The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman
Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members
Rafael Bello Perez, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at http://www.springer.com/series/11156
Proceedings of the Future Technologies Conference (FTC) 2018

Volume 2
Editor’s Preface

Future Technologies Conference (FTC) 2018 was held on November 13–14, 2018, in Vancouver at the Marriott Pinnacle Downtown Hotel, with sweeping views of the coastal mountains, Coal Harbour, and Vancouver’s city skyline. The city of Vancouver is considered as one of the most beautiful cities in the world.

With great privilege, we present the Proceedings of FTC 2018 in two volumes to the readers. We hope that you will find it useful, exciting, and inspiring. FTC 2018 aims at producing a bright picture and charming landscape for future technologies by providing a platform to present the best of current systems’ research and practice, emphasizing innovation and quantified experience. The ever-changing scope and rapid development of future technologies create new problems and questions, resulting in the real need for sharing brilliant ideas and stimulating good awareness of this important research field.

Researchers, academics, and technologists from leading universities, research firms, government agencies, and companies from 50+ countries presented the latest research at the forefront of technology and computing. After the double-blind review process, we finally selected 173 full papers including six poster papers to publish.

We would like to express our gratitude and appreciation to all of the reviewers who helped us maintain the high quality of manuscripts included in this conference proceedings. We would also like to extend our thanks to the members of the organizing team for their hard work. We are tremendously grateful for the contributions and support received from authors, participants, keynote speakers, program committee members, session chairs, organizing committee members, steering committee members, and others in their various roles. Their valuable support, suggestions, dedicated commitment, and hard work have made FTC 2018 a success. Finally, we would like to thank the conference’s sponsors and partners: Western Digital, IBM Research, and Nature Electronics.

We believe this event will help further disseminate new ideas and inspire more international collaborations.
We hope that all the participants of FTC 2018 had a wonderful and fruitful time at the conference and that our overseas guests enjoyed their sojourn in Vancouver!

Kind Regards,

Kohei Arai
# Contents

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking and Visualization of Experts for Communication Using Linkedin</td>
<td>1</td>
</tr>
<tr>
<td>Muhammad Iqbal and Mubashar Ahmad</td>
<td></td>
</tr>
<tr>
<td>A Modern Book Archive Geared for Custom Publishing</td>
<td>13</td>
</tr>
<tr>
<td>Liam Borgstrom</td>
<td></td>
</tr>
<tr>
<td>WOWcube Puzzle: A Transreality Object of Mixed Reality</td>
<td>22</td>
</tr>
<tr>
<td>Ilya V. Osipov and Evgeny Nikulchev</td>
<td></td>
</tr>
<tr>
<td>Wetware and the Cyborg Era: The Future of Modifications on the Human Body According to Science Fiction</td>
<td>34</td>
</tr>
<tr>
<td>Bianca Helena Ximenes and Hermano Perrelli de Moura</td>
<td></td>
</tr>
<tr>
<td>Investigating the Impact of Data Volume and Domain Similarity on Transfer Learning Applications</td>
<td>53</td>
</tr>
<tr>
<td>Michael Bernico, Yuntao Li, and Dingchao Zhang</td>
<td></td>
</tr>
<tr>
<td>Teaching Quantum Computing</td>
<td>63</td>
</tr>
<tr>
<td>Lewis Westfall and Avery Leider</td>
<td></td>
</tr>
<tr>
<td>Hareth Zmezm, Hamzah F. Zmezm, Mustafa S. Khalefa, and Hamid Ali Abed Alasadi</td>
<td></td>
</tr>
<tr>
<td>Collaborative Intelligence Analysis for Industrial Control Systems Threat Profiling</td>
<td>94</td>
</tr>
<tr>
<td>Ke Li, Jianzhou You, Hui Wen, Hong Li, and Limin Sun</td>
<td></td>
</tr>
<tr>
<td>Copyright Declaration for Leaked Numeric Data</td>
<td>107</td>
</tr>
<tr>
<td>Che-Wei Lee</td>
<td></td>
</tr>
</tbody>
</table>
Ontology-Based Intelligent Security Framework for Smart Video Surveillance ......................................... 118
Amna Shifa, Mamoona Naveed Asghar, Martin Fleury, and Muhammad Sher Afgan

The Role of Trust and Control in Managing Privacy When Photos and Videos Are Stored or Shared .................................................. 127
Srinivas Madhisetty and Mary-Anne Williams

Cloud Data Security Solution Based on Data Access Classification, Advanced Encryption Standard and Message Authentication Code .... 141
Mohammed Saeed Jawad

Vehicle Communication Using Secrecy Capacity ......................... 158
Na-Young Ahn, Donghoon Lee, and Seong-Jun Oh

A Lightweight DDoS Attack Mitigation System within the ISP Domain Utilising Self-organizing Map ........................................ 173
Ili Ko, Desmond Chambers, and Enda Barrett

Secure Feature Extraction in Computational Vision Using Fully Homomorphic Encryption ................................. 189
Thomas Shortell and Ali Shokoufandeh

Behavioral Detection of Scanning Worm in Cyber Defense ........... 214
Mohammad M. Rasheed and Munadil K. Faaeq

Network Aware Defenses for Intrusion Recognition and Response (NADIR) .......................................................... 226
Nont Assawakomenkool, Yash Patel, and Jonathan Voris

