a framework for integrating current information technologies into a platform enabling the blind and visually impaired people to access computer graphics based on the annotated SVG format. We also present a technique enabling the conversion of any graphical object to the annotated SVG format and easy annotation supported by OWL based ontology. This approach is not limited to vector graphics only, but enables also the flexible annotation and application of raster graphics (e.g. photographs). We briefly describe the architecture of the GATE (Graphics Accessible To Everyone) project, which contains the corresponding implemented modules. As an illustration, we provide an example showing how the blind can access the annotated graphics.

Paper C

Dialogue-based Information Retrieval from Images

Pavel Hamřík, Ivan Kopeček, Radek Ošlejšek, Jaromír Plhák

Masaryk University, Faculty of Informatics, Brno, Czech Republic

ICCHP – 14th International Conference on Computers Helping People with Special Needs, Springer LNCS, volume 8547, 2014, p. 85-92, 8 pp. https://doi.org/10.1007/978-3-319-08596-8_13

Abstract

Our concept of communicative images aims to provide graphical information by means of dialogue interaction, which is suitable for people with various disabilities. Communicative images are graphical objects integrated with a dialogue interface and linked to an associated knowledge database which stores the semantics of the objects depicted. This paper deals with the utilization of formal ontologies for the process of image annotation and dialogue-based investigation in the context of assistive technologies.

Paper D

Hybrid Approach to Sonification of Color Images

Ivan Kopeček, Radek Ošlejšek

Masaryk University, Faculty of Informatics, Brno, Czech Republic

ICCIT – 3rd International Conference Convergence and Hybrid Information Technology, IEEE Computer Society, 2008, p. 722-727, 6 pp. https://doi.org/10.1109/ICCIT.2008.152

Abstract

This paper deals with the accessibility of graphics for visually impaired people. It presents a novel method of sonification of complex graphical objects, such as color photographs, based on a hybrid approach combining sound and speech communication. This approach is supported by a special color model, called semantic color model, which is introduced in the paper. The semantic color model possesses suitable properties that can be used to deliver the relevant graphical information in sound or speech in a convenient form. The integration of this approach with the annotated SVG format developed within the ongoing GATE project, which is also briefly described in the paper, enhances the efficiency of the system.

Paper E

Annotating and Describing Pictures – Applications in E-learning and Accessibility of Graphics

Ivan Kopeček, Radek Ošlejšek

Masaryk University, Faculty of Informatics, Brno, Czech Republic

ICCHP – 12th International Conference on Computers Helping People with Special Needs, Springer LNCS, volume 6179, 2010, p. 124-130, 7 pp. https://doi.org/10.1007/978-3-642-14097-6_21

Abstract

The paper describes the ontology based approach to the annotation of graphical objects in relation to the accessibility of graphics. Some applications in e-leaning are also discussed. The problem concerning the optimality of graphical ontologies in relation with the annotation process is addressed and an optimality measure enabling algorithmic solution of this problem is proposed. Finally, an approach to generating picture description is presented.

Paper F

Integrating Dialogue Systems with Images

Ivan Kopeček, Radek Ošlejšek, Jaromír Plhák

Masaryk University, Faculty of Informatics, Brno, Czech Republic

TSD – 12th International Conference on Text, Speech and Dialogue, Springer LNCS, volume 7499, 2012, p. 632-639, 8 pp. https://doi.org/10.1007/978-3-642-32790-2_77

Abstract

The paper presents a novel approach, in which images are integrated with a dialogue interface that enables them to communicate with the user. The structure of the corresponding dialogue system is supported by graphical ontologies and enables the system learning from the dialogues. The Internet environment is used for retrieving additional information about the images as well as for solving more complex tasks related with exploiting other relevant knowledge. Further, the paper deals with some problems that arise from the system initiative dialogue mode and discusses the structure and algorithms of the dialogue system. Some examples and applications of the presented approach are presented as well.

Paper G

KYPO: A Tool for Collaborative Study of Cyberattacks in Safe Cloud Environment

Zdenek Eichler, Radek Ošlejšek, Dalibor Toth

Masaryk University, Faculty of Informatics, Brno, Czech Republic

HCI International 2015 – Human Aspects of Information Security, Privacy, and Trust. Springer LNCS, volume 9190, 2015, p. 190-199, 10 pp. https://doi.org/10.1007/978-3-319-20376-8_17

Abstract

This paper introduces the KYPO – a cloud-based virtual environment faithfully simulating real networks and enabling users to study cyber attacks as well as to train users in isolated and controlled environment. Particularly, the paper focuses on the user environment and visualizations, providing views and interactions improving the understanding of processes emerged during experiments. Web user interface of the KYPO system supports several collaboration modes enabling the participants to experiment and replay different types of security related tasks.

