Zbornik 21. mednarodne multikonference

INFORMACIJSKA DRUŽBA - IS 2018

Zvezek E

Proceedings of the 21st International Multiconference

INFORMATION SOCIETY - IS 2018

Volume E

Delavnica AS-IT-IC
AS-IT-IC Workshop

Uredila / Edited by
Matjaž Gams, Jernej Zupančič

http://is.ijs.si

8.–12. oktober 2018 / 8–12 October 2018
Ljubljana, Slovenia

Zbornik 21. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2018

Zvezek E

Proceedings of the 21st International Multiconference

INFORMATION SOCIETY – IS 2018

Volume E

Delavnica AS-IT-IC
AS-IT-IC Workshop

Uredila / Edited by

Matjaž Gams, Jernej Zupančič

http://is.ijs.si

8.–12. oktober 2018 / 8–12 October 2018
Ljubljana, Slovenia

Urednika:

Matjaž Gams
Odsek za inteligentne sisteme
Institut »Jožef Stefan«, Ljubljana

Jernej Zupančič
Odsek za inteligentne sisteme
Institut »Jožef Stefan«, Ljubljana

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Lana Zemljak
Oblikovanje naslovnice: Vesna Lasič

Dostop do e-publikacije:
http://library.ijs.si/Stacks/Proceedings/InformationSociety

Ljubljana, oktober 2018

Informacijska družba
ISSN 2630-371X

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID=31893031
ISBN 978-961-264-139-9 (pdf)

PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2018

Multikonferenca Informacijska družba (http://is.ijs.si) je z enaindvajseto zaporedno prireditvijo osrednji srednjeevropski dogodek na področju informacijske družbe, računalništva in informatike. Letošnja prireditev se ponovno odvija na več lokacijah, osrednji dogodki pa so na Institutu »Jožef Stefan«.

Informacijska družba, znanje in umetna inteligenca so še naprej nosilni koncepti človeške civilizacije. Se bo neverjetna rast nadaljevala in nas ponesla v novo civilizacijsko obdobje ali pa se bo rast upočasnila in začela stagnirati? Bosta IKT in zlasti umetna inteligenca omogočila nadaljnji razcvet civilizacije ali pa bodo demografske, družbene, medčloveške in okoljske težave povzročile zadušitev rasti? Čedalje več pokazateljev kaže v oba ekstrema – da prehajamo v naslednje civilizacijsko obdobje, hkrati pa so notranji in zunanji konflikti sodobne družbe čedalje težje obvladljivi.

Letos smo v multikonferenco povezali 11 odličnih neodvisnih konferenc. Predstavljenih bo 215 predstavitev, povzetkov in referatov v okviru samostojnih konferenc in delavnic. Prireditev bodo spremljale okrogle mize in razprave ter posebni dogodki, kot je svečana podelitev nagrad. Izbrani prispevki bodo izšli tudi v posebni številki revije Informatica, ki se ponaša z 42-letno tradicijo odlične znanstvene revije.

Multikonferenco Informacijska družba 2018 sestavljajo naslednje samostojne konference:

- Slovenska konferenca o umetni inteligenci
- Kognitivna znanost
- Odkrivanje znanja in podatkovna skladišča – SiKDD
- Mednarodna konferenca o visokozmogljivi optimizaciji v industriji, HPOI
- Delavnica AS-IT-IC
- Soočanje z demografskimi izzivi
- Sodelovanje, programska oprema in storitve v informacijski družbi
- Delavnica za elektronsko in mobilno zdravje ter pametna mesta
- Vzgoja in izobraževanje v informacijski družbi
- 5. študentska računalniška konferenca
- Mednarodna konferenca o prenosu tehnologij (ITTC)

Soorganizatorji in podporniki konference so različne raziskovalne institucije in združenja, med njimi tudi ACM Slovenija, Slovensko društvo za umetno inteligenco (SLAIS), Slovensko društvo za kognitivne znanosti (DKZ) in druga slovenska nacionalna akademija, Inženirska akademija Slovenije (IAS). V imenu organizatorjev konference se zahvaljujemo združenjem in institucijam, še posebej pa udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

V letu 2018 bomo šestič podelili nagrado za življenjske dosežke v čast Donalda Michieja in Alana Turinga. Nagrado Michie-Turing za izjemen življenjski prispevek k razvoju in promociji informacijske družbe bo prejel prof. dr. Saša Divjak. Priznanje za dosežek leta bo pripadlo doc. dr. Marinki Žitnik. Že sedmič podeljujemo nagradi »informacijska limona« in »informacijska jagoda« za najbolj (ne)uspešne poteze v zvezi z informacijsko družbo. Limono letos prejme padanje državnih sredstev za raziskovalno dejavnost, jagodo pa Yaskawina tovarna robotov v Kočevju. Čestitke nagrajencem!

Mojca Ciglarič, predsednik programskega odbora

Matjaž Gams, predsednik organizacijskega odbora

FOREWORD - INFORMATION SOCIETY 2018

In its 21st year, the Information Society Multiconference (http://is.ijs.si) remains one of the leading conferences in Central Europe devoted to information society, computer science and informatics. In 2018, it is organized at various locations, with the main events taking place at the Jožef Stefan Institute.

Information society, knowledge and artificial intelligence continue to represent the central pillars of human civilization. Will the pace of progress of information society, knowledge and artificial intelligence continue, thus enabling unseen progress of human civilization, or will the progress stall and even stagnate? Will ICT and AI continue to foster human progress, or will the growth of human, demographic, social and environmental problems stall global progress? Both extremes seem to be playing out to a certain degree – we seem to be transitioning into the next civilization period, while the internal and external conflicts of the contemporary society seem to be on the rise.

The Multiconference runs in parallel sessions with 215 presentations of scientific papers at eleven conferences, many round tables, workshops and award ceremonies. Selected papers will be published in the Informatica journal, which boasts of its 42-year tradition of excellent research publishing.

The Information Society 2018 Multiconference consists of the following conferences:

- Slovenian Conference on Artificial Intelligence
- Cognitive Science
- Data Mining and Data Warehouses - SiKDD
- International Conference on High-Performance Optimization in Industry, HPOI
- AS-IT-IC Workshop
- Facing demographic challenges
- Collaboration, Software and Services in Information Society
- Workshop Electronic and Mobile Health and Smart Cities
- Education in Information Society
- 5th Student Computer Science Research Conference
- International Technology Transfer Conference (ITTC)

The Multiconference is co-organized and supported by several major research institutions and societies, among them ACM Slovenia, i.e. the Slovenian chapter of the ACM, Slovenian Artificial Intelligence Society (SLAIS), Slovenian Society for Cognitive Sciences (DKZ) and the second national engineering academy, the Slovenian Engineering Academy (IAS). On behalf of the conference organizers, we thank all the societies and institutions, and particularly all the participants for their valuable contribution and their interest in this event, and the reviewers for their thorough reviews.

For the sixth year, the award for life-long outstanding contributions will be presented in memory of Donald Michie and Alan Turing. The Michie-Turing award will be given to Prof. Saša Divjak for his life-long outstanding contribution to the development and promotion of information society in our country. In addition, an award for current achievements will be given to Assist. Prof. Marinka Žitnik. The information lemon goes to decreased national funding of research. The information strawberry is awarded to the Yaskawa robot factory in Kočevje. Congratulations!

Mojca Ciglarič, Programme Committee Chair

Matjaž Gams, Organizing Committee Chair

KONFERENČNI ODBORI
CONFERENCE COMMITTEES

International Programme Committee

Vladimir Bajic, South Africa
Heiner Benking, Germany
Se Woo Cheon, South Korea
Howie Firth, UK
Olga Fomichova, Russia
Vladimir Fomichov, Russia
Vesna Hljuz Dobric, Croatia
Alfred Inselberg, Israel
Jay Liebowitz, USA
Huan Liu, Singapore
Henz Martin, Germany
Marcin Paprzycki, USA
Karl Pribram, USA
Claude Sammut, Australia
Jiri Wiedermann, Czech Republic
Xindong Wu, USA
Yiming Ye, USA
Ning Zhong, USA
Wray Buntine, Australia
Bezalel Gavish, USA
Gal A. Kaminka, Israel
Mike Bain, Australia
Michela Milano, Italy
Derong Liu, USA
Toby Walsh, Australia

Organizing Committee

Matjaž Gams, chair
Mitja Luštrek
Lana Zemljak
Vesna Koricki
Mitja Lasič
Blaž Mahnič
Jani Bizjak
Tine Kolenik

Programme Committee

Franc Solina, co-chair
Viljan Mahnič, co-chair
Cene Bavec, co-chair
Tomaž Kalin, co-chair
Jozsef Györkös, co-chair
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Marko Grobelnik
Nikola Guid
Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Mitja Luštrek
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Vladislav Rajkovič
Grega Repovš
Ivan Rozman
Niko Schlamberger
Stanko Strmčnik
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Tanja Urbančič
Boštjan Vilfan
Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah

KAZALO / TABLE OF CONTENTS

Zbornik 21. mednarodne multikonference

INFORMACIJSKA DRUŽBA – IS 2018

Zvezek E

Proceedings of the 21st International Multiconference

INFORMATION SOCIETY – IS 2018

Volume E

Delavnica AS-IT-IC
AS-IT-IC Workshop

Uredila / Edited by

Matjaž Gams, Jernej Zupančič

http://is.ijs.si

12. oktober 2018 / 12 October 2018
Ljubljana, Slovenia

PREDGOVOR

Delavnica AS-IT-IC omogoča predstavitev primerov uporabe ter izmenjavo izkušenj med znanstveniki in drugimi deležniki na področju pametnega turizma, ki ga omogočajo inteligentna orodja in storitve, podprte z informacijsko-komunikacijskimi tehnologijami (IKT), predvsem umetno inteligenco (UI). Delavnica omogoča krepitev vezi in sodelovanja med izvajalci praktičnih turističnih storitev in znanstveno-raziskovalno sfero in spodbuja uporabo naprednih rešitev v turizmu. Delavnica je ena izmed aktivnosti projekta *Avstrijsko-Slovenski Turistično-Informacijski Center* (AS-IT-IC), ki je bil sprejet na Programu sodelovanja Interreg V-A Slovenija-Avstrija 2014-2020. Glavni cilj projekta je operativni center, kjer ponudniki turističnih informacij in virtualni asistenti sodelujejo pri odgovarjanju na turistično orientirana vprašanja turistov in jim pomagajo pri načrtovanju izletov na Slovensko-Avstrijskem čezmejnem območju.

Sprejeti prispevki opisujejo stanje projekta AS-IT-IC eno leto pred zaključkom projekta. Prispevek *Avstrijsko-Slovenski Inteligentni Turistično-Informacijski Center: Poročilo o napredku projekta* povzame napredek glede na projektne in programske kazalnike, med tem ko se ostali prispevki osredotočajo na posamezne komponente končnih projektnih rezultatov. V *Pregledu IKT orodij v turizmu* so predstavljene različne IKT rešitve za pomoč turistom in ponudnikom turističnih infomacij. V *Podatkovnih zbirkah AS-IT-IC* so predstavljeni podatki ter podatkovni servisi, ki so na voljo na platformi AS-IT-IC. V prispevku *API za podatke*, je predstavljen sistem za pridobivanje ter dostop do podatkov. V e-Turist2.0 je predstavljena nadgrajena verzija sistema za načrtovanje in priporočanje izletov. V prispevku *Testiranje varnosti pogovornih asistentov z uporabo planiranja* avtorji opisujejo napredni sistem za ugotavljanje varnostnih pomanjkljivosti pogovornih asistentov.

INTRODUCTION

The AS-IT-IC Workshop is a forum for presenting the use cases and exchanging experience among academic and service industry partners on deploying intelligent information communication technology, in particular artificial intelligence, supported tools and services for enabling smarter tourism, as well as stimulating further adoption of such solutions through promotional activities and establishing direct collaboration between academia and industry.

The workshop is an activity of the Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020 project Austrian-Slovenian Intelligent Tourist-Information Center (AS-IT-IC project). The main project output will be the operational center with human and virtual assistants enabling automatic answering to the tourism-oriented questions in natural language and performing services to enable trip planning in the Slovenian-Austrian cross-border region.