Critical Workload Deployment in Public Clouds with Guaranteed Security Levels and Optimized Resource Usage and Energy Cost ..... 240
Soamar Homsi, Gang Quan, and Laurent Njilla

Malware Forensic Analytics Framework Using Big Data Platform .... 261
Suriayati Chuprat, Aswami Ariffin, Shamsul Sahibuddin, Mohd Naz’ri Mahrin, Firham M. Senan, Noor Azurati Ahmad, Ganthar Narayana, Pritheega Magalingam, Syahid Anuar, and Mohd Zabri Talib

Efficient LFSR Based Distance Bounding Protocol for Contactless EMV Payments .................................................. 275
Trishla Shah and Srinivas Sampalli

Performance Evaluation of a Pseudo-Random Number Generator Against Various Attacks ........................................ 291
Trishla Shah, Srinivas Sampalli, Darshana Upadhyay, and Priyanka Sharma
Contents ix

Detection: Definition of New Model to Reveal Advanced Persistent Threat .................................................. 305
M. Maccari, A. Polzonetti, and M. Sagratella

Web Password Recovery: A Necessary Evil? ................................. 324
Fatma Al Maqbali and Chris J. Mitchell

The Online Identity Detection via Keyboard Dynamics .................. 342
Merylin Monaro, Marta Businaro, Riccardo Spolaor, Qian Qian Li,
Mauro Conti, Luciano Gamberini, and Giuseppe Sartori

Using Kali Linux Security Tools to Create Laboratory Projects for Cybersecurity Education ......................... 358
Ahmad Ghafarian

Financial Risks of the Blockchain Industry:
A Survey of Cyberattacks ......................................................... 368
Aleksandr Lazarenko and Sergey Avdoshin

OBJS: Forking Javascript to OOP Paradigms ............................... 385
Alessandro Rosa

SIMS: A Stakeholder Information Management System for Successful Projects ............................................. 402
Belmir Patkovic, Kassem Saleh, and Paul Manuel

Solving the Linear Ordering Problem via A Memetic Algorithm ...... 421
Jingyu Song, Haidan Zhao, Taoqin Zhou, Ye Tao, and Zhipeng Lü

A Usability Framework for Diabetic Health Applications in South Africa ..................................................... 431
Bulelani Ngamntwini and Liezel Cilliers

Meta-requirement Method Towards Analyzing Completeness of Requirements Specification .......................... 444
Muhamad Idaham Umar Ong, Mohamed Ariff Ameedeen, and Imran Edzereiq Kamarudin

Research and Develop of AGV Platform for the Logistics Warehouse Environment ........................................ 455
Ha Quang Thinh Ngo, Thanh Phuong Nguyen, and Hung Nguyen

Transformation of UML Activity Diagram for Enhanced Reasoning ......................................................... 466
Irfan Chishti, Artie Basukoski, Thierry Chaussalet, and Neeraj Beeknoo

Estimation of Phase, Range, Doppler of Targets Using Maximum Likelihood Estimator ................................. 483
Nauman Anwar Baig, Muhammad Anwar Baig, Thawban Anwar Baig, Adnan Anwar Baig, and Abdullah Anwar Baig
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixelphonics: Colocating Sound and Image in Media Displays</td>
<td>491</td>
</tr>
<tr>
<td>Michael Filimowicz</td>
<td></td>
</tr>
<tr>
<td>Technological and Regulatory Developments for Electromagnetic</td>
<td>506</td>
</tr>
<tr>
<td>Transmission into the Millimeter Wave and Terahertz Wave Spectrum</td>
<td></td>
</tr>
<tr>
<td>Salim Hanna</td>
<td></td>
</tr>
<tr>
<td>SwICh: A Channel Assignment Technique for Self Coexistence of Cognitive Radio Networks</td>
<td>519</td>
</tr>
<tr>
<td>Komalpreet Kaur and Manish Wadhwa</td>
<td></td>
</tr>
<tr>
<td>A Comparison of Modern Localization Techniques in Wireless Sensor Networks (WSNs)</td>
<td>535</td>
</tr>
<tr>
<td>Maryum Hamdani, Usman Qamar, Wasi Haider Butt, Fatima Khalique, and Saad Rehman</td>
<td></td>
</tr>
<tr>
<td>A Resource Allocation Scheme for Multi-user MmWave Vehicle-to-Infrastructure Communication</td>
<td>549</td>
</tr>
<tr>
<td>Ramanathan Subramanian</td>
<td></td>
</tr>
<tr>
<td>Network Design, Simulations and Improvement, Using Riverbed Academic Edition, Version 17.5 at the Univers</td>
<td>568</td>
</tr>
<tr>
<td>Christopher Udeagha, Ashley Clarke, Shadrick Francis, Howard Perry, Kino Elliot, and Zhane Aflick</td>
<td></td>
</tr>
<tr>
<td>Improving Vacant Channel Utilization in Shared TV White Spectrum</td>
<td>586</td>
</tr>
<tr>
<td>George Atta-Boateng, Patrick Otoo Bobbie, K. O. Boateng, and E. K. Akowiah</td>
<td></td>
</tr>
<tr>
<td>Investigation of the User Redial and Its Effect on the Network</td>
<td>598</td>
</tr>
<tr>
<td>Jennifer Mintah Gyamfuuaa, Willie K. Ofosu, K. A. Dotche, and K. Diawuo</td>
<td></td>
</tr>
<tr>
<td>An Incentive Compatible Mechanism for Lowest-Cost Routing</td>
<td>608</td>
</tr>
<tr>
<td>Donghong Qin, Ting Lv, Jiahai Yang, Lina Ge, and Zhenkun Lu</td>
<td></td>
</tr>
<tr>
<td>Path Loss Propagation Prediction and Optimization Using Wallisch-Bertoni Model at 900 and 1800 MHz Over Macro-Cellular Western Regions of Nigeria</td>
<td>623</td>
</tr>
<tr>
<td>Zero-Copy Receive for Virtualized Network Devices</td>
<td>638</td>
</tr>
<tr>
<td>Kalman Meth, Joel Nider, and Mike Rapoport</td>
<td></td>
</tr>
<tr>
<td>Bandwidth Estimation for Admission Control in MANET: Review and Conceptual MANET Admission Control Framework</td>
<td>651</td>
</tr>
<tr>
<td>Folayo Aina, Sufian Yousef, and Opeyemi Osanaiye</td>
<td></td>
</tr>
</tbody>
</table>
Investigation of Wireless Communication from Under Seawater to Open Air with Xbee Pro S2B Based on IEEE 802.15.4 (Case Study: West Java Pangandaran Offshore Indonesia) .......... 672
Nurul Hiron, Asep Andang, and Nundang Busaeri

Towards Filling the Gap of Routing Changes in Software-Defined Networks ........................................ 682
Ali Malik, Benjamin Aziz, and Mo Adda

Local Conditioning: Exact Message Passing for Cyclic Undirected Distributed Networks .................................. 694
Matthew G. Reyes

Proof of Human Engagement on Decentralized Networks .............. 712
Qifeng Chen, Shiyu Zhang, and Wilson Wei

Performance Analysis of Software Defined Network with Three Controllers .................................................. 718
Shishupal Kumar, Nidhi Lal, and Vijay Kumar Chaurasiya

Reconfiguring Computer-Generated Holographic Coding Patterns to Enhance Optical Access Network Security ......... 733
Jen-Fa Huang, Chun-Chieh Liu, Chao-Chin Yang, and Chung-Hao Li

A PAL in Need Is a PAL Indeed: An Opioid Use Case for a Peer Alerting Lifeline (PAL) to Mitigate Risk Exposure .......... 743
Nicholas D. Preston, Derek Jacoby, and Yvonne Coady

Solving Startup-Delay-QoE Dilemma for Video Streaming Services in 5G Networks ........................................ 758
Khadija Bouraqia, Wissal Attaoui, and Essaid Sabir

Design and Analysis of Cycloconverter to Run Split Phase Induction Motor Using PWM Control .......................... 771
Iftakher Ahmed and Shubhalaxmi Kher

Robust Control of Physiologically Relevant Sit-to-Stand Motion Using Reduced Order Measurements .................. 783
Samina Rafique, A. Mahmood, and Muhammad Najam-ul-Islam

Real Time Comparative Analysis of Several Intelligent Speed Controllers Applied to the Direct Torque Control of Asynchronous Machine ........................................ 797
Chaymae Laoufi, Ahmed Abbou, and Mohammed Akherraz

Honey Bees Repellent Device: Preliminary Experimental Research with the Bees Hearing Sensitivity ................. 827
David Costa, João Paulo Teixeira, and Vasco Cadavez
Design of Photonic Crystal Fiber Surface Plasmon Resonance Sensor with External Channel Approach ........................................... 841
Veerpal Kaur and Surinder Singh

Infrared Sensor Controlled Wheelchair for Physically Disabled People ........................................................................ 847
Nadia Nowshin, Md Moontasir Rashid, Tasneema Akhtar, and Nafisa Akhtar

S2A Architecture with Sift-Out at the Sensor Level ........................................ 856

Embedded System for Access Control Based on Facial Biometry and RFID ........................................................................ 868
Cristian Henrique M. Souza, Ivanilson França V. Júnior, Robinson Luis S. Alves, Arthur Henrique C. Costa, and Melquiades P. Lima Júnior

On the Performability of Fault-Tolerant NCSs with Video Sensors . . . 884
Hadeer Ahmed, Tarek K. Refaat, Hassan Halawa, Ramez M. Daoud, and Hassanein H. Amer

Comparative Study on Standardized Tests to Determine Energy Efficiency in Three-Phase Induction Motors Operating at Partial Load ................................................................. 896
Alejandro Paz Parra, Manuel Vicente Valencia, and Carlos Alberto Lozano Espinoza

RF Biometric for Wireless Devices ................................................................. 904
R. Page Heller, Thomas G. Pratt, Joe Loof, and Eric Jesse

Remotely Operated Underwater Vehicle (ROV) Using Wireless Communication Protocol over a Floating Unit ....... 912
Mustajab Ahmad, Muhammad Ali, Muhammad Imran, Zoha Hashem, Shaikh M. Anas, Saddam Hussain, Abdul Saboor Khan, and Bilal M. Yousuf

A Digital Transmitter Architecture with Enhanced Delay, Power and Noise Performance for Sensors and IoT Applications .......... 925
Mohsen Sarraf and Fatemeh Forati

Systolic Design Space Exploration of Polynomial Division over $GF(3^m)$ ......................................................... 933
Ibrahim Hazmi, Fayez Gebali, and Atef Ibrahim

Electromagnetic Coupling Optimization by Coil Design Improvements for Contactless Power Transfer of Electric Vehicles .......... 944
Junlong Duan and Weiji Wang
Just-In-Time Project Task Scheduling ........................................... 959
Malcolm Ferrier

Commercialization Skills: Necessity for High Technology Entrepreneurs in Digital Era ...................................................... 965
Saheed A. Gbadegeshin

How Bots Have Taken over Brand Choice Decisions ................. 976
Adam Labecki, Phil Klaus, and Judith Lynne Zaichkowsky

A Cluster Based Collaborative Filtering Method for Improving the Performance of Recommender Systems in E-Commerce ...... 990
Bahman Sassani (Sarrafpour), Alaa Alahmadi, and Hamid Sharifzadeh

RecSmart: Data Augmentation to Facilitate Recommendation Using Skewed and Sparse Data of Restaurant Loyalty Programs .... 1002
Ishani Chakraborty

Modern Chatbot Systems: A Technical Review .......................... 1012
Abbas Saliimi Lokman and Mohamed Ariff Ameedeen

The Impact of Distributed Data in Big Data Platforms on Organizations .............................................................................. 1024
Oded Koren, Matan Binyaminov, and Nir Perel

The Blockchain: A New Framework for Robotic Swarm Systems .... 1037
Eduardo Castelló Ferrer

A Wearable General-Purpose Solution for Human-Swarm Interaction ... 1059
Eduardo Castelló Ferrer

Personal Food Computer: A New Device for Controlled-Environment Agriculture ................................................................. 1077

Input Password Using Simple Input Device ................................. 1097
Manabu Okamoto

Searching of Self-similar Spaces .................................................. 1106
Leszek Kaliciak, Hans Myrhaug, and Ayse Goker

LCC Passengers’ Willingness-to-Pay for the Baggage Check-in Additional Service: A Case Study on the Taiwan-Japan Route .......... 1125
Rong-Chang Jou, Chung-Wei Kuo, and Yi-Chun Chiu

Human Factors in e-Learning ...................................................... 1140
Marcel Simonette, Vera Queiroz, and Edison Spina

NoObesity Apps – From Approach to Finished App .................... 1145
Denyse King, Em Rahman, Alison Potter, and Edwin van Teijlingen
Strategic Technology Adoption in Networked Markets ................. 1158
Samira Hossein Ghorban

MedMatcher: A Mobile Game Designed to Improve Medication Understanding at Hospital Discharge ......................... 1176
Dawn M. Becker and Dave Becker

Author Index .................................................. 1181
Abstract. Now-a-days, social and professional networks have become main focus of interest for the research community to extract useful information. Many researchers have explored different features of social networks which help the experts to make discoveries in an easy way. LinkedIn is a professional network and there are more than one hundred millions of registered users on the LinkedIn. These users have different kinds and levels of expertise in various domains. Although the data available on the LinkedIn is in semi-structured form, however, still it’s a big challenge for the organizations to find the required expertise in such huge collection of data. In this paper, we proposed an automated technique which collects structured information from the LinkedIn profiles. An innovative algorithm has been designed and developed which ranks professionals based on their expertise level according to user selected criteria. Our proposed methodology (“Float Search”) also provides a user-friendly interface and an interactive visualization of the experts. Float Search also provides option for users to weight each required expertise according to their preferences. Results of Float Search have been compared with that of LinkedIn search and evaluated through user study. The results show that 70.49% of the reviewers considered our ranking approach better and 28.35% reviewers regarded it as the best approach for searching required experts.

Keywords: Experts ranking · LinkedIn · Experts finding · Visualization
Float search

1 Introduction

Experts are the people with specialized experience and knowledge [1]. Finding experts is critical to support a number of tasks such as: (1) finding experts in organizations and industries; (2) finding experts in academia; (3) finding experts in software projects; and (4) finding experts in enterprises. Although the information is available on the World Wide Web, expertise finding is a tough task. Almost all recruitment processes require having a resume in the digital form and these files are processed by the organizations manually [2]. The task of expert finding is achieved using both manual and automated approaches. The manual approaches give high accuracy but difficult to maintain, however, the automated approaches give low accuracy but efficient to compute. Most of the time, automatic techniques use a single dimension to find experts. Google
scholar uses citation to rank researchers; H-index is used to rank the expertise in the academia. But the use of single metric techniques is not enough to rightly and accurately reflect a particular expertise [3].

Despite the presence of huge collection of online information the organizations still face a great challenge to find experts in different domains. The information available on the social web portals and the social activities on these portals can be used for the recruitment processes. These activities help in creating the professional profiles on particular portals [4]. Many Researchers have explored different features of Social Networks which help the experts to discover in an easy way. Unfortunately data available on the social web are mostly unstructured and provide experts information in generalized form with few vocabularies. It is still a challenge for organizations and academia to find experts in a specific domain. Further this data is in a form which is not directly readable for the machines [1].

Social networks are further divided into two categories: (1) Internal social network, which is used only by a particular group or within the company or organization; (2) External social network, which is not specific to a group, company or organization, but is open to all communities and individuals [5]. LinkedIn is an external social network which provides a platform for the professionals to interact and share information with one another [6]. According to [7] there are more than 135 million of registered users on the LinkedIn. These users have different kinds and levels of expertise in various domains. Although the data available on the LinkedIn is in semi structured form, but still it’s a great challenge for the organizations to find the required expertise in such huge collection of data. Further this information is not available in machine readable format.

We proposed an automated technique which collects structured information from the LinkedIn profiles. The proposed technique extracts information from individual’s profiles like Name, expertise, Employer, Designation and duration in months and stores them in an XML file. This information is then stored in the database. Afterwards we employed a method to identify and rank expertise, which facilitate and enables the individuals and organizations to search and find expertise with minimum efforts. This methodology demonstrates dataset of engineers in Islamabad, Pakistan, However this may be extended for broader domain.

Finally, to enhance the methodology findings and visibility, our proposed methodology provides a user-friendly interface and an interactive visualization of the experts. The methodology provides our own distinct and unique way to visualize experts using HTML 5.0 canvas. The proposed approach is providing the opportunity for users to enter weights for each expertise according to their preference but LinkedIn does not provide such a facility. To evaluate the data extracted by our proposed methodology, we have used a data set of 100 top level of expertise and manually validated from project managers and Human resource managers working in different software industries.

The rest of this paper is organized as follows: the detailed literature survey on discoveries and visualization of expertise is provided in Sect. 2. The proposed methodology is explained in Sect. 3. Evaluation is discussed in Sect. 4 and finally, we conclude in Sect. 5 with some potential future directions.
2 Related Work

Pérez-Rosés et al. [8] have proposed a method for computing authority score using the number and quality of endorsement of a particular skill provided on a user profile. The problem was that the endorsement was not being calculated for people profile in social network website like LinkedIn. This method uses endorsement digraph to enhance the user profile for a particular skill. It applies the digraph ranking method like PageRank for ranking the user skills. Synthetic network consisting of 1493 nodes was used for testing and evaluating the proposed method. Basu et al. [5] presented that recommending ranked items to the user is still a big challenge for categorically different but interdependent items. Author has constructed an optimization formulation for restricted multi-slots which has the capability to model the items interactions belonging to various slots. To provide efficient solution, the solution was organized based on problem parameters and identified the best conditions. An algorithm was developed which provide efficient solution with least compromise on QCQP restrictions. The experiment results compared to other state of the art methods prove its efficiency and accuracy.

According to Wongsuphasawat et al. [9], it is difficult for naive analyst to view the automated visualized data in different views. To assist manual chart development, a tool “Voyager” that is based on mixed initiative and provides a browser with different features for recommended charts based on perceptual and statistical measures. The principle designs of this tool are motivating and recommend the visualization for generating interaction. The tool enhances data visualization and covers data variable with large aspect. Adoption of visualization tools was designed for balancing the fast exploration of specific answering of questions. According to Ananya et al. [10], many educational purpose software tools are available to enhance student skills of problem solving by hints and practicing. The student learning velocity of a problem can be defined, if user can significantly define the difference between the right solution and that of wrong attempt in problem solving. Using the finite automata construct, the author has extended the “JFLAP” a learning software tool for computing learning velocity of student working on a problem. This tool enables instructors for ranking students based on their learning capabilities and difficulty based ranking of the problem.

Fazel-Zarandi and Fox [11] have proposed a technique for creating and enhancing expert profiles for individuals. This technique employs First-Order Logic (FOL) that uses notations to infer and validate skills and competencies. It not only constructs and updates the skill of a person but it also finds the proficiency level for a particular skill. It will make the system highly expensive, complex, and inefficient. In the future author has planned for new system with the capability of evolving expert profile and conduct gap analysis. The conduct gap analysis will help to understand that who need what type of training and to evaluate these systems in real world.

Hristoskova et al. [12] has presented a framework, which collect information from different sources like Google scholar to build and maintain a dataset for expert finding. The internal functionality of this framework is divided into three parts including, (1) data gathering: collecting initial information like authors list, title and abstracts of articles; (2) data cleaning: merging and disambiguation of authors; and (3) analyzing and clustering of this data to relevant author. It improves the DBLP result by 17%.
In the future author intends to put the negative weights to the graph and increase the no of resources for collecting information. Balog et al. [13] has found that the existing techniques for finding experts in an organization are costly and need effort. Author proposed two cost effective models based on candidate’s model and document model. The proposed system employs rule-based technique that introduces four methods of binary associations, based on names and emails of candidates. Its performance has evaluated on World Wide Web Consortium (W3C) TREC Enterprise Track 2005. The results show that proposed models are among top five best models and Model 2 outperforms than Model 1. However, unstructured data and different data formats make the extraction methods more complex and degrade their performance remarkably.

Afzal et al. [3] has highlighted that to find and assign an expert to knowledge domain is needed in various era and situations. Author has used the multifaceted approach to create the expert profile by assigning weights to the expertise. To visualize the expertise, the proposed system uses the combination of hyperbolic and spiral visualization techniques. Hyperbolic browser visually represents ACM classification hierarchies. Spiral is used to visualize high profile authors in a sequence by selecting a node in the ACM hierarchy. This proposed approach is only implemented for Authors and reviewers of Journal of Universal Computer Science (JUCS).

3 Methodology

We propose a methodology for mining software engineers using LinkedIn. The remainder of sections, we define that how our proposed methodology extract data from the LinkedIn professional network and other operations performed by our methodology to rank different individuals in the field of software engineering according to their expertise level. There is more than one hundred millions of record available on the LinkedIn, so a lot of effort and time is needed to find an individual with particular expertise [2]. Therefore, we proposed a methodology which enables users to find individual with required level of expertise in an automated way. The main target of our methodology is to mine only structure information of Software engineers available on the LinkedIn. We use java selenium server to crawl LinkedIn web pages of experts. By using PHP XML generator we extract relevant fields/attributes from the crawled LinkedIn pages and generate the XML file. To further process these information are loaded to the database. Expertise calculation engine is used to compute the expertise dynamically rank experts based on the user provided information. Complete architecture of our proposed methodology is given in Fig. 1.

3.1 XML Generation

Our main contribution is to provide ranked list of software engineers in an XML form and we make this XML openly available to the research community. Secondly, we will have to provide an interactive visualization of the ranked experts. For this purpose we have developed a crawler which crawls LinkedIn profiles of each software engineers and store their HTML contents. Afterward we extract only relevant information like Name, expertise, Employer, Designation and duration in months from the crawled
HTML pages. However an individual has expertise with more than one employer in single language/tool. To compute the total expertise we use the following formula.

$$E_t = \sum_{j=0}^{n} E_n$$ (1)

After calculating the expertise, this data is then stored in an XML file. Figure 2 shows the sample XML, which consist of different nodes. The reason for storing expertise information in XML file is that, it is compatible for almost all kind of technologies, tools and environments. It can be easily loaded to different DBMS like Mysql, sql server, etc. for further processing.

### 3.2 Experts Ranking

After generating the XML, the next phase of the proposed method is to rank the experts according to experience. For ranking the experts, the methodology provide the top experts to the user for the experience he required. User will select the experience name in the user interface and he will provide the weight of experience according to his requirement. The first step is used to store the user entered weights in an array using (2).

$$W[i] =: \prod_{i=1}^{n} W[i] \quad \forall i \in Z^+ \text{ and } i = 1, 2, \ldots, n$$ (2)

In (2), $(W[i])$ represent the array and “$\prod_{i=1}^{n}$” is used for iteration and the colon (:) sign in “=:” shows that this equation has been defined in this paper for iteration purpose.

$$W_t = \sum_{i=1}^{n} W_i \quad \forall i \in Z^+ \text{ and } i = 1, 2, \ldots, n$$ (3)

$$W_{a[i]} =: \prod_{i=1}^{n} W[i] W_t \quad \forall i \in Z^+ \text{ and } i = 1, 2, \ldots, n$$ (4)
Equations (3) and (4) are used to sum the total weight and store them in $W_t$ and then computes the actual weights ($W_a[i]$) and assign them to an array ($W[i]$).

$$I_{m}^{n} = \left\{ I_{j=1}^{m} E_{j}[c] \forall j \in Z^+ \text{ and } j = 1, 2, \ldots, m \rightarrow 4.2 \left\{ I_{k=1}^{p} E_{j}[c] \right\} \right\}$$

$$= \left\{ E_{js} \times W_a[i] \quad W_a[i] \iff E_{js} \rightarrow 4.1 \right\}$$

where $c = n \times m \times p$, $\forall i, j, \text{ and } k \in Z^+$ and

$$i = 1, 2, \ldots, n, j = 1, 2, \ldots, m, k = 1, 2, \ldots, p$$

(5)

In (5), there are three nested loops. The outer loop is executed “n” time where n is the total number of users in the dataset. Initially in this paper a dataset consist of two hundred (200) experts. Second loop will repeat “m” and “m” represent the total expertise like Linux, oracle, php, etc. of each user from the dataset. Third loop which is nested iterates “p” times where “p” shows the number expertise/languages/technologies selected by the employer using user interface (UI) provided. Inside the loop (iii), “if”
condition is provided which check to select only those expertise of an experts which match with expertise selected by the employer in the UI.

If the given condition is successful then duration of the matched expertise \( E_{js} \) is multiplied with \( W_{a[i]} \). Each time at the end of loop (ii) the overall expertise of an individual is calculated and stored in array \( E_{tw[j]} \).

\[
\text{SORT}(E_{tw})
\]

In (6) “SORT ()” function is applied on array “\( E_{tw[j]} \)” to sort expertise in ascending order.

\[
E_{tw}[i] := \sum_{i=1}^{5} E_{tw}[i] \quad \forall i \in \mathbb{Z}^+ \text{ and } i = 1, 2, \ldots, 5
\]

Finally, after calculating total expertise weight for each individual, the loop shown at (7) is used to select first five (5) experts form sorted array and store them in array “\( E_{ts}[i] \)”, which is then used to visualize these experts.

3.3 Visualize the Top Experts Based on Expertise Level

The proposed methodology provides a user-friendly interface and an interactive visualization technique which uniquely visualize expertise using PHP and HTML 5 canvas.

\[
\text{Max} = E_{ts}[1]
\]

Equation (8) uses the array \( E_{ts} \) from to find maximum expertise using Max function and stored in variable max.

\[
\text{CC} = 100/\text{Max}
\]

In (9) max is then used to find the circle co-efficient (Cc) by using formula: Cc = 100/max. Where 100 is the radius of main circle which is constant.

\[
I_{i=1}^{n} I_{j=1}^{m} RExp := \{ E_j \times CC \quad \forall i \text{ and } j \in \mathbb{Z}^+ \text{ and } i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, m \}
\]

\[
\text{makeCircle}(RExp)
\]

The inner loop is used to calculate the radius of each expertise and then makeCircle() is used to draw inner circle. The main interface of the expertise calculation engine is presented in Figs. 3 and 4.

As shown in Figs. 3 and 4 user is able to select the company name from the provided multiple selection list. When a user is unable to select a company, by default all the companies in the list would be used for calculation. Subsequently user selects the programming language/Technology in which user looking for expert. For example user can select php, C#, Asp.Net, etc. To further filter the results the user can select
experience ranges. For example in Fig. 3 user has selected experience from 3 month to 7 years. In case a user need an individual(s) with expertise in more than one language, tool or technology, the interface provide the facility to the user to enter required weight for each language/tool/technology in percentage. The default value for the weight will be 100. Afterward, when user submits their query, a list of ranked experts is visualized in the form of circles. The outer circle represents a single individual. The experts are displayed from left to right according to their ranks. Inner circles represent their expertise level in different areas/fields. The size of the inner circle shows level of expertise of an individual.

This interface provide facility to the users to select individuals with multiple expertise like expertise in PHP, C#, Asp.Net, Mysql and Sql Server, etc. User can select as many expertises as needed by just clicking the button labelled as “Add Another row” and then select experience and enter weight against each expertise. Figure 3 shows the User interface with visualization. In this figure user selects more than on expertise like Oracle and Linux. The user is an employer of the company working on various kinds of project using different languages/technologies. For example the projects are Oracle based, Java based and DotNet based projects. Employer need a team lead for Oracle based project. The expert must have experience in oracle and Linux. The proposed methodology gives five experts based on their experiences, in these experts, the ranked no 5 expert is the right person for the user to
be selected because he has experience in Oracle and Linux which is mandatory condition but he is also has expertise in Oracle financials, Oracle developer, java and DotNet. The reason is that this expert can easily be assigned to other projects when it is required. This means that the methodology not only help users to retrieve high level experts according to users query but also helps to find experts in related domain which is the implicit query of the users.

4 Results and Discussions

To evaluate the data extracted and ranked expertise by the proposed methodology to find the answers of the following questions:

1. Whether the individuals ranked by the methodology are correct?
2. Is this mined information will be helpful to find and expert?

To evaluate and validate the result of proposed methodology (Float Search), three different search results of the Float search have been evaluated and validated each one from the 87 different experts in the software industry manually. Table 1 shows the detailed results (All Reports Data) of the evaluation.

<table>
<thead>
<tr>
<th></th>
<th>Oracle &amp; Linux</th>
<th>PHP</th>
<th>Dot Net</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Agree</td>
<td>65</td>
<td>65</td>
<td>54</td>
</tr>
<tr>
<td>Strongly Agree</td>
<td>21</td>
<td>21</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 1 shows that among 261 total feedbacks, 70.49% consider “Agree”, 28.35% selected “Strongly Agree” and only 1.15% has considered “Disagree”. The overall result of the evaluation strongly supports “Float Search” criteria instead of LinkedIn search technique for finding experts. As shown in Fig. 5, when we searched with the keyword “php”, in LinkedIn we found these 5 top experts. But the experts number 4 “Syed Junaid Ali” is most experience in php. It is not showing 1st top expert in LinkedIn. But in float search, “Syed Junaid Ali” is showing top 1st expert. So float search is better to provide ranking of experts on the basis of their experience.

As shown in Fig. 6, when we searched with the keyword “oracle and linux”, in LinkedIn we found these 5 top experts. But the experts number 4 “Atif Humayun” is most experience in php. It is not showing 1st top expert in LinkedIn. But in float search, “Atif Humayun” is showing top 1st expert. So float search is better to provide ranking of experts on the basis of their experience.
**Fig. 5.** Graph comparison of float search and LinkedIn search with benchmark by using filter “php”.

**Fig. 6.** Graph comparison of float search and LinkedIn search with benchmark by using filter “Oracle, Linux”.
5 Conclusion

LinkedIn is a professional social network. There are more than 135 million registered users on the LinkedIn. These users have different kinds and levels of expertise in various domains. Although the data available on the LinkedIn is in semi-structured form, but still it’s not an easy task for the organizations to find the required expertise in such massive collection of data. Further this information is not in a format which is directly readable for the machine. The proposed methodology (Float Search) has provided an automated technique which collects only structured information from individual’s profiles and stores them in an XML file. For further processing this information is then loaded to the database. Moreover, this methodology has provided a user-friendly interface and an interactive visualization of the experts. Float search has given the facility for user to give weight for each expertise according to their preference but LinkedIn does not provide option to give weight for each searching preference.

In future it is intended to extend the scope of this research and to further enhance the current methodology by providing an efficient and user-friendly interface. It also needed to get the unstructured data of LinkedIn and use it for searching results. It will be helpful for the organizations to get the suitable expert for the required job. Many experts have not added the experience in structured format, the enhancement of the methodology will also add those experts in the search results.

References


11. Fazel-Zarandi, M., Fox, M.S.: Constructing expert profiles over time for skills management and expert finding. Paper Presented at the 11th International Conference on Knowledge Management and Knowledge Technologies (i-KNOW 2011), Graz, Austria (2011)


A Modern Book Archive Geared for Custom Publishing

Liam Borgstrom

University of Pretoria, Pretoria, South Africa
liam.borgstrom@up.ac.za

Abstract. The world of book publishing has changed drastically with the adoption of XML technologies, by imposing regular structures on authors and allowing production departments to design, alter, and redistribute books in a fraction of the time it would have taken previously. However, as we move to make our content more machine readable, we need to consider how we apply metadata in a way that supports custom publishing solutions. The author proposes a new way of looking at XML schemas for archiving books, and suggests how this may be implemented with code examples. While an archive generally preserves content as it was produced, because publishers’ archives are potential sources of revenue, they should be compiled in such a way that allows for calculated mining, supported by semantic mark-up describing not only the book function, but the intended uses.

Keywords: Custom publishing · XML · Higher education · XML workflow · Contextual publishing

1 Introduction

So often when looking into digital innovation we find that the focus is on the big advances, such as artificial intelligence and big data. Attention is focused on how this can be used to manage cars and cities, create smart recommendations for advertisers and assist us in our daily lives. But how do we apply this to books?

Yes, Siri and Google Assist can tell us about anything with a Wikipedia article, but what if we say, “Ok Google, I want to get into Deep Learning and Neural Networks, I’m in high school and have done some programming, and I think that this may be a potential career move. Can you give me an introduction to the field?”

Typing this is now gives you links to AI blogs, MIT and Coursera Courses, Newspaper articles, some adverts for data science, and eventually a 2208 book on page four of the results. However in Frederick Pohl’s Day of the Boomer Dukes (1956) [1] the protagonist – Foraminifera 9-Hart Bailey’s Beam – asks the Learning Lodge index computer to find some information for him based on certain criteria and, “Then there was a hiss and a crackle, and in the receiver of the desk a book appeared. I unzipped the case, took it out, and opened it to the pages marked on the attached reading tape.”

This idea, that a book can be created to the needs of a specific reader is not a new one, and is being attempted by educational companies today venturing into the practice of Custom Publishing.
Custom publishing has its greatest potential in the higher education market, where books are increasingly being compiled from front and backlist content according to the requirements of a specific course.

However, in order to achieve this, we need more than a business practice; we need a new way of looking at our content. Brian O’Leary has addressed this in his context-first publishing philosophy [2]. However, I am not sure how many publishers are really looking into a practical way of implementing this. Publishing in XML with DocBook and custom schemas is certainly the first step, but how should we preparing our books so that the content can be intelligently repurposed based on its content rather by its formatting.

I find that the theme I am constantly repeating to my students is that digital content needs to be machine-readable. Programmers understand this. They understand that for any system to work, a precise vocabulary must be used in defining different classes, functions and variables, in this way the software can work according to its programmed logic. The software used in publishing is designed in this way. InDesign (still the book design staple) works best when character, paragraph, and object styles are used consistently. This allows for valid nesting and importing of content and for eventual compatibility with a publishing system designed to access content defined in XML.

While XML schemas are becoming more semantic, custom publishing solutions are still reliant on describing the book as part of a book. For Pohl’s scene in the Learning Lodge to occur, an AI must be able to examine a piece of text and understand not only its semantic meaning, but the intent of the writing. This is not only a tool for training AI, but to assist in human searches as well. While search algorithms are constantly adapting, they must still respond to usage patterns and direct language. Full text searching is still our best method quickly discerning the content of a book. However, this is limited by our own language and individual means of expressing our objectives. For custom publishing to truly work, we need to be able to extract content based on its purpose and function, and be able to locate suitable content timeously from our backlist.

My proposal here is for an archiving system with the potential to catalogue content down to the paragraph level. Content should be semantically defined so that it is possible to make a search query such as, “explanations and exercises in geometry for students written in a conversational and accessible tone”.

As publishers, we must strive to defend the unique aspects of our books. The profile of the author, or the marketing blurb is not enough, (especially in educational environments) and it is up to us as publishers to create textual content that engages the intended readership as much as possible. The query above as a search request could only realistically pull results from the marketing information or introduction. However, with granular enough metadata this query could extract the exercises themselves based on their intended audience.

 Practically this comes down to rigorous metadata. However, mention metadata in the book world, and thoughts immediately fly to ONIX, MARC, Dublin-Core, XMP and so forth. However, these formats – valuable as they are – are primarily useful in distribution, discovery, and access of the complete volume. They assist the sales division, but how can we use them in production?

If we are to truly redefine publishing in the 21st century, we need to define content right down to the paragraph, so that as publishers we are truly able to create a content
database of relatable, and chunkable material, and that can be thoughtfully collated and curated according to need, and possibly even by machine.

2 An XML Framework

While O’Leary’s explanation of contextual publishing focusses on how content is interconnected, and needs to be aware of how it is being consumed, I think what we can add to this is, “who is reading it?”

For this we need to start thinking practically about how we store our content. The best technology we have for this so far is XML, it’s ubiquitous, scalable, and suitably rigorous.

If we consider how most modern publishers are working with XML schemas we will probably find that books are defined as such:

```xml
<Book>
  <FrontMatter>
    <dc:identifier id="isbn-id">urn:isbn:9782541387655</dc:identifier>
    <dc:creator id="author">John Andrews</dc:creator>
    <meta refines="#author" property="role" scheme="marc:relators">aut</meta>
    <dc:title>Book of Wonders</dc:title>
    ...
  </FrontMatter>
  <MainMatter>
    <Chapter>
      <Title>What is Wonder?</Title>
      <p class="first">When I first began...</p>
    </Chapter>
    ...
  </MainMatter>
  <EndMatter>
    <Index>
      <Iheading>A</Iheading>
      <Iterm id="aardvark">Aardvark <a href="#IT-aardvark">IpageRef/</a></Iterm>
    </Index>
    ...
  </EndMatter>
</Book>
```

Fig. 1. A typical book schema.