Paper H

KYPO Cyber Range: Design and Use Cases

Jan Vykopal[1], Radek Ošlejšek[2], Pavel Čeleda[1], Martin Vizváry[1], Daniel Tovarňák[1]

[1] Masaryk University, Institute of Computer Science, Brno, Czech Republic
[2] Masaryk University, Faculty of Informatics, Brno, Czech Republic

ICSOFT – 12th International Conference on Software Technologies. SciTePress, volume 1, 2017, p. 310-321, 12 pp. http://doi.org/10.5220/0006428203100321

Abstract

The physical and cyber worlds are increasingly intertwined and exposed to cyber attacks. The KYPO cyber range provides complex cyber systems and networks in a virtualized, fully controlled and monitored environment. Time-efficient and cost-effective deployment is feasible using cloud resources instead of a dedicated hardware infrastructure. This paper describes the design decisions made during it's development. We prepared a set of use cases to evaluate the proposed design decisions and to demonstrate the key features of the KYPO cyber range. It was especially cyber training sessions and exercises with hundreds of participants which provided invaluable feedback for KYPO platform development.

Paper I

Lessons Learned From Complex Hands-on Defence Exercises in a Cyber Range

Jan Vykopal[1], Martin Vizváry[1], Radek Ošlejšek[2], Pavel Čeleda[1], Daniel Tovarňák[1]

[1] Masaryk University, Institute of Computer Science, Brno, Czech Republic
[2] Masaryk University, Faculty of Informatics, Brno, Czech Republic

FIE – IEEE Frontiers in Education Conference. IEEE, 2017, p. 1-8, 8 pp. http://doi.org/10.1109/FIE.2017.8190713

Abstract

We need more skilled cybersecurity professionals because the number of cyber threats and ingenuity of attackers is ever growing. Knowledge and skills required for cyber defence can be developed and exercised by lectures and lab sessions, or by active learning, which is seen as a promising and attractive alternative. In this paper, we present experience gained from the preparation and execution of cyber defence exercises involving various participants in a cyber range. The exercises follow a Red vs. Blue team format, in which the Red team conducts malicious activities against emulated networks and systems that have to be defended by Blue teams of learners. Although this exercise format is popular and used worldwide by numerous organizers in practice, it has been sparsely researched. We contribute to the topic by describing the general exercise life cycle, covering the exercise's development, dry run, execution, evaluation, and repetition. Each phase brings several challenges that exercise organizers have to deal with. We present lessons learned that can help organizers to prepare, run and repeat successful events systematically, with lower effort and costs, and avoid a trial-and-error approach that is often used.

Paper J

Timely Feedback in Unstructured Cybersecurity Exercises

Jan Vykopal[1], Radek Ošlejšek[2], Karolína Burská[2], Kristína Zákopčanová[2]

[1] Masaryk University, Institute of Computer Science, Brno, Czech Republic
[2] Masaryk University, Faculty of Informatics, Brno, Czech Republic

SIGCSE – Proceedings of Special Interest Group on Computer Science Education. ACM, 2018, p. 173-178, 6 pp. http://doi.org/10.1145/3159450.3159561

Abstract

Cyber defence exercises are intensive, hands-on learning events for teams of professionals who gain or develop their skills to successfully prevent and respond to cyber attacks. The exercises mimic the real-life, routine operation of an organization which is being attacked by an unknown offender. Teams of learners receive very limited immediate feedback from the instructors during the exercise; they can usually see only a scoreboard showing the aggregated gain or loss of points for particular tasks. An in-depth analysis of learners' actions requires considerable human effort, which results in days or weeks of delay. The intensive experience is thus not followed by proper feedback facilitating actual learning, and this diminishes the effect of the exercise.

In this initial work, we investigate how to provide valuable feedback to learners right after the exercise without any unnecessary delay. Based on the scoring system of a cyber defence exercise, we have developed a new feedback tool that presents an interactive, personalized timeline of exercise events. We deployed this tool during an international exercise, where we monitored participants' interactions and gathered their reflections. The results show that learners did use the new tool and rated it positively. Since this new feature is not bound to a particular defence exercise, it can be applied to all exercises that employ scoring based

on the evaluation of individual exercise objectives. As a result, it enables the learner to immediately reflect on the experience gained.

Paper K

Evaluation of Cyber Defense Exercises Using Visual Analytics Process

Radek Ošlejšek[1], Jan Vykopal[2], Karolína Burská[1], Vít Rusňák[2]

[1] Masaryk University, Faculty of Informatics, Brno, Czech Republic
[2] Masaryk University, Institute of Computer Science, Brno, Czech Republic

FIE – IEEE Frontiers in Education Conference. IEEE, 2018, 9 pp. to appear

Abstract

This Innovative Practice Full Paper addresses modern cyber ranges which represent unified platforms that offer efficient organization of complex hands-on exercises where participants can train their cybersecurity skills. However, the functionality targets mostly learners who are the primary users. Support of organizers performing analytic and evaluation tasks is weak and ad-hoc. It makes harder to improve the quality of an exercise, particularly its impact on learners. In this paper, we present an application of a well-structured visual analytics process to the organization of cyber exercises. We illustrate that the classification derived from the adoption of the visual analytics process helps to clarify and formalize analytical tasks of educators and enables their systematic support in cyber ranges. We demonstrate an application of our approach on a particular series of eight exercises we have organized in last three years. We believe the presented approach is beneficial for anyone involved in preparation and execution of any complex exercise.