Accepted papers describe the AS-IT-IC project state one year before the project conclusion. *Austrian-Slovenian Intelligent Tourist-Information Center: Project Progress Report* summarizes the project progress with respect to project and programme indicators, while other contributions focus on specific modules of the final project results. In *Tourism Related ICT Tools: a Review* different ICT solutions with the aim to help tourist and tourist information providers are presented. In *AS-IT-IC Databases* the data and data services made available through the AS-IT-IC Platform are described. In the *Content API* paper the system for retrieving and serving the data is presented. In *e-Tourist2.0* authors write about the upgraded trip planning and recommendation solution. Finally, the smart security testing for security leaks for common attack scenarios is presented in *Planning-based Security Testing for Chatbots* paper.

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Matjaž Gams, IJS (chair)

Franz Wotawa, IST (co-chair)

Josip Božič, IST

Jernej Zupančič, IJS

Tomaž Šef, IJS

Oliver August Tazl, IST

Dieter Hardt-Stremayr, GRAZ

Katarina Čoklc, ZOS

Marija Lah, SPOTUR

Austrian-Slovenian Intelligent Tourist Information Center: Project Progress Report 2018

Jernej Zupančič
Jožef Stefan Institute and
Jožef Stefan International Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia
jernej.zupancic@ijs.com

Matjaž Gams
Jožef Stefan Institute and
Jožef Stefan International Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia
matjaz.gams@ijs.si

ABSTRACT
Austrian-Slovenian Intelligent Tourist Information Center (AS-IT-IC) is a project that was accepted in the Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020 call. The project goal is to create a joint Austrian-Slovenian center – an information and communication technology (ICT) supported network of service providers and tourist offices, municipalities, tourists and citizens to enhance continuous cooperation between them. The main project output will be the AS-IT-IC operational center with humans involved, having support of the ICT tools for communication, automatic question answering in natural language, information provision, trip recommendation and trip planning. This paper overviews the current state of the project progress.

Keywords
virtual assistants; chatbots; chat platforms; tourism; natural language understanding; AS-IT-IC project

1. INTRODUCTION
According to [1] a tourist cannot get the desired information in an integrated way from both humans and Web services, and much less the joint Austrian-Slovenian services. Typically, Slovenian or Austrian tourist office will provide only predefined national tours and not user-centric cross-border tours. As a consequence, tourists may miss locations they might be interested in visiting and tourist locations get less visits. The goal of the project is to create a joint Austrian-Slovenian center – an ICT supported network of service providers and tourist offices, municipalities, tourists and citizens to enhance continuous cooperation between them. Thus, cross-border tourist exchange, collaboration and expertise transfer between providers will largely increase with respect to the current state. The main project output will be an operational center having support of the following tools: Virtual assistant (providing automatic answering in natural language to the questions and performing services according to demands from tourists), Communication service (ICT solution that will enable conversation between the tourists, virtual assistants, tourist information workers and local communities), Information sources (inclusion of existing information sources), Recommender system for tour planning, Network of tourist services and services from local communities.

The system will help tourists better plan their cross-border visits, discover less popular sites that would otherwise be missed, stay longer and better satisfy their needs. Local communities will easily offer local services and information to visitors, e.g. a tour might include visiting a specialized craftsman and boost the selling of local products. Tourist officers will get better access to tourists. AS-IT-IC project (Table 1) provides the integration of virtual and human services from Austria and Slovenia with the uniform functionality – to provide most relevant information, attract tourists, and prolong their stay.

Table 1: Project information card

Title	Austrian-Slovenian Intelligent Tourist Information Center
Partners	1. Institut "Jožef Stefan" (lead partner) 2. Technische Universität Graz, Institut fűr Softwaretechnologie 3. Javni zavod za turizem, šport, mladinske in socialne programe SPOTUR Slovenj Gradec 4. Združenje občin Slovenije 5. Graz Tourismus und Stadtmarketing GmbH
Duration	From 1. 7. 2016 to 30. 6. 2019

The rest of the paper goes as follows. In Section 2 we describe the state of deliverables and project workpackages, in Section 3 we describe the project idea, while Section 4 overviews the state of the prototypes and provides information on what has been accomplished by now. Sections 5 and 6 describe the project dissemination activities and project impact, respectively, so far. Section 7 concludes the paper.

2. PROJECT PROGRESS
The project has entered the last year of implementation (Figure 1). While the majority of the technical details (Table 2) have been resolved, the project results are still under active implementation and testing. Additionally, the dissemination strategy and sustainability plan will be addressed in more details in the coming months.

3. PROJECT IDEA
The AS-IT-IC project tries to combine several solutions that already provide partial solutions for smarter tourism: attraction discovery, trip planning [2], and communication

Figure 1: Project Gantt chart

Table 2: Project deliverables status

Management	Project reports	3/6
Communication	Dissemination and promotion report	not started
	Promotion material	✓
	Publications	2/4
	Scientific publications	7/1
	Workshops on AS-IT-IC	4/3
	Participation in tourist related events	0/3
	Project website	✓
Tourist information platform	System requirements and specification	✓
	Tourist information platform	in progress
	Content items	✓
	Content creation guidelines	✓
	Communication applications	in progress
Virtual assistant	Virtual assistant requirements	✓
	Virtual assistant service prototype	in progress
	Virtual assistant service	not started
Tour planning	Tour planning requirements	✓
	Tour planning service prototype	in progress
	Tour planning service	not started
AS-IT-IC Center	AS-IT-IC Deliverable	✓
	AS-IT-IC Center	in progress

with human and virtual assistants [3, 4]. In order to provide the state-of-the art platform that enables smarter tourism several open source technologies, data sources and internal tools and services were examined, upgraded and are in the process of integration into one tourism platform – the AS-IT-IC platform. Using the open source software enables us to start with a solid working solution and provide necessary modifications as required by the project. A simplified reference architecture is presented in Figure 2.

3.1 Communication platform

The communication platform enables the users to communicate with each other (tourist – tourist, tourist – tourist information provider, tourist – virtual assistant) over the chat based interface. Increased popularity of chat applications (Facebook messenger[1], WhatsApp[2]) prove that this is a valid communication option used for exchanging and obtaining information. The main benefit being the option to upgrade the communication by integrating various virtual assistant services.

3.2 Virtual assistants

Virtual assistants (also chatbots or conversational robots [3]) are computer programs that can process input in natural language and provide a reply. The input can either be voice or text and the answer is usually a combination of a response in natural language and an action that was carried out by taking into account the user input. An example would be as follows. User asks "What are some cultural heritage sights I could visit near me?". The virtual assistant would then first identify the intent (the user would like to execute a search) and the arguments (location: near the user, type of sight: cultural heritage). Then it would acquire user location and user preferences from the system and issue a request to the system database in order to obtain relevant attractions. The

[1] `https://www.messenger.com/`
[2] `https://www.whatsapp.com/`

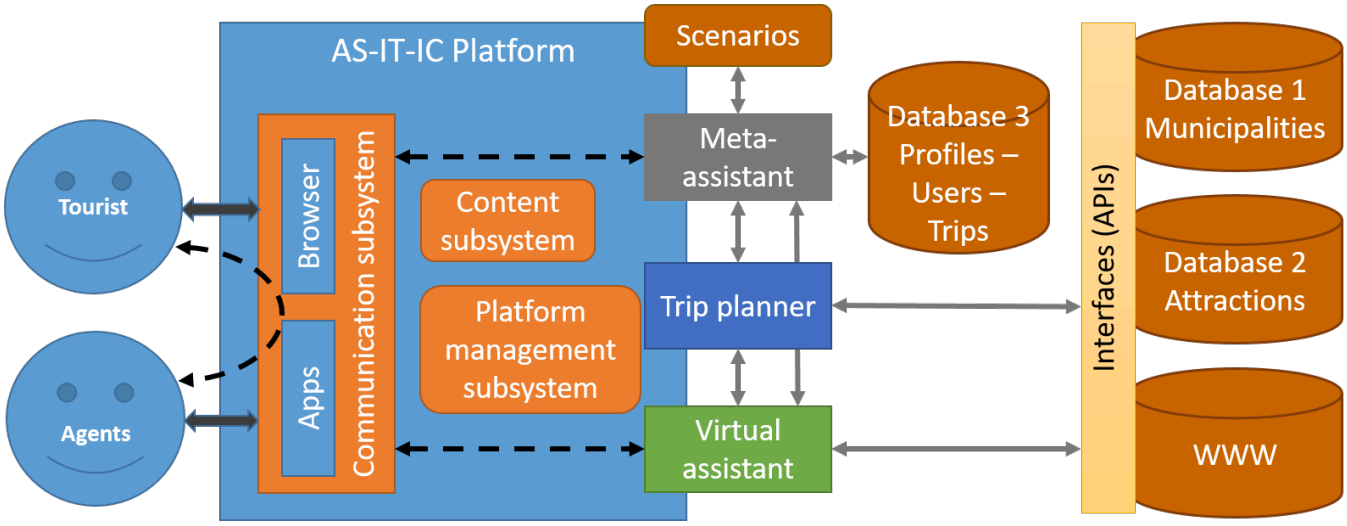

Figure 2: Simplified architecture corresponding to the project idea

results would then be properly formatted and presented to the user: "I have found the following attractions that match your search: Ljubljana castle, Cankar memorial house and Manor pavilions. Would you like to learn more about a specific attraction?".

3.3 Trip planner

Trip planners [2] help the user in planning the trip by keeping track of the places the user intents to visit, recommending attractions and points of interest relevant for the user, and automatically arranging the itinerary in order to optimize the travel between the items on the itinerary. Trip planners usually also enable the user a visual overview of the whole trip and sometimes even enable the navigation.

3.4 Databases

At the heart of the platform are the databases that provide all the information required by the AS-IT-IC services. The databases provide structured data that can be used by several services for further processing. The databases consist of: the information about the attractions and other points-of-interest (castles, caves, restaurants, etc.); the information about the geographical entities (places, regions, rivers, municipalities, etc.); and information scraped from the useful webpages (municipality information, opening hours, etc.).

Beside the "Content" databases the system also requires databases for user management and storing of the system states and user generated content (conversation, saved trip itineraries, etc.).

4. PROTOTYPES

The AS-IT-IC platform will consists of services deployed either using Docker[3] virtualization technologies or Flynn[4] – a self-hosted platform-as-a-service.

4.1 Communication platform

As a base an open source team communication software Rocket.Chat[5] was chosen. In order to meet the project requirements, several additions were developed: a custom home dashboard; a message modification for improved user experience; a custom information tab with information about the trip; notification modifications for better operation of mobile communication application; custom application programming interface for automatic message processing and posting. Screenshot of the conversation user interface is presented in Figure 3.

4.2 Virtual assistants

The virtual assistant used in the AS-IT-IC platform comprises several modules. Two approaches were used when designing the assistant modules: rule-based approach (which is an upgrade of virtual assistants deployed at Jožef Stefan Institute and the majority of Slovenian municipalities [3]); and natural language based approach [4]. The rule based models are more stable and easier to debug and understand, however, they have the issue of rule design, since every rule has to be designed by hand, which is why they take a long time to implement. The natural language based modules, however, enables one to produce a virtual assistant that transforms the natural language input into a structured format that can be further used by computer programs. The main disadvantage of such systems is the need of a language model (which is an active area of research, especially for smaller languages such as Slovenian) and the need for a large set of training data.

Within the AS-IT-IC Platform the rule-based approach is used for the virtual assistant action that results from the user interaction with the uniquely identifiable objects present in the user interface (for instance buttons) and for common text input provided by the user. The natural language based approach is used for intelligent search capabilities and in cases where the rule-based approach fails to work. Two backends are currently used for parsing the user input and

[3] https://www.docker.com
[4] https://flynn.io/
[5] https://rocket.chat/

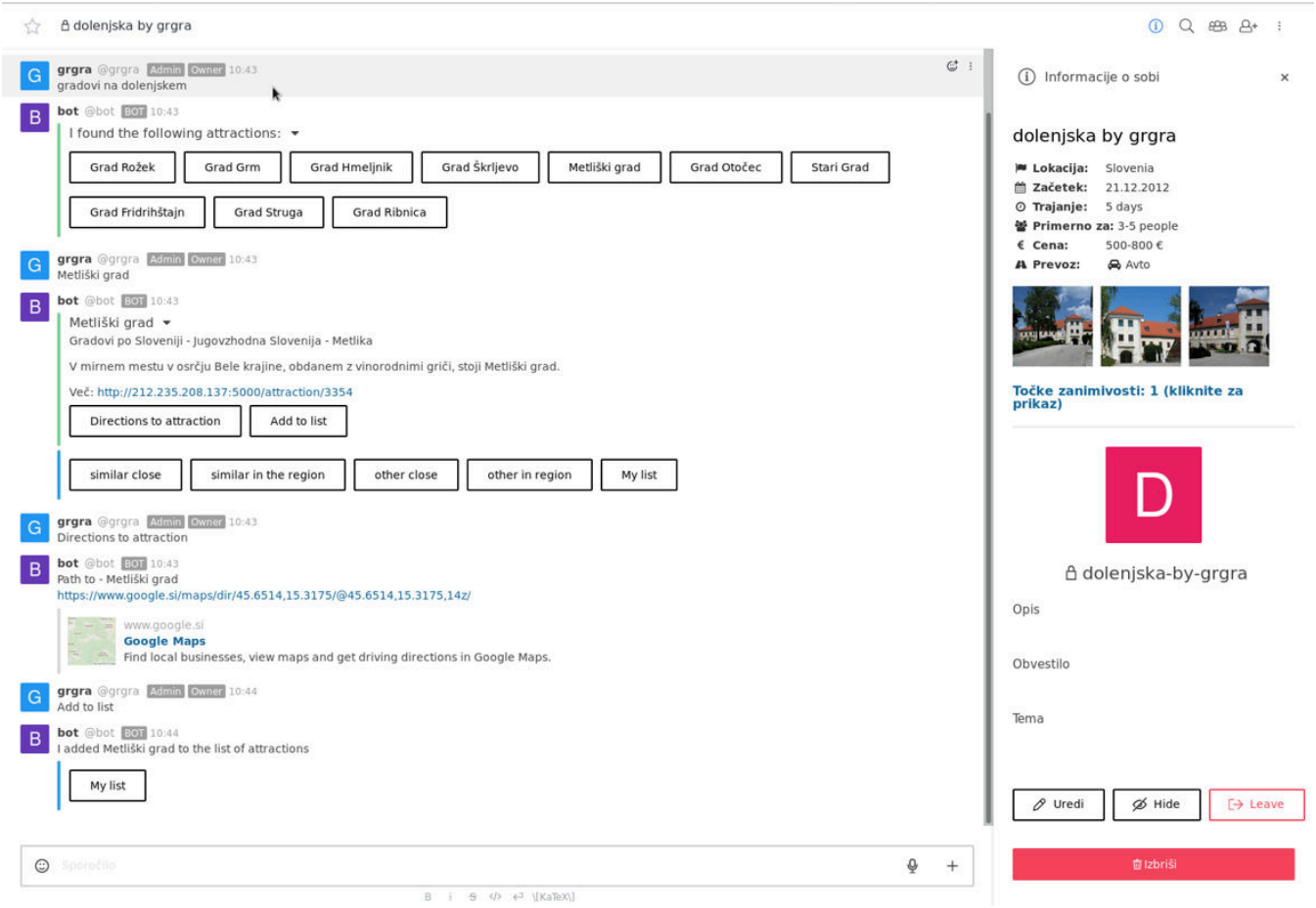

Figure 3: Conversation user interface

translating it into structured text – Dialogflow[6] and Rasa[7]. Additionally, in some cases we take the advantage of the full-text search capability of the PostgreSQL[8] database – we have made the required adjustments for the full-text search to work in all three project languages: Slovenian, English and German.

Adapters were developed that enable the interaction between the communication system and the virtual assistant services: reading the user input, processing the text, performing required actions and returning a response to the user.

4.3 Trip planner

A basis for the trip planner used in the AS-IT-IC platform was the e-Tourist application [2]. In order to meet the project requirements heavy modifications were made to the e-Tourist codebase: the databases were extended to support additional attraction and geographical data; the routing was adjusted so that the trip planner can be used by third party applications; user management was upgraded to enable AS-IT-IC platform users to automatically log-in into the system; several functions that were previously coded by hand were moved to the database, which significantly reduced the project size; and all the modules used by the application were upgraded to the latest versions, therefore increasing the application security.

4.4 Databases

Webpages and datasources that contain relevant tourism-oriented information and enable the use of the content for non-commercial purpose were reviewed and gathered. While there are many datasources with relevant information (Slovenia.info[9], DEDI[10], OPSI[11], Geoportal ARSO[12], e-Geodetski podatki[13], eVode[14], register kulturne dediščine[15]) there was additional work needed to unify the data formats, remove the data that was not of sufficient quality and integrate all

[6] https://dialogflow.com/
[7] https://rasa.com/
[8] https://www.postgresql.org/
[9] https://www.slovenia.info/en
[10] http://www.dedi.si/
[11] https://podatki.gov.si/
[12] https://gis.arso.gov.si
[13] https://egp.gu.gov.si/egp/
[14] http://www.evode.gov.si/sl/vodni-kataster/zbirka-vode/zbirka-podatkov-o-povrsinskih-vodah/hidrografija/
[15] http://www.mk.gov.si/si/storitve/razvidi_evidence_in_registri/register_nepremicne_kulturne_dediscine/

the data into one database.

5. DISSEMINATION

Project partners have been active in disseminating the project results, producing scientific papers at conferences and international journals, producing publications for general public, maintaining the project website, hanging project posters at partner sites, and organizing workshops.

Scientific papers and publications for general public introduced the project to the wider audience by presenting the project idea, describing the need for such a project, and presenting tools, services and prototypes, developed within the project. Project partners have so far contributed to 7 scientific papers and 2 publications for general audience.

Project website[16] was deployed in the first half year of the project. It presents all relevant information about the project and project partners, together with the project news. Additionally it enables the visitors of the website to contact the project partners.

Project posters were designed according to the Cooperation Programme Slovenia-Austria rules and posted at partner sites (Jožef Stefan Institute, GUT Institute for Software technologies, and Association of municipalities of Slovenia).

Workshops are one of the main dissemination channels, where partners invite general audience to attend or the partners present the project and project results to the general audience. Several workshops were organized, where AS-IT-IC project was presented: "AS-IT-IC Workshop" within the Information society 2017 Multiconference, "Presentation of tourism applications" workshop on behalf of the invitation in Nazarje, "Artificial Intelligent into every municipality" workshop organized at Jožef Stefan Institute, and a "Site visit" workshop organized at Jožef Stefan Institute, where Jožef Stefan Institute employees were invited.

6. PROJECT IMPACT

The AS-IT-IC project enables the project partners to greatly increase the cooperation in the cross border area. The partners have organized five cross-border meetings, which resulted in the exchange of information, data, examples of good practice in the field of tourism and also in additional project application. The partners also cooperated in the organization of workshops, which enabled the project partners to reach several third party stakeholders: 7/50 (7 reached out of 50 promised) representatives from local public authorities – entities were reached so far (ministries and municipalities); 10/30 representatives from interest groups including non-governmental organizations (development centers, tourist organizations); 7/5 small and medium sized companies; and 89/3000 interested individuals.

The main communication goals of the project are:

1. Integrated tourist communication service: to raise awareness of the AS-IT-IC platform. AS-IT-IC was mentioned five times in press and web page news.
2. Tourist workers networking: to change behaviour and increase the cooperation of tourist workers and providers of ICT tools for tourism. Nine meetings and workshops were organized to promote AS-IT-IC and enable such networking.
3. eHeritage: to raise the awareness of the need to make the description of heritage sights and attractions available on the Internet in a way that enables easy search and the inclusion of such attractions into the tourists' itineraries. So far cca. 1000 heritage attractions were inserted into the AS-IT-IC content database.

[16] `https://as-it-ic.ijs.si`

7. CONCLUSION

In the paper we presented the AS-IT-IC project, its goals and the issues it addresses. Project partners from the Slovenian-Austrian cross-border area came together in order to enable smarter tourism by integrating several tools and services into one platform. The project has entered its last year of implementation and so far the development has gone according to the plan. The prototypes for communication platform, virtual assistant, and trip planners are under development and will soon be ready for integration into the AS-IT-IC platform and ready for testing by end users.

The majority of the future work will be on disseminating the project results and attracting a larger user base. To this end the tourist partners are in contact with the major tourist organizations (such as Slovenian tourist board), which will enable us to reach a wider target audience and receive useful user feedback.

Additionally, the partners will look into sustainable options for transferring the AS-IT-IC Platform management to interested organizations. This will enable additional growth of the AS-IT-IC Center, while the partners will maintain the functionality of the AS-IT-IC platform as developed for the purpose of the project.

8. ACKNOWLEDGMENTS

The work was co-funded by Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020, project AS-IT-IC.

9. REFERENCES

[1] AS-IT-IC Project partners. About the project. `https://as-it-ic.ijs.si/about/`, 2016. [Online; accessed 16-September-2018].

[2] B. Cvetković, H. Gjoreski, V. Janko, B. Kaluža, A. Gradišek, M. Luštrek, I. Jurinčič, A. Gosar, S. Kerma, and G. Balažič. e-turist: An intelligent personalised trip guide. *Informatica*, 40(4):447, 2016.

[3] D. Kužnar, A. Tavčar, J. Zupančič, and M. Duguleana. Virtual assistant platform. *Informatica*, 40(3):285, 2016.

[4] J. Zupančič, G. Grasselli, A. Tavčar, and M. Gams. Virtual assistants for the austrian-slovenian intelligent tourist-information center. In *Proceedings of the 12th International Multiconference Information Society - IS 2017*, volume E, pages 27–30, Ljubljana, Slovenia, 2017. Jožef Stefan Institute.

Tourism Related ICT Tools: a Review

Gregor Grasselli
Jožef Stefan Institute and
Jožef Stefan International Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia
gregor.grasselli@ijs.si

Jernej Zupančič
Jožef Stefan Institute and
Jožef Stefan International Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia
jernej.zupancic@ijs.com

ABSTRACT

In the paper we review the existing information and communication technology (ICT) based tools and services that empower tourists and tourist information providers in obtaining and providing information needed for trip planning. We define four tourism-related service categories: search with booking, trip planners, chatbots, and forums. We summarize the good practices identified in the reviewed tools and expose the issues stemming from the fragmentation of the tools and data. In order to overcome the identified problems we propose the AS-IT-IC Platform – a project result of the Austrian-Slovenian Intelligent Tourist Information Center (AS-IT-IC) project, accepted in the Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020 call. The AS-IT-IC project aims to integrate several functionalities usually offered by distinct services, unifying the databases, and providing free access to tools and services for tourists and tourist information providers.

Keywords

virtual assistants; chatbots; tourism; trip planning; search; review; AS-IT-IC project

1. INTRODUCTION

Due to the increase in the Internet usage it has become paramount that organizations and their tools are accessible through the Internet. In the last few years a steep increase in the mobile usage has additionally fueled the development of mobile applications and mobile friendly web applications, several of those in the domain of tourism.

In this paper we present the tools provided by organizations that operate on a global scale. We categorize the tools into four categories:

1. Search with booking
2. Trip planners
3. Chatbots – conversational robots
4. Forums

Each category is unique: in the ways it helps the user in the process of trip planning; in the type of data it uses; and in the target users. This, however, leads to the fragmentation of data and the need for: tourists to check several web sites and applications, when searching for the trip-related information; tourists to call or go to local tourist information offices for additional info; tourist information providers to keep their information available in several systems all over the internet and structured in different – often incompatible ways; tourist information providers to pay for the presence on tourism-tools and services. There is clearly room for improvement [3] and the first step would be the establishment of an online platform that would in one place integrate several services and tools and enable the users to gather and provide all necessary information related to trip planning. One such platform is being developed within the Austrian-Slovenian Intelligent Tourist Information Center (AS-IT-IC) project.

The rest of the papers continues with the description of four categories of tools and services in Sections 2, 3, 4 and 5. Good practices are summarized in Section 6 and the need for an integrated solution is addressed in Section 7. Section 8 concludes the paper.

2. SEARCH WITH BOOKING

The group offering the least functionality is the one of search engines for hotels, plane flights, restaurants and such. The user is mostly required to input a predetermined set of parameters, and is then referred to a list of matching options they can choose from. This list can sometimes be further filtered and/or sorted based on additional criteria. After deciding on what suits them best, the user is relegated to booking, which is how the site makes money.

Of course most if not all hotels, airlines, tourist agencies and other tourist providers have their own websites, which allow a prospective user to directly book their services. The examples below are of search engines that accumulate data from many such sites, or get it directly from the end providers and try to make it more convenient for a prospective tourist to find what they are looking for.

- Sabre – `https://en.eu.sabretravelnetwork.com/home/page/book_flights_hotels`
- Expedia – `https://www.expedia.com`
- TripAdvisor – `https://www.tripadvisor.com`
- Priceline – `https://www.priceline.com`
- Yelp – `https://www.yelp.com`

- OpenTable – https://www.opentable.com/start/home
- Booking.com – https://www.booking.com
- Skyscanner – https://www.skyscanner.com

3. TRIP PLANNERS

A slightly more interesting group is the group of trip planners. They also require the user to input a set of predetermined parameters, but unlike in the previous group where the parameters are mostly the same on all websites, here they vary quite a lot from service to service. What is different is also that these services mostly output a single trip plan which can later be modified. After a user decides on their final trip plan, they are offered accommodation and transportation booking necessary for the trip's realization. The booking stage on some of the websites utilizes services from the first category. Some interesting examples of such services are provided below.

- Roadtrippers – https://roadtrippers.com
- Mapquest – https://www.mapquest.com/routeplanner
- Triphobo Tripplanner – https://www.triphobo.com/tripplanner
- Inspirock – https://www.inspirock.com
- Sygic Travel – https://travel.sygic.com
- routeperfect – https://www.routeperfect.com/trip-planner
- wanderapp – https://www.wanderapp.me
- Go Real Europe – https://www.gorealeurope.com

4. CHATBOTS

With the emergence of natural language processing tools like dialogflow[1], wit[2] and rasa[3] that make it quite easy for developers to implement chat based interfaces in a growing number of languages, there has been an explosion in the number of applications and websites that offer a chat based interface. There are even companies like Botflux[4] that offer their customers custom made chatbots.

Facing the consumer, chatbots offer a more dynamic (request values for different parameters based on the user's input so far) and interactive approach to defining the user's requirements than classical forms and menus. They also offer a completely new experience when using a voice interface through a smartphone or a specialized device such as Amazon's Echo or Google Home. A good resource for finding interesting chatbots based on the messaging platform, where one can converse with them, and their area of expertise is botlist.co[5].

[1] https://console.dialogflow.com
[2] https://wit.ai
[3] https://rasa.com
[4] https://www.botflux.com/tourism
[5] https://botlist.co/

4.1 Customer service chatbots

One of the trends in using chatbots is to automate customer service on a company's website. Things like providing answers to frequently asked questions, or finding relevant information on the website without having to manually search for it can easily be automated by chatbots. The aim of these bots is to reduce the load on human customer service staff and to provide a better customer experience by making it easier and faster to get answers from data on the website. In this context a chatbot is an addition to an existing website, often appearing as a chat window in one of the corners.

Examples of such chatbots are:

- Ana – https://connectmiles.copaair.com/en/web/guest/ask-ana
- Julie – https://www.amtrak.com/about-julie-amtrak-virtual-travel-assistant

4.2 Instant messaging chatbots

With rising popularity of instant messaging platforms like Facebook's messenger[6] and Telegram[7] that expose APIs for bots to converse with users, many bots now exploit this channel for access to users. By responding to messages in group chats (possible on Facebook's messenger for instance), chatbots are a new way for making a product more discoverable and for making the user's trip from wish to purchase shorter.

In many cases chatbots present on an instant messaging platform are just a different interface to an existing service. Examples of such bots are:

- Expedia – https://viewfinder.expedia.com/features/introducing-expedia-bot-facebook-messenger
- Skyscanner – https://www.skyscanner.net/news/tools/skyscanner-facebook-messenger-bot
- Cheapflights – https://www.cheapflights.co.uk/news/cheapflights-chat-awards
- Hello Hipmunk – https://www.hipmunk.com/hello

There are a number of chatbots that act as aggregators over different services. They provide a conversational interface for searching offers from many sources and providing the user with the result that most closely matches their requirements. An example of such a bot is Assist[8], a chatbot that aggregates several services for making hotel reservations, ride hailing, making table reservations and online shopping. It is the only product of a start up with the same name and can be used through messenger, Telegram, SMS, Twitter, Google Assistant and Slack.

According to statista[9], business travel in 2016 amounted to 1.3 trillion USD and represented about 10% of all travel

[6] https://www.messenger.com/
[7] https://telegram.org/
[8] http://www.assi.st/
[9] https://www.statista.com/topics/2439/global-business-travel-industry

spending in 2015. Thus it makes sense that a number of travel chatbots are specifically targeting business travellers:

- Carla – `https://www.cwtcarla.com/CarlaWeb`
- Pana – `https://pana.com`
- MEZI – `https://mezi.com`

A different kind of application is a so called virtual concierge. Its main goal is to assist in communication between hotel staff and their guests by providing an interface for checking into a hotel, ordering room service, and requesting information about the hotel. They have automatic translation integrated into the service, so customers can interact with the staff in their native language. These applications are in this category because they mainly function through a conversational interface and use instant messaging technology. Also some of their functions are fully automated so they are bots and not just messaging apps with translations. An example of such a virtual concierge is The Besty[10], a phone app to help people communicate with their hotel's staff as well as find and book tours, restaurants and activities at the "lowest" prices. They also offer tour guides and live chat with local tour experts. The MEZI chatbot, mentioned earlier also offers a virtual concierge service of this kind as part of its capabilities. Another example of this kind of chatbot is Hi Jiffy[11]. It is available on messenger, and allows searching for hotels and making reservations in addition to its customer care functionality. It employs a model where queries that cannot be answered automatically are forwarded to hotel staff. The query and the provided answer are then included in the bot's training set so that it can answer automatically when a similar query is input by a user. According to its website 77% of its answers are provided automatically at the time of this writing.

5. FORUMS

When people want opinions, recommendations or advice, they turn to the forums, where they can ask questions related to their planned excursions and get answers from travellers who have been there before. On some of the forums, travel agents seem to be quite involved as well, answering questions by prospective tourists while advertising their services. All of the forums we came across cover travelling to the whole world, but usually have a different section for each continent which is then further divided by country.

Most forums also have a section dedicated to posting longer accounts of travellers who believe they have experienced something worth sharing. An interesting website that collects longer posts by travellers as well as photos is Travelblog[12]. In addition to their forum for discussing travel plans and asking for advice, they also have a blog section, where anyone can write about their experiences or post a photo they think is particularly eye catching, and the rest of the users will vote on the best blog and photo of the week.

In most cases travel related forums appear as part of a bigger travel related website. The popular TripAdvisor website also includes a typical travel forum[13] of this kind. Questions and responses are checked for destinations and attractions TripAdvisor knows and if any are found, they are displayed in a card, below the user's post, showing their name, a picture and their ranking. Clicking on the card will show the site for that attraction.

Another example of a website that also includes a forum is Lonely Planet[14]. According to Wikipedia[15] they are the largest travel guide publisher in the world. They also provide a website that would fall into the search with booking category, coupled with their newsletter and of course selling travel guides. This rather expansive website also includes a forum for exchanging "travel advice, hints and tips" as they put it. The forum is not limited to country based discussions but also has (among others) sections about equipment, travel health and vaccinations, searching for travelling companions, house sitting and swapping as well as people selling and buying stuff through the forum.

According to the quick analysis of posts, Fodor's Travel Talks Forum[16] seems to be the most popular. They also sell guidebooks and have a very extensive website that also offers hotel bookings.

As we are living in the age of social media, it would be remiss not to mention the `#travel` tag on Twitter[17], used to post about travel experiences, as well as the existence of quite a number of twitter accounts that are dedicated solely to travel news[18].

6. GOOD PRACTICES

By reviewing the existing ICT solutions available on the internet, the following commonalities and good practices were observed.

1. Since almost 53% of all internet traffic in 2017 was produced through mobile devices[19], having a mobile application or a different way of making the application work on a mobile phone (like through messenger, or a mobile-first web application) is a must.
2. When possible it is a good idea to integrate with applications users are already using in their everyday life, like calendars. This allows to get user data without needing the user to type everything, as well as enables the user to use the results of an application more conveniently.
3. Integrating multiple data sources into a single view is very helpful for users as they get more complete

[10] `https://thebesty.com/`
[11] `http://hijiffy.com/`
[12] `https://www.travelblog.org`
[13] `https://www.tripadvisor.com/ForumHome`
[14] `https://www.lonelyplanet.com/thorntree/welcome`
[15] `https://en.wikipedia.org/wiki/Lonely_Planet`
[16] `https://www.fodors.com/community`
[17] `https://twitter.com/`
[18] `http://mashable.com/2012/08/04/travel-twitter/#KtwUdPjm_Gqw`
[19] `http://www.trendreports.com/article/technology-in-tourism`

information about a destination and do not have to check multiple sources on their own.

4. Many popular applications scrape provider websites for information and special offers. Others rely on providers to manually enter all information.
5. Sites that cover a wider geographical area are more useful, since they provide a one-stop shop for the whole trip as opposed to having to visit several websites to get informed on each destination individually.
6. Availability through multiple channels. Having a website is fine, but also being available through other channels, especially instant messaging platforms really helps with discoverability.
7. Until the invention of general AI, machines will be limited in what they can do, so to minimize customer frustration, keeping humans in the loop on the provider side can be very helpful.
8. Making customization of automatically generated trip plans and other suggestions as easy and complete as possible, or the users will only use the tool to get the suggestions then they will use more low level tools to actualize the parts they liked. This lowers the conversion rate of the tool and makes customers less happy.
9. Allowing users to filter and sort displayed information based on their interests. A good example is how Roadtrippers allows users to set what kinds of points of interest they want to see on the map.

7. RELATION TO AS-IT-IC

The AS-IT-IC project tries to combine several partial solutions already implemented by the ICT tools mentioned in this paper: attraction discovery, trip planning [1], and communication with human and virtual assistants [2, 4]. In order to provide the state-of-the art platform that enables smarter tourism several open source technologies and data sources were utilized and integrated into one tourism platform – AS-IT-IC platform.

AS-IT-IC project empowers tourists by: helping in obtaining all the required information related to trip planning in one place; and enabling discovery of local and less known but still relevant attractions. Further, it empowers tourist information providers by: providing an integrated way of exposing the tourism content in his or her area to the Internet; and enabling the access to the tourists in an asynchronous, modern chat-based style.

The reviewed tools, together with traditional communication methods, indeed already offer the same or at least very similar functionality as is planned for the AS-IT-IC platform. However, even disregarding the obvious benefit of the functionality integration into one platform, there are still advantages of the AS-IT-IC project results, for the time being mainly for the Slovenian-Austrian cross-border area:

1. Larger database of attractions.
2. Inclusion of path-based attractions – for instance wine roads, or walking trips and geographical information.
3. Cooperation of several local tourist information centers with vast knowledge on the touristic offers in their area.
4. Free access for tourists and tourist information providers.
5. Open access to data and data services.

The AS-IT-IC Platform, however, does have one big disadvantage compared to the rest of the services – it is a publicly founded project with reserved funds for the development and initial activities to raise the project awareness and disseminate results. After the end of the project, no resources have been granted yet to further promote project results. While the partners have committed to maintain the project results for another 5 years after the end of the project, the issue of getting sustainable funds to enable further promotion and dissemination is yet to be solved.

8. CONCLUSION

In this paper the ICT tools provided in order to empower smarter tourism are presented. Providers from around the world were taken into account. The tools were classified into categories in order to provide a sense of what is available for tourists and tourist service providers. Additionally, the tools were critically assessed and good practices were identified. Further, the AS-IT-IC Platform was compared to existing tools and main similarities and differences were pointed out.

The review provides a basis for anyone interested in the deployment of tourism-oriented services. One has to take into account, however, that not every problem in tourism has a technological solution. One of the main components of the AS-IT-IC project is the networking one, where the goal is to connect several stakeholders that provide technology solutions to the users in need of such solutions.

9. ACKNOWLEDGMENTS

The work was co-funded by Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020, project AS-IT-IC.

10. REFERENCES

[1] B. Cvetković, H. Gjoreski, V. Janko, B. Kaluža, A. Gradišek, M. Luštrek, I. Jurinčič, A. Gosar, S. Kerma, and G. Balažič. e-turist: An intelligent personalised trip guide. *Informatica*, 40(4):447, 2016.

[2] D. Kužnar, A. Tavčar, J. Zupančič, and M. Duguleana. Virtual assistant platform. *Informatica*, 40(3):285, 2016.

[3] B. Peischl, O. A. Tazl, and F. Wotawa. Open questions of technology usage in the field of tourism. In *Proceedings of the 12th International Multiconference Information Society - IS 2017*, volume E, pages 29–22, Ljubljana, Slovenia, 2017. Jožef Stefan Institute.

[4] J. Zupančič, G. Grasselli, A. Tavčar, and M. Gams. Virtual assistants for the austrian-slovenian intelligent tourist-information center. In *Proceedings of the 12th International Multiconference Information Society - IS 2017*, volume E, pages 27–30, Ljubljana, Slovenia, 2017. Jožef Stefan Institute.

AS-IT-IC Databases

Jernej Zupančič
"Jožef Stefan" Institute and
Jožef Stefan International
Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia
jernej.zupancic@ijs.com

Oliver A. Tazl
Institute for Software
Technology
Graz University of Technology
Inffeldgasse 16b
Graz, Austria
oliver.tazl@ist.tugraz.at

Blaž Mahnič
"Jožef Stefan" Institute
Jamova cesta 39
Ljubljana, Slovenia
blaz.mahnic@ijs.si

Gregor Grasselli
"Jožef Stefan" Institute and
Jožef Stefan International
Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia
gregor.grasselli@ijs.si

ABSTRACT

Austrian-Slovenian Intelligent Tourist Information Center (AS-IT-IC) is a project that was accepted in the Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020 call and has two main goals: one is to build information and communication technnology (ICT) tools to support the tourist when he or she creates personalized itinerary for the visit of Slovenian-Austrian cross-border area; and the second is to create a sustainable community that will support the use of the tools. In this paper we describe the provision, cleaning, integration and deployment of data and data services needed by the ICT tools in tourism. Data and data services form one of the main pillars that enables the AS-IT-IC platform to provide tools and services, which could serve tourism-related information to end users – tourists and tourist information providers.

Keywords

web spider; data; tourism; databases; web services; AS-IT-IC project

1. INTRODUCTION

There is an increasing number of services and applications available for tourists and tourist information providers across the Internet. The services could be roughly categorized into the following categories: *Search with booking*; *Trip planners*; *Virtual assistants*; and *Forums*.

Search with booking enables the user to search for a type of accommodation, transport or adventure, specifying the time interval of using the services and in some cases even buying the service or reserving it. Examples are Expedia[1] and OpenTable[2].

Trip planners enable the user to view attractions in certain area, obtain additional attractions by clicking on them and forming a customized trip. Examples of such services are Roadtrippers[3] and Triphobo Tripplanner[4].

Virtual assistant enables the user to obtain tourism-related information using a rich-text based interface, similar to the ones provided by Facebook Messenger[5] . Examples include Hello Hipmunk[6] and MEZI[7].

Forums provide a place where usually users but sometimes also professionals provide descriptions of their trips, express their opinions about attractions and places to visit, and provide helpful advice to fellow travellers. Examples include Fodor's Travel Talks Forum[8] and Lonely Planet[9].

In order to provide such services developers need data. The data can be obtained in several ways:

1. The information about attractions and other points of interest can be obtained in advance (as is the case with some trip planners and virtual assistants).
2. By making an application programming interface (API) call to an external service (search with booking).
3. By relying on users to provide the content when the application is already live (forums).

The ICT tools that will be used directly by the users and are developed within the AS-IT-IC project integrate several services that belong to the before-mentioned categories: communication platform that enables communication of tourists with tourist information providers and virtual assistants; virtual assistant, which provides useful information 24/7;

[1] https://www.expedia.com
[2] https://www.opentable.com/start/home
[3] https://roadtrippers.com
[4] https://www.triphobo.com/tripplanner
[5] https://www.messenger.com/
[6] https://www.hipmunk.com/hello
[7] https://mezi.com
[8] https://www.fodors.com/community
[9] https://www.lonelyplanet.com/thorntree/welcome

and tour planning for the automatic creation of a trip. In order to provide useful tourist information services, we had to combine several data procurement options, also used by other systems that usually cover only a part of the AS-IT-IC functionality.

Additional difficulty in obtaining the data was the fact that AS-IT-IC covers the Slovenian-Austrian cross-border area, for which little structured information is available. This is especially true for natural and cultural heritage attractions, which are the focus of the Interreg Slovenia-Austria Programme. We reviewed several data sources that are available for non-commercial use and tried to include the most relevant and quality ones.

According to the review, the following types of data was identified as useful: attraction data and tourism-related points-of-interest (natural and cultural heritage, sights, activities, accommodation, places to eat etc.); geographical data (structured representation of geographical entities such as rivers, lakes, municipalities, cities etc.); and services related to getting from place A to place B, so called routing services.

In the rest of the paper we review available data sources in Section 2. Further, we describe the data related to attractions in Section 3, the geographical data in Section 4, and the chosen routing system in Section 5. Section 6 provides a brief description of how the data will be used within the AS-IT-IC Platform and Section 7 concludes the paper.

2. DATABASES AND DATA SOURCES

The following data sources were identified as the most relevant:

1. Slovenia.info website[10] (Figure 1). The data source comprises 8798 tourist attractions and is still growing, due to the fact that the tourist information providers are constantly uploading and updating new attraction descriptions.

2. Dedi.si website[11]. The data provided by this website comprises only natural and cultural heritage, therefore being very suitable for the purpose of the project. However, due to the incompatible data formats, the inclusion of the Dedi.si data is postponed for the time being.

3. Europeana website[12]. This data source stores the data about artworks, artifacts, books, films and music from European museums, galleries, libraries and archives from around the world. The number of all entries is around 58 millions. However, due to the automatic data collection the information is very often wrong. We have decided to not include the Europeana data.

4. Open data portal of Slovenia[13]. The portal provides the information, tools, and useful resources, which can be used in web and mobile applications. There are fourteen data categories on the web site: Population and Society, Justice, the legal system and public safety, Public Sector, Education, Culture and Sport, Social and employment, Health, Environment and Spatial Planning, Transport and infrastructure, Agriculture, fisheries, Forestry and nutrition, Finance and Taxes, Economy, Energy, Science and Technology and International Affairs. Some of the datasets available on the portal are of interest also for the tourism domain, for instance a computer readable map of bodies of water, where bathing is possible.

5. Slovenian Cultural heritage register[14]. This is an official database of cultural heritage on the territory of the Republic of Slovenia, provided by the Ministry of Culture of Republic of Slovenia. The registry contains 30.095 entries of several types. The big disadvantage of this database, however, is that the use of the database is prohibited for online applications, which is a big drawback in the information age – especially since the data is of public interest.

6. Graz Tourism database. The data of this website comprises the tourism sights, attractions and offers of the city of Graz and its neighbouring regions. The data is available via a back-end using a REST-JSON API. The data is maintained by Graz Tourism and the tourism partners to provide detailed and high-quality data.

Figure 1: Tourist attractions from Slovenia.info

3. ATTRACTIONS DATA

This sections describes all the attributes used to represent an attraction datum. Data was structured in an appropriate way that will enable the AS-IT-IC services to provide different kind of functionality.

The data was imported into the PostgreSQL[15] database, where each datum insert is represented as follows (the data structure is based on the already developed e-Tourist systems [1]):

- title (sl, en, de, it): the title of attraction stored in four languages – Slovenian, English, German, Italian. For example "Gostišče Kimovec"

[10] https://www.slovenia.info/en/map
[11] http://www.dedi.si/
[12] https://www.europeana.eu/portal/en
[13] https://podatki.gov.si/
[14] http://www.mk.gov.si/si/storitve/razvidi_evidence_in_registri/register_nepremicne_kulturne_dediscine/
[15] https://www.postgresql.org/

- description (sl, en, de, it): description of tourist attraction in four languages
- category (sl, en, de, it): for example "Adrenaline sports"
- subcategory (sl, en, de, it): for example "Paragliding"
- location: GPS coordinates of tourist attraction, for example "(45.94,13.71"
- figure: image that represents the tourist attraction (web path)
- trip advisor: attraction rating retrieved from TripAdvisor[16] web site
- address: "Zgornji Hotič 15, 1270 Litija"
- recommended viewing duration time: "1:30:00"
- price range: how much does it cost to visit the attraction "1-5"
- expert evaluation: what is the expert opinion in the quality of the attraction "1-5"
- parking: parking options
- campers: availability of camper parking
- web page: "www.gostisce-kimovec.com"
- phone: "05 458 654"
- working hours: "mo-fr: 8:00-18:00"
- working hours comment: "Always opened"
- accessibility: how can one visit the attraction "(car, walk, bike, boat, bus)"
- keywords: few keywords that relate most to the attraction

According to the data analysis, the data categories are presented in Table 1.

Table 1: Category counts for the attraction data

Category	Count
Accommodation	2040
Adrenaline sports	63
Casinos	59
Cities	649
Culture	2337
Cycling and biking	326
Food and wine	1857
Hiking	142
Nature	703
Spas and health resorts	287
Sports	72
Water activities	208
Winter sports	55

There are additional 103 subcategories that further classify each attraction datum or point-of-interest, however, they are not listed here due to the space reasons.

4. GEOGRAPHICAL DATA

The database of Geographical data is composed of two parts; geographical data and statistical regions. Geographical data contains geographical data for Slovenia such as lakes, rivers, caves etc. The data related to the water bodies was obtained from the "eVode portal"[17] (eng. eWaters), while the data regarding other natural bodies was retrieved from the ARSO Geoportal[18].

Water bodies data contains the information in the GeoJSON format[19]:

1. 48 bathing areas.
2. The Slovenian coast.
3. 20 lakes and larger bodies of water.
4. 165 rivers.

Additional data on natural heritage also contained the data in the GeoJSON format:

1. 17 protected areas not included into Natura 2000.
2. 307 areas of ecologic importance.
3. 10.730 caves with descriptions added.
4. 357 Natura 2000 areas.
5. 2.657 items from the registry of natural heritage.
6. 517 additional protected areas.

The geographic information data on statistical regions and settlements has been acquired from GURS (Slovenian Geodetic administration) portal e-Surveying data[20]. All the data was obtained in the GeoJSON format and it included: Boundaries of *12 statistical regions*; Boundaries of *6037 settlements*.

5. ROUTING SERVICE

There are several routing services available on the Internet, the most popular being the Google maps[21]. While being practical the subscription services are not cost effective and not in line with the project goals. We have therefore looked into open source solutions available. The most popular open source solution identified was the Open Street Routing Machine [2] (OSRM).

We have downloaded the map data[22] and combined the Austrian and Slovenian maps into one file using the recommended `osmconvert`[23] tool. Then we processed the data according to the official OSRM instructions[24]. The authors of the OSRM tool also provide a Docker[25] image that can be used for the processing of the maps data and for serving the routing back-end. We have utilized the OSRM docker image in order to process and deploy three distinct routing services: for walking, cycling and car riding. This enables the service to recommend routes to the user based on his or her preferred way of traveling.

Services are currently available through the API calls, for instance, when the service requires a route from point A to point B, it issues an API call to:

```
http://docker-e9.ijs.si:5007/route/v1/driving/LON-A,
LAT-A;LON-B,LAT-B?steps=false
```

[16] `https://www.tripadvisor.com/`

[17] `http://www.evode.gov.si/sl/vodni-kataster/zbirka-vode/zbirka-podatkov-o-povrsinskih-vodah/hidrografija/`

[18] `https://gis.arso.gov.si/geoportal/catalog/main/home.page`

[19] `https://tools.ietf.org/html/rfc7946`

[20] `https://egp.gu.gov.si/egp/`

[21] `https://cloud.google.com/maps-platform/`

[22] `http://download.geofabrik.de/europe.html`

[23] `https://wiki.openstreetmap.org/wiki/Osmconvert`

[24] `https://github.com/Project-OSRM/osrm-backend/wiki/Docker-Recipes`

[25] `https://www.docker.com/`

The service returns a JSON response with the most important objects: route specifications in the form of an encoded polyline; route distance in meters; and route travel time in seconds. Additional information about the API can be obtained on the official OSRM website[26].

6. DATA SERVICES IN AS-IT-IC

The described data and data services will enable the AS-IT-IC services to provide the functionality as required by the project.

Th attractions database enables: the virtual assistant to search for points-of-interest that best match the users query, recognize points-of-interest entities, and fetch information about the attraction; the tour planning service to take into account the attractions locations and provide recommendations to the user based on the attraction category, subcategory, location and similarity to other attractions based on the attraction description; the communication platform to present the data about attractions through the familiar user interface.

The geographical data enables: the virtual assistant to recognized geographical entities, search using the geographical position qualifiers (e.g. "cultural heritage in the Ljubljana city"); the tour planning service to enable recommendation based on the exact location and geographical area boundaries.

The routing service enables: the virtual assistant to take into account the tourist travel options (e.g. "show me natural heritage sites that I can reach in one hour by a bicycle"); the tour planning service to calculate optimal travel plan (since it takes into account the geographical position of the attractions on the itinerary and the transport option chosen by the tourist) and to provide a preview of the trip on a map.

7. CONCLUSION

In the paper we described the data sources, procurement, structure and types of data made available for the AS-IT-IC platform. Additionally we provided a short description of services, which are possible due to the data availability.

The main problem with the tourism-related data procurement is the unavailability of data in a structured, easily accessible format. The portal "Odprti Podatki Slovenije"[27] (eng. Slovenian Open Data) for instance is a good start, however, there are still problems with the discoverability, data formats and data availability. Several good data sources were identified only after weeks of searching for relevant data over the Internet. In order to obtain relevant information users with non-commercial intent can still use web-scrapping in order to obtain the desired data, add the authorship notice and link to the original data source, however, this leads to fragmentation of data structures, additional stress on Internet bandwidth and non-optimal solutions to keep the data updated. Additional problem is the unavailability of data for the use in public interest. This is the case with the registry of cultural heritage.

In the future we plan to integrate additional data sources into the AS-IT-IC databases – by performing data-fusion procedures we will merge the data into a single, richer database. Additionally, we will try to provide the data services to third party developers that now have to go through the same procedure as we did, in order to obtain similar data. Open data and services was one of the main project goals from the start, since we want to improve the tourist experience not only directly but also indirectly by providing services that will enable third party developers to come up with their own innovative solutions.

[26] https://github.com/Project-OSRM/osrm-backend/blob/master/docs/http.md

[27] https://podatki.gov.si/

8. ACKNOWLEDGMENTS

The work was co-funded by Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020, project AS-IT-IC.

9. REFERENCES

[1] B. Cvetković, H. Gjoreski, V. Janko, B. Kaluža, A. Gradišek, M. Luštrek, I. Jurinčič, A. Gosar, S. Kerma, and G. Balažič. e-turist: An intelligent personalised trip guide. *Informatica*, 40(4):447, 2016.

[2] D. Luxen and C. Vetter. Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 513–516, New York, NY, USA, 2011. ACM.

Content API - A Cloud-based Data Source for the AS-IT-IC Platform

Oliver A. Tazl
Institute of Software Technology
Graz University of Technology
Inffeldgasse 16b/II, 8010 Graz, Austria
oliver.tazl@ist.tugraz.at

Franz Wotawa
Institute of Software Technology
Graz University of Technology
Inffeldgasse 16b/II, 8010 Graz, Austria
wotawa@ist.tugraz.at

ABSTRACT
This paper introduces the design and implementation of an element of a microservice application for supporting a modern application for tourist information. Therefore, we introduce the Content Application Pogramming Interface (API) system, a microservice, which collects tourism relevant data from multiple sources and provides it to serveral services within the Austrian-Slovenian Intelligent Turist Information Center (AS-IT-IC) plattform in turn. Content API is built using modern technologies and frameworks, like Docker[1], Spring [2] or Vaadin [3].

Categories and Subject Descriptors
H.2 [**Database Management**]: Systems; I.2 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*; K.6 [**Management of Computing and Information Systems**]: Software Management—*Software development*

General Terms
Database

Keywords
User interface, data aquisition, REST API

1. INTRODUCTION
The AS-IT-IC platform[4] was introduced in order to enable tourists and tourism information provider to interact and collaborate with each other. Several data sources exist from which the AS-IT-IC partners can retrieve information of sights, natural heritage as well as other events and tourism offers in the program area. The information from these databases is not stored in an unified format, so in order to integrate this information into the AS-IT-IC platform it is nesseary to integrate those data into a fitting format. The content subsystem, called Content API, integrates this data and provides it to the AS-IT-IC ecosystem. The system collects information from other web sources in order to combine and complete the information for a specific offering.

The core items of these databases are:

1. Sights
2. Tourism companies (e.g. hotels, restaurants,...)
3. Natural Heritage (e.g. rivers, mountains,...)
4. Events
5. Cultural heritage

The goal of the content subsystem is to integrate various information sources about tourism offers into a single database. This database is meant to be updateable from these sources, like Google or other tourim websites. The integrated user interface also allows human-computer interaction in order to update and review the integrated data. This allows a collaborative approach that enables users to add new information pieces to our database and helps to keep the database up-to-date as well.

The remainer of the paper is organised as follows: In the next section, we present the chosen design of the system. In Section 3, we get into details of the architecture. Afterwards, Section 4 provides an overview about the implementation and the deployment of the system. Finally, we show some related research and conclude the paper.

2. DESIGN & ARCHITECTURE
To fit the domain specific data requirements, we figured out a data structure that allows to represent the data needed in our platform, as shown in Figure 2.

The Core Service Layer contains the main functionalities of the web service. The layer hosts the data storage logic, the data aquisition along with the merger functionalities. They are encapsulated to ensure easy extensibility for new data sources to be included. We call these modules within the initial startup sequence as well as on demand.

The possiblity to merge data from different sources together is crucial in order to present a good and up-to-date data quality. This functionality is contained within the Core Service Layer.

It is also possible to integrate new data sources in order to add new information aspects into the database.

The system is integrated into the architecture of the AS-IT-IC platform as discussed in [1].

[1]see `docker.com`
[2]see `spring.io`
[3]see `vaadin.com`
[4]see `as-it-ic.ijs.si`

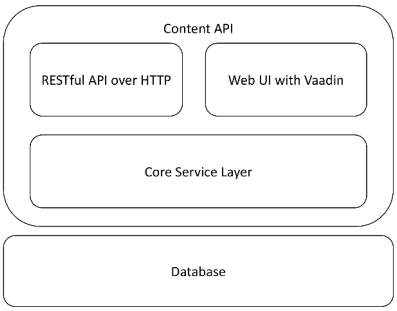

Figure 1: Layer of the Content API

2.1 Update Process

The update process collects several web data sources and integrates it into the database in an automated manner. Every data source needs a specific update routine that is implemented to integrate the collectable data. These routines compare the recently collected data with the stored data stored and update the information in case that there is any change provided by the external source. The process can be scheduled automatically or triggered manually.

2.2 User Interface

The web-based user interface (UI) allows the community to contribute information to the platform. It is possible to add new information items, such as sights or tourism offers, to the database. Reviewing existing items and updating them represents also a very important opportunities for the user community in order to allow a wiki-like contribution model.

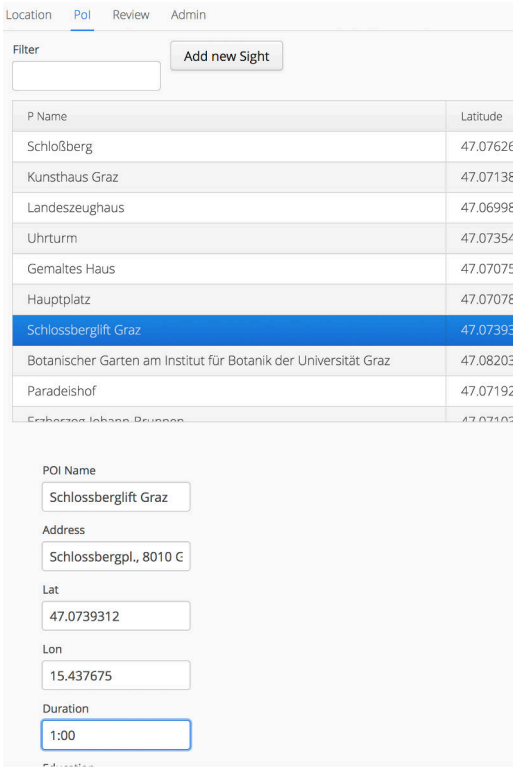

Figure 2: Screenshot of the web-based UI

In Figure 2, we show the Vaadin based UI containing several subforms to interact with the different entries from the database.

2.3 REST Interface

The Representational State Transfer (REST) interface can be used to implement operations like create, read, update and delete (CRUD) the information of the content subsystem.

There exist endpoints that represent the information in the database and allow to access them via JSON objects, i.e.:

- Location
- Points of interest (POI)
- Equipment
- Transportation

We use these endpoints in order to access and modify the data in the system via HTTP-JSON calls.

3. IMPLEMENTATION & DEPLOYMENT

We use Docker within the implementation and the deployment step. Docker Compose is a tool for defining and running multi-container Docker applications. In this project, Docker Compose is used on the developer machine to set up a testing environment, which configurations corresponds to the productive pendant.

4. CONCLUSION

In this paper, we presented the Content API, the content microservice and database of the AS-IT-IC platform. Here we focussed on the architecture and design of the system, as well as the implementation and the deployment. We also highlighted the use of the web-based user interface and the REST API as interaction possiblies. Finally we described the automated deployment using an application container technology as well as continous integration and deployment tools.

5. ACKNOWLEDGMENTS

Research presented in this paper was carried out as part of the AS-IT-IC project that is co-financed by the Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020, European Union, European Regional Development Fund.

6. REFERENCES

[1] J. Zupancic, G. Grasselli, A. Tavcar, and M. Gams. Virtual Assistants for the Austrian-Slovenian Intelligent Tourist-Information Center. In *Proceedings of the 20th International Multiconference INFORMATION SOCIETY - IS 2017, Volume E*, 2017.

e-Tourist 2.0: an Adaptation of the e-Tourist for the AS-IT-IC Project

Gregor Grasselli
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Jamova 39
1000 Ljubljana, Slovenia
gregor.grasselli@ijs.si

ABSTRACT

This article presents a new version of the `e-Tourist` system called `e-Tourist2.0`. The `e-Tourist2.0` system is developed within the Austrian-Slovenian Intelligent Tourist-Information Center (AS-IT-IC) project, accepted in the Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020 call. The new `e-Tourist2.0` system brings a number of additional features with respect to the `e-Tourist`, such as location aware search, an implicit recommendation engine and a more interactive interface for trip planning. In the paper we briefly explain the need for a new system, and present the architecture and functionality of the `e-Tourist2.0`.

Keywords

trip planner; AS-IT-IC project; tourism; attractions; tourism database; recommendation engine; location aware search

1. INTRODUCTION

Trip planners are applications that require a user to input a number of predetermined parameters and then respond by offering the user a trip plan that can later be modified. The Austrian-Slovenian Intelligent Tourist-Information Center (AS-IT-IC) platform provides a trip planner with a chat based user interface. It is built from three major components, each contributing to the final user experience. They are:

1. the front-end provided by a slightly modified version of `Rocket.Chat`[1]
2. `e-Tourist2.0` that fulfils the role of the back-end
3. a conversational program (a bot) that takes user inputs, be they free form text input or button clicks, from `Rocket.Chat` and generates responses, using data acquired from `e-Tourist2.0` when necessary

In its function as the back-end for the AS-IT-IC platform, `e-Tourist2.0` needs to support an assortment of queries over textual and geographical data (i.e. the ability to provide an answer to questions like "Which 5 points of interest are most similar to the Bled castle based on their descriptions?" or "List all castles in the Gorenjska region"), a recommendation engine, that uses implicit data about user interest, as well as trip planning functionality. It also needs to provide the pieces of the user interface that are missing from `Rocket.Chat`.

Initially we believed that this role could be filled by the existing `e-Tourist` after some light modifications. However, a more thorough inspection of its capabilities and architecture convinced us that extensive adjustments to the code base are required. Additionally some of the functionality originally implemented in the application code is now supported by the database itself which would lead to the deletion of several lines of code. The resulting conclusion was to write a new system using the latest tools and design patterns, while reusing the old code as much as possible.

This rest of this article describes the need for a new system in Section 2, followed by a presentation of the new system Section 3, where we present the features available through the application programming interface (API) of `e-Tourist2.0`. In section 4 we present the system architecture and 5 concludes the paper.

2. WHY A NEW SYSTEM

The required back-end functionality includes:

1. Trip planning
2. Location aware search
3. Recommendation system

Initially `e-Tourist` was designed to plan trips visiting the coastal region of Slovenia.It was provided with a relatively small list of tourist attractions in both regions and is capable of planning a trip to either of them. The user can specify a number of constraints like the exact start of their trip, and whether they want to eat along the way and the `e-Tourist` produces a nice trip recommendation. Based on how well this works, we were convinced that all that needed doing was to add additional data with points for other regions of Slovenia and Austria, spruce up the API so the bot would be able to use the system from outside, and we would have our back end. This was presented in [1].

We are porting some of the trip planning capabilities of `e-Tourist` to `e-Tourist2.0`, and in this respect `e-Tourist` mostly meets our requirements, except for not being RESTful, that is calls to the API needed session specific state with

[1] `https://rocket.chat/`

them in order for it to return the correct result. Figure 1 shows what the original user interface for trip planning looks like. It shows the trip on a map and adds controls to add or remove points from it.

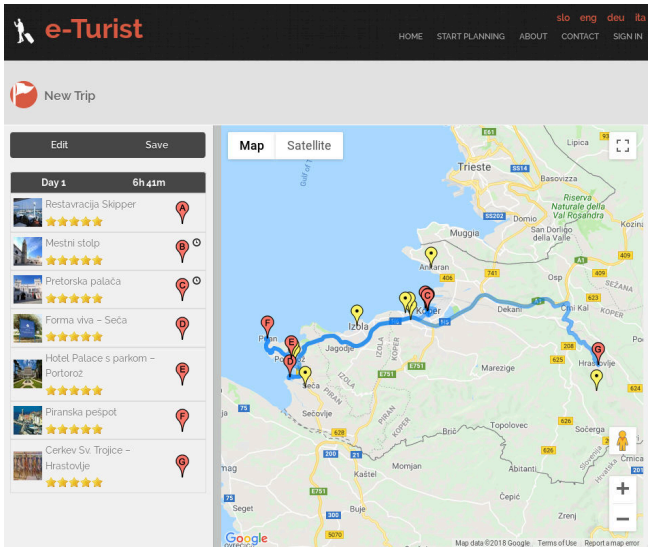

Figure 1: `e-Tourist` trip planning interface

As it turns out, the `e-Tourist` database stores each point of interest in a tree-based data structure, based on its location – a point-of interest belongs to (or is a child in a tree-based structure of) a settlement that in turn belongs to a region. This proves to be a very brittle way for storing larger amounts of points where some of the settlements have the same names and some areas of interest belong to multiple settlements and even regions. Also the `e-Tourist` does not make it possible to ask queries based on actual distance, like "Which restaurants are within 1km of lake Bohinj?". Therefore, a rewrite of the location aware search aspects of the application would be required.

In terms of the recommendation system `e-Tourist` requires users to rate the points of interest on a 1 to 5 scale or to input a lot of data about themselves like their education, age, gender so forth. It also requires data about points of interest that would be hard to acquire in an automated way, like which age groups a point of interest is most appealing to or the level of education of the people who most enjoy their visit to the point of interest. Since the implementation of the `e-Tourist` several modules have been developed that implement various recommendation algorithms and were released as open source software[2]. For reasons of stability, security and ease of maintenance we opted for one of those instead of using our own recommendation engine implementation.

According to the code analysis we reached the following conclusions:

1. A recommendation system needs to be rewritten to also include implicit information about user interaction with the system.
2. Location aware search needs to be included.
3. Proper database migrations have to be implemented.
4. Database architecture needs to be reworked.
5. The way travel time was computed was very storage and time intensive.
6. `e-Tourist` is not REST-ful.
7. We need a more stateless way of user authentication.
8. Code base has to be updated to the newest framework versions.

Besides the tree based structure mentioned in the introduction, that becomes unnecessary when storing geographic data correctly, the database schema of `e-Tourist` also failed to make use of several data structures offered by the PostgreSQL database engine that would make some queries simpler and more performant. The main problem with the database however was the lack of an initial migration that would create the relations required in the application. The authors assumed that future developers would use a copy of a pre created database for that, an approach that made creating a new development environment as well as deployment an unnecessarily complicated task.

Because `e-Tourist` does not comply with REST best practices, it is quite hard to use programmatically, which is essential for us since most of it is going to be used by the `Rocket.Chat` bot, a program. The latter also means that session cookies are not a viable way for us to check user credentials and a different way of authenticating calls to the API is needed.

Finally the code used some features of an older version of the framework that were discarded in the newer ones. In order to be able to guarantee long term support for the project we decided that a newer framework version with longer support was required.

Tallying up all of the above we figured it would take more work to make the necessary changes to the old system then writing a new one from scratch with an eye out for code reuse whenever possible.

3. FEATURES

In addition to the features already mentioned at the start of the Section 2, the new `e-Tourist2.0` supports:

1. Finding similar points of interest
2. A user interface that allows quickly adding points to a trip plan and deleting them from it by using a map
3. Exporting trips to Google Maps
4. Full text search for points of interest based on their descriptions, and limited by their location
5. A recommendation engine that uses implicit data about user's interaction with the system to find points that might interest a particular user based on the user's history

[2] `https://github.com/benfred/implicit/`, `http://surpriselib.com/`

We will continue with a discussion of the features, what they do and why we need them.

3.1 Trip planning

Trip planning means that given a list points of interest a potential tourist wants to visit, the system can plan a route on a map that visits all the points listed by the tourist. The new `e-Tourist2.0` does this by using the `open source routing machine` [2] which uses open street map[3]data to plan the route. In addition to being able to plan the route, `e-Tourist2.0` is also capable of exporting that route to a Google Maps link, so that users may conveniently follow along using the Google Maps app on their devices. The trip planning user interface will also display potentially interesting attractions near the route already chosen so that a tourist may quickly add them to their trip plan.

3.2 Recommendation

Recommendations are a way to present the users with more of the relevant content based on their interests as shown through their history of using the system. This will also allow a registered user to simply ask the system "What is interesting in Koroška?", as well as provide additional suggestions along the planned path.

Another use of a recommendation system is to compare attractions based on user behaviour. This allows us to *find similar attractions* to the one picked by the user. Another way to find similar attractions is by comparing their descriptions and `e-Tourist2.0` uses both of them. This feature enables the users to quickly narrow in on what they want to see or to just explore their options more conveniently.

3.3 Full text search

Full text search is a way to quickly search a large database of documents for the ones containing given words or phrases. In `e-Tourist2.0` this is coupled with some sentence analysis that attempts to produce more relevant results. It searches through attraction descriptions in German, English and Slovenian.

All attractions and geographical features in `e-Tourist2.0` carry complete information about their location. Complete in the sense that regions and settlements are saved as polygons describing their borders, rivers are saved as lines describing their entire flow and so on. This enables all sorts of *location based search* queries, as well as constraining other queries to certain locations. Examples of such queries are "List all museums in Slovenj Gradec" and "Find all attractions similar to the Bled castle near Klagenfurt".

The most important feature from the standpoint of the tourist providers is the *administration interface*, which will allow them to add new points of interest to the system and make corrections to the data on those already there. To request administration access, a tourism provider should fill in a form available on the `e-Tourist2.0` website[4].

[3] `https://www.openstreetmap.org/`

[4] `https://e-turist.docker-e9.ijs.si/tourism-provider/request-admin`

4. `e-Tourist2.0` ARCHITECTURE

In order to make it easier to port code from `e-Tourist` to `e-Tourist2.0` we chose to implement `e-Tourist2.0` in the same framework as `e-Tourist`, *Django*[5]. Besides allowing us to more easily port code from the old project it also comes with a built in administration interface that made implementing the user interface for tourist information providers a lot easier, since we just had to customise the one provided with the framework.

Our data storage is provided by the *PostgreSQL*[6] database. By using the *Postgis*[7] extension we were able to save geographical data and use it for several spatial queries. It also supports full text search. In order to enable support for the Slovenian language we provided some language specific configuration and files, while the German and English languages are supported out-of-the-box.

5. CONCLUSIONS

We have presented a short description of the new `e-Tourist2.0`, describing the need for a new trip planner implementation, its features and its architecture. While some of the `e-Tourist` code base was reused, the new `e-Tourist2.0` is mostly a new program. Most of the features presented here are already fully functional, however, the program is not yet entirely complete and changes to existing features, or additional features are possible in the future.

6. ACKNOWLEDGMENTS

We thank students Tadej Petrič, Aljaž Glavač and Martin Češnovar, who contributed to `e-Tourist2.0` development. The work was co-funded by Cooperation Programme Interreg V-A Slovenia-Austria 2014-2020, project AS-IT-IC.

7. REFERENCES

[1] B. Cvetković, H. Gjoreski, V. Janko, B. Kaluža, A. Gradišek, M. Luštrek, I. Jurinčič, A. Gosar, S. Kerma, and G. Balažič. e-turist: An intelligent personalised trip guide. *Informatica*, 40(4):447, 2016.

[2] D. Luxen and C. Vetter. Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 513–516, New York, NY, USA, 2011. ACM.

[5] `https://www.djangoproject.com/`

[6] `https://www.postgresql.org/`

[7] `https://postgis.net/`

Planning-based Security Testing for Chatbots

Josip Bozic
Institute for Software Technology
Graz University of Technology
A-8010 Graz, Austria
jbozic@ist.tugraz.at

Franz Wotawa
Institute for Software Technology
Graz University of Technology
A-8010 Graz, Austria
wotawa@ist.tugraz.at

ABSTRACT

Chatbots are of increasing importance in modern day communication between users and industrial applications. For example, providers of financial and medical institutions make use of intelligent agents in order to provide accessibility on a 24/7 basis. The human-like communication, often realized in an entertaining way, represents one of these advantages that chatbots offer. Eventually, chatbots make use of artificial intelligence methods in order to learn from past communication interactions, to provide better and more personalized responses. Often chatbots are deployed as part of web applications. As a consequence, this makes them vulnerable to typical security attacks on websites. Planning-based techniques can help to identify security leaks for common attack scenarios in a smart way. In this paper, we present such an approach that relies on artificial intelligence planning for security testing of chatbots that are accessible using web applications.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: General—*Security and protection*; D.2 [**Software Engineering**]: Testing and Debugging; I.2 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents*

General Terms

Theory, Security

Keywords

Planning, security testing, chatbots

1. INTRODUCTION

ELIZA [21] is first known computer program that communicated in a natural language with a person, which was developed in 1966. Over the years, further improvements have been added to such similar applications, called chatbots [9, 20]. Chatbots are deployed in a stand-alone or online fashion, i.e., as part of websites in form of virtual assistants. Such programs offer the advantage of human-like communication and as well an almost unlimited accessibility. Deploying chatbots offer financial advantages for service providers as well. They can be used in order to respond to customers' inquiries, e.g., provide information about certain goods or services without the need of human intervention.

In contrast to the initial versions of chatbots, where the programs usually responded to an inquiry according to stored natural language patterns, intelligent agents can be trained to learn from the communication with clients by relying on machine learning. In the long term, these smart chatbots refine their responses and provide better answers over time. Advanced chatbots offer the client the possibility to authenticate herself or himself, thus personalizing the connection between machine and person. Here the chatbot stores potentially sensible data from the user. In this case, the chatbot must guarantee user authenticity and data integrity. The failure to fulfill such promises might result in personal and financial consequences for the user.

This requirement and the fact that chatbots can be deployed as part of web applications leads to security issues, since web applications are known to be vulnerable to several hacking attacks. Vulnerabilities like SQL injections (SQLI) and cross-site scripting (XSS) are still very common [4], despite security measures and new security testing approaches. It is interesting to know that the security issue in the context of chatbots have almost not been considered before.

Automated planning and scheduling, or simply planning, is a branch of artificial intelligence that was initially used in robotics and intelligent agents [17]. There planning is used for guiding an agent by responding to encountered conditions. Some approaches have applied planning to security testing in specific domains as well [14, 16]. In order to contribute to the mentioned security issues, we introduce an automated planning-based security testing approach for chatbots in this paper.

The paper is organized as follows. Section 2 introduces planning to security testing of chatbots. Then, Section 3 gives an overview about a test execution framework. Finally, Section 4 concludes the work and discusses potential goals for the future.

2. PLANNING FOR CHATBOTS

Planning has been already used in security testing to a small degree. The authors of this paper have applied planning for testing of web applications (e.g. [7]) and the TLS protocol [6]. However, the application to chatbots is novel in this sense. The main motivation behind using planning for testing is the fact that attacks against applications can be depicted in form of a sequence of steps that could be applied against every program. Actually, a plan acts as a blueprint for an attack. In this paper, we applied the planning speci-

fication from our previous work [5] but use it for identifying vulnerabilities in the case of chatbots.

In general, the planning problem was initially given in [11] and can be defined as follows.

Definition 1. *A planning problem is defined by the tuple* (I, G, A). *A state is defined by a set of first order logic predicates. I represents the initial state, the goal state is G and the set of actions is given by A. Every action* $a \in A$ *comprises a precondition and an effect. The functions* $pre(a)$ *and* $eff(a)$ *connect the individual preconditions and effects, respectively.*

If the precondition $pre(a)$ of an action a is satisfied the current state S, then this action will be selected for the solution of the planning problem. The execution of this action will lead to a new state S', namely $S \xrightarrow{a} S'$. This procedure will continue until the execution reaches the goal state G, i.e. fulfills its preconditions. The program that reads the planning specification and searches for a solution according to a planning algorithm is called a planner.

Definition 2. *The solution for the planning problem* (I, G, A) *is returned in form of a plan, which is given by a sequence of actions* $\langle a_1, \ldots, a_n \rangle$ *such that* $I \xrightarrow{a_1} S_1 \xrightarrow{a_2} \ldots \xrightarrow{a_{n-1}} S_{n-1} \xrightarrow{a_n} G$.

The planning problem is implemented in the Planning Domain Definition Language (PDDL) [15]. Here, two specifications have to be provided:

- Domain definition: Data that is present for every problem definition.
- Problem definition: Data that defines one specific problem.

PDDL supports a type-object hierarchy of data and uses it in conjunction with first-order logic predicates. Every object correspondents to a specific type, which relates to variables and classes in object-oriented programming, respectively. The individual action definitions are built from parameters and pre- and postconditions, which are defined with one or more predicates. For example, an excerpt from the the domain definition for chatbot testing is depicted below.

```
(define (domain chatdomain)
 (:requirements
    :strips :typing :equality :fluents
    :adl)
 (:types
    status address server status-si
    status-se type expect result username
    password action method integer sqli
    xssi response script)
 (:constants
    init - status two - status-si sqli
    xss - type get post head - method
    username - username
    password - password)
 (:predicates
    (inInitial ?x)
    (inAddressed ?x)
    (inSentReq ?x)
    (inRecReq ?x)
    (inSQLI ?x)
    (inXSS ?x)
    (inAttackedSQL ?x)
    (inAttackedXSS ?x)
    (inFound ?x)
    (Empty ?url)
    (FoundScript ?script - script ?resp -
       response))
 (:action Start
    :parameters(
      ?x - status
      ?url - address)
    :precondition (and
      (inInitial ?x)
      (not (Empty ?url)))
    :effect (and
      (inAddressed ?x)
      (not (inInitial ?x))))
 (:action SendReq
    :parameters(
      ?x - status
      ?se - status-se
      ?si - status-si)
    :precondition (inAddressed ?x)
    :effect (and
      (inSentReq ?x)
      (not (inAddressed ?x))
      (assign (sent ?se) 1)
      (statusinit two)))
 (:action ParseRespXSSCheck
    :parameters(
      ?x - status
      ?script - script
      ?resp - response)
    :precondition (and
      (inRecRespRXSS ?x)
      (not(FoundScript ?script ?resp)))
    :effect (and
      (FoundScript ?script ?resp)
      (inFound ?x)
      (not(inRecRespRXSS ?x)))))
```

Domain description in PDDL

As can be seen, the PDDL definition encompasses, among others, types, predicates and actions. Again, the individual action definitions make use of the predicates and apply parameters in order to check if the predicate is valid. The specification uses the parameter x to denote the current state of execution. As mentioned, the above domain, due to space reasons, does not include our entire specification. On the other hand, the problem definition is defined as follows.

```
(define (problem chatproblem)
 (:domain chatdomain)
 (:objects
    x - status
    s - server
    url - address
    m - method
    exp - expect
    script - script
    resp - response)
(:init
   (inInitial x)
   (not (Empty url))
   (Method post)
   (Response resp)
   (not (FoundScript script resp))
(:goal
   (inFinal x)))
```

Problem description in PDDL

The problem definition comprehends the definition of objects and, most important, the initial state. This state represents the starting point from which the planner will start the search. Modification of the initial state will result in the generation of a different plan. If no plan can be generated, then the planner returns an error. A generated plan looks as follows:

```
0: START X URL
1: SENDREQ X SE SI
2: RECREQ X SI
3: PARSE X M USERNAME PASSWORD TYPE
4: CHOOSEXSS X TYPE
5: ATTACKXSS X XSSI M UN PW
6: PARSERESPXSS X SCRIPT RESP
7: PARSERESPXSSCHECK X SCRIPT RESP
8: FINISH X
```

Generated plan for XSS

As mentioned before, the plan is represented by a sequence of actions and corresponding parameters picked from the domain definition. In our case, we used the planner `Metric-FF` [3]. Now, this sequence of steps acts as an abstract test case that will be executed by an executioner against the system under test (SUT). The purpose of the plan is to guide the test execution process that, in the best case, will lead to the detection of a vulnerability. The main idea here is to apply this plan against every chatbot that corresponds to the scenario as described in the next section.

3. SECURITY TESTING OF CHATBOTS

Security plays a major role for every software system. Failure to fulfill security requirements might lead to severe private, financial and reputation consequences. For this reason, programs have to be tested during the development lifecycle and after release of the software. Until now, many manual and automated approaches have been introduced in order to test for vulnerabilities. For example, model-based approaches usually rely on a model of the SUT [19, 18], whereas fuzzing and combinatorial testing put emphasis on test case generation from a pentesting aspect [10, 13].

SQLI and XSS represent two common vulnerabilities for many years and need further addressing for this reason. Detailed information about these two vulnerabilities can be found in [8] and [12], respectively. Chatbots, as already mentioned, when deployed as part of a web application, inherit the vulnerabilities as well. A scenario that depicts the entire planning-based security testing system is depicted in Figure 1.

Attack vectors are malicious input strings that an attacker or tester submits against an application. For XSS, the list of attack vectors consists of JavaScript code, whereas malicious SQL statements are used for SQLI. As already mentioned in Section 2, a generated plan is used as an abstract test case. The reason for this is the fact that PDDL is limited with regard to setting of concrete values for parameters. For this reason, we define a test execution framework, that encompasses, among others, an executioner. This framework is implemented in Java and contains concrete Java methods that correspond to the individual actions from PDDL. More information about this mechanism can be found in [5].

The executioner reads the abstract plan and searches for the concrete counterpart of the individual actions. Then, HTTP requests are created with the help of `HttpClient` [1] and instantiated with an attack vector. Then, the attack is carried out in form of the HTTP request. The SUT is a deployed chatbot that encompasses a database. The chatbot has a user input field, e.g. an HTML element for textual inputs, that represents the target for the executioner. The test oracles specify what test output is expected and provide the final test verdict. We rely on our previously implemented oracles from [7] for this purpose. After an attack, a parser reads the response from the SUT. It searches the HTML structure for critical vulnerability indicators, as specified in the oracles. In this scenario, we rely on `jsoup` [2]. The testing process continues as long as the plan has been executed for every attack vector.

4. CONCLUSION

In this paper, we introduced a security testing approach for chatbots that relies on planning. After manually defining the specification and generating the plan, a test execution implementation executes the plan in an automated manner. The approach is meant for testing of chatbots against two common web vulnerabilities, namely XSS and SQLI. Under the assumption that chatbots will play a major role in the future, it remains important to address this issue. In the future, the proposed security testing approach will be evaluated against real-world applications and compared to other testing techniques.

Acknowledgments

The research presented in the paper has been funded in part by the Cooperation Programme Interreg V-A Slovenia-Austria under the project AS-IT-IC (Austrian-Slovene Intelligent Tourist Information Center).

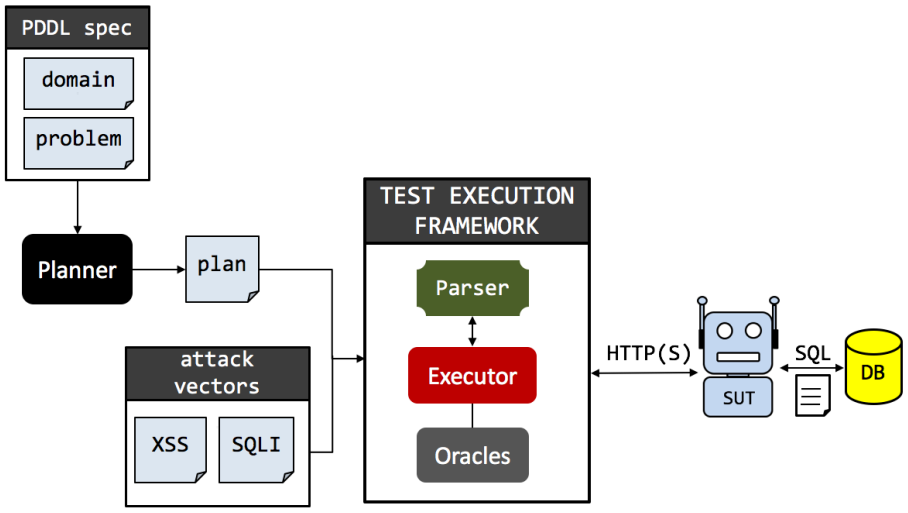

Figure 1: Planning-based Chatbot Security Testing

5. REFERENCES

[1] Apache HttpComponents - HttpClient. `https://hc.apache.org/httpcomponents-client-ga/`. Accessed: 2018-09-06.

[2] jsoup: Java HTML Parser. `https://jsoup.org/`. Accessed: 2018-02-02.

[3] Metric-FF. `https://fai.cs.uni-saarland.de/hoffmann/metric-ff.html`. Accessed: 2018-09-06.

[4] OWASP Top Ten Project. `https://www.owasp.org/index.php/Category:OWASP_Top_Ten_Project`. Accessed: 2018-01-31.

[5] J. Bozic and F. Wotawa. Plan It! Automated Security Testing Based on Planning. In *Proceedings of the 26th IFIP International Conference on Testing Software and Systems (ICTSS'14)*, pages 48–62, September 2014.

[6] J. Bozic and F. Wotawa. Planning the Attack! Or How to use AI in Security Testing? In *Proceedings of First International Workshop on AI in Security (IWAIse)*, 2017.

[7] J. Bozic and F. Wotawa. Planning-based Testing of Web Applications. In *Proceedings of the 13th International Workshop on Automation of Software Test (AST'18)*, 2018.

[8] J. Clarke, R. M. Alvarez, D. Hartley, J. Hemler, A. Kornbrust, H. Meer, G. OLeary-Steele, A. Revelli, M. Slaviero, and D. Stuttard. *SQL Injection Attacks and Defense*. Syngress, Syngress Publishing, Inc. Elsevier, Inc., 30 Corporate Drive Burlington, MA 01803, 2009.

[9] K. Colby. Artificial Paranoia: A Computer Simulation of Paranoid Process. In *Pergamon Press, New York*, 1975.

[10] F. Duchene, S. Rawat, J.-L. Richier, and R. Groz. KameleonFuzz: Evolutionary Fuzzing for Black-Box XSS Detection. In *CODASPY*, pages 37–48. ACM, 2014.

[11] R. E. Fikes and N. J. Nilsson. STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving. In *Artificial Intelligence*, pages 189–208, 1971.

[12] S. Fogie, J. Grossman, R. Hansen, A. Rager, and P. D. Petkov. *XSS Attacks: Cross Site Scripting Exploits and Defense*. Syngress, 2007.

[13] D. Kuhn, D. Wallace, and A. Gallo. Software Fault Interactions and Implications for Software Testing. In *IEEE Transactions on Software Engineering 30 (6)*, 2004.

[14] A. Leitner and R. Bloem. Automatic Testing through Planning. Technical report, Technische Universität Graz, Institute for Software Technology, 2005.

[15] D. McDermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins. PDDL - The Planning Domain Definition Language. In *The AIPS-98 Planning Competition Comitee*, 1998.

[16] A. M. Memon, M. E. Pollack, and M. L. Soffa. A Planning-based Approach to GUI Testing. In *Proceedings of the 13th International Software / Internet Quality Week (QW'00)*, 2000.

[17] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentic Hall, 1995.

[18] I. Schieferdecker, J. Grossmann, and M. Schneider. Model-Based Security Testing. In *Proceedings of the Model-Based Testing Workshop at ETAPS 2012. EPTCS*, pages 1–12, 2012.

[19] M. Utting and B. Legeard. *Practical Model-Based Testing - A Tools Approach*. Morgan Kaufmann Publishers Inc., 2006.

[20] R. S. Wallace. The Anatomy of A.L.I.C.E. In *ALICE A.I. Foundation*, 2004.

[21] J. Weizenbaum. ELIZA–A Computer Program For the Study of Natural Language Communication Between Man and Machine. In *Communications of the ACM Volume 9, Number 1 (January 1966)*, 1966.

Indeks avtorjev / Author index

Konferenca / Conference

Uredila / Edited by

Delavnica AS-IT-IC / AS-IT-IC Workshop

Matjaž Gams, Jernej Zupančič