

IC3K 2016

Final Program and Book of Abstracts

8th International Joint Conference on Knowledge Discovery,
Knowledge Engineering and Knowledge Management

Porto - Portugal
November 9 - 11, 2016

Sponsored by

INSTICC - Institute for Systems and Technologies of Information, Control and Communication

Co-Sponsored by

FCT - Fundação para a Ciência e Tecnologia

In Cooperation with

ACM SIGMIS - ACM Special Interest Group on Management Information Systems

ACM SIGAI - ACM Special Interest Group on Artificial Intelligence

APPIA - Portuguese Association for Artificial Intelligence

APRP - Associação Portuguesa de Reconhecimento de Padrões

AI*IA - Associazione Italiana per l'Intelligenza Artificiale

ERCIM - The European Research Consortium for Informatics and Mathematics

AAAI - Association for the Advancement of Artificial Intelligence

GI - SIG KM - German Society for Computer Science - Special Interest Group on Knowledge Management

Table of Contents

Foreword	5
Social Event and Banquet	7
General Information	8
Rooms Layout	9
Program Layout	10
Wednesday Sessions: November 9	19
Thursday Sessions: November 10	33
Friday Sessions: November 11	45
Author Index	59

Foreword

This volume contains the final program and paper abstracts of the eighth International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2016) which was sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC) and held in Porto, Portugal and co-organized by Fundação para a Ciência e a Tecnologia (FCT). IC3K was organized in cooperation with the ACM SIGMIS - ACM Special Interest Group on Management Information Systems, ACM SIGAI - ACM Special Interest Group on Artificial Intelligence, APPIA - Portuguese Association for Artificial Intelligence, APRP – Associação Portuguesa de Reconhecimento de Padrões, Associazione Italiana per l'Intelligenza Artificiale, ERCIM - European Research Consortium for Informatics and Mathematics, AAAI – Association for the Advancement of Artificial Intelligence and the German Society for Computer Science – Special Interest Group on Knowledge Management.

The main objective of IC3K is to provide a point of contact for scientists, engineers and practitioners interested on the areas of Knowledge Discovery, Knowledge Engineering and Knowledge Management. IC3K is composed of three co-located complementary conferences, each specialized in one of the aforementioned main knowledge areas. Namely: - International Conference on Knowledge Discovery and Information Retrieval (KDIR) - International Conference on Knowledge Engineering and Ontology Development (KEOD) - International Conference on Knowledge Management and Information Sharing (KMIS) The International Conference on Knowledge Discovery and Information Retrieval (KDIR) aims to provide a major forum for the scientific and technical advancement of knowledge discovery and information retrieval. The International Conference on Knowledge Engineering and Ontology Development (KEOD) provides a point of contact for scientists, engineers and practitioners interested in the scientific and technical advancement of methodologies and technologies for Knowledge Engineering and Ontology Development both theoretically and in a broad range of application fields. The goal of the International Conference on Knowledge Management and Information Sharing (KMIS) is to provide a major meeting point for researchers and practitioners interested in the study and application of all perspectives of Knowledge Management and Information Sharing.

The joint conference, IC3K received 211 paper submissions from 48 countries in all continents, of which 18% were accepted as full papers. The high quality of the papers received imposed difficult choices in the review process. To evaluate each submission, a double blind paper review was performed by the Program Committee, whose members are highly qualified independent researchers in the three IC3K Conferences topic areas.

Moreover, the conference also featured a number of keynote lectures delivered by internationally well-known experts, namely Dieter A. Fensel (University Innsbruck, Austria), Frans Coenen (University of Liverpool, United Kingdom), Una-May O'Reilly (MIT Computer Science and Artificial Intelligence Laboratory, United States) and Marijn Janssen (Delft University of Technology, Netherlands), thus contributing to increase the overall quality of the conferences and to provide a deeper understanding of the conferences interest fields. The IC3K program also includes 2 workshops and 1 doctoral consortium.

To recognize the best submissions and the best student contributions, awards based on the best combined marks of paper reviewing, as assessed by the Program Committee, and the quality of the presentation, as assessed by session chairs at the conference venue, were conferred at the closing session of the conference.

All presented papers will be submitted for indexation by Thomson Reuters Conference Proceedings Citation Index (ISI), INSPEC, DBLP, EI (Elsevier Engineering Village Index) and Scopus, as well as being made available at the SCITEPRESS Digital Library. Additionally, a short list of presented papers will be selected to be expanded into a forthcoming book of IC3K Selected Papers to be published by Springer Verlag.

Building an interesting and successful program for the conference required the dedicated effort of many people. We would like to express our thanks, first of all, to all authors including those whose papers were not included in the program. We would also like to express our gratitude to all members of the Program Committee and auxiliary reviewers, who helped us with their expertise and valuable time. Furthermore, we thank the invited speakers for their invaluable contribution and for taking the time to synthesize and prepare their talks. Moreover, we thank the workshop chairs whose contribution to the diversity of the program was decisive. Finally, we gratefully acknowledge the professional support of the INSTICC team for all organizational processes.

Conference Chair

Joaquim Filipe, Polytechnic Institute of Setúbal / INSTICC, Portugal

KDIR Program Chair

Ana Fred, Instituto de Telecomunicações/IST, Portugal

KEOD Program Co-chairs

Jan Dietz, Delft University of Technology, Netherlands

David Aveiro, University of Madeira / Madeira-ITI, Portugal

KMIS Program Co-chairs

Kecheng Liu, University of Reading, United Kingdom

Jorge Bernardino, Polytechnic Institute of Coimbra - ISEC, Portugal

Social Event and Banquet

**Venue: Guided Visit to the “Caves Taylor” followed by Dinner with a Music show
Thursday 10, 19:00 - 23:30**

Port is one of the great classic European wines and its history is a long and fascinating one.



Port wine is produced in the mountainous eastern reaches of the Douro Valley in northern Portugal, one of the world's oldest and most beautiful vineyard areas where wine has been made for at least two thousand years. In 1756 the Port wine vineyards of the Douro became the first vineyard area in the world to be legally demarcated. Like other great classic wines, Port owes its distinctive character to a unique association of climate, soil, grape variety and wine making tradition. The unique terroir of the Douro Valley and its remarkable wines cannot be replicated elsewhere.

Traditionally it is served towards the end of the meal with cheese, as a dessert wine or as an after dinner drink although some styles, like white Port, can also be enjoyed as an aperitif.

Many creative chefs also enjoy pairing Port wine with main dishes and it is one of the best wines to enjoy with chocolate.



Port is regarded as one of the most civilized and sociable of wines which will help to make any occasion special, whether a quiet evening by the fireside, an informal gathering of friends or a sophisticated formal meal.

For many, Taylor's is the archetypal Port house and its wines the quintessential Ports. Established over three centuries ago in 1692, Taylor's is one of the oldest of the founding Port houses. It is dedicated entirely to the production of Port wine and in particular to its finest styles.

Taylor's Port is the last of the original English founding port companies to remain family owned. It has never been bought, sold or taken over. The company is run by descendants of the founders. This ensures its outlook and philosophy remains focused on the production of top quality ports.



Consumers everywhere are becoming more knowledgeable and curious and therefore more likely to be attracted to wines such as Port which represent genuine quality and tradition and which have fascinating stories to tell.

General Information

Welcome Desk/On-site Registration

Tuesday, November, 8 – Open from 16:30 to 18:15
Wednesday, November, 9 – Open from 08:30 to 18:30
Thursday, November, 10 – Open from 08:30 to 18:30
Friday, November, 11 – Open from 10:15 to 18:45

Opening Session

Wednesday, November 9, at 10:15 in the Pinhão room.

Welcome Drink

Wednesday, November 9, at 18:30 in the Foyer.

Closing Session

Friday, November 11, at 18:30 in the Pinhão room.

Farewell Drink

Friday, November 11, at 18:45 in the Foyer.

Meals

Coffee-breaks will be served in the Foyer to all registered participants.
Lunches will be served in the Restaurant to all registered participants. Please check the hours in the Program Layout.

Communications

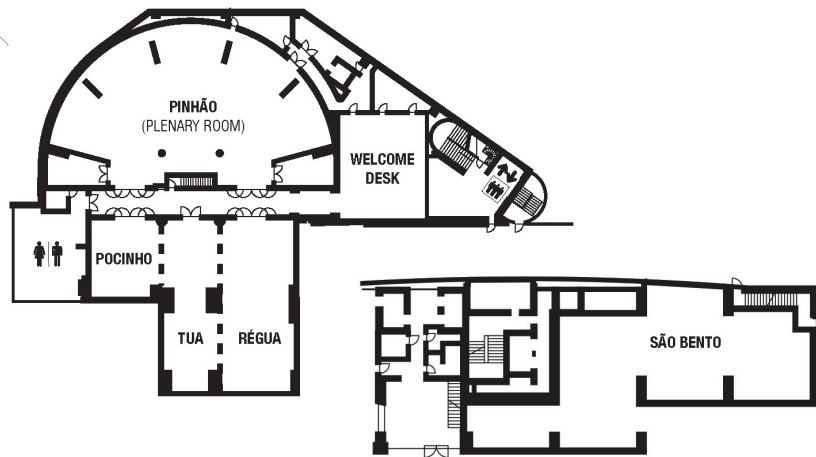
Wireless access will be provided free of charge to all registered participants.

Secretariat Contacts

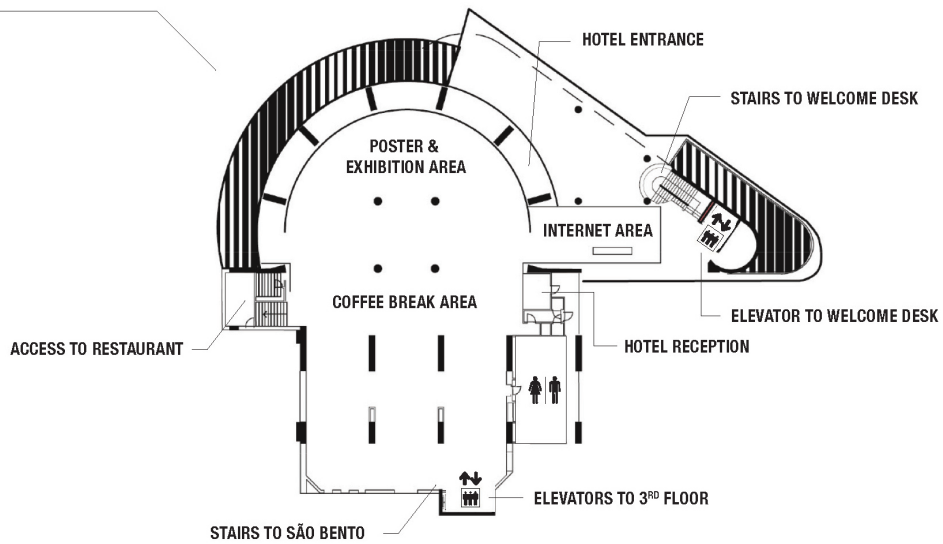
IC3K Secretariat
Address: Av. D. Manuel I, 27A 2° Esq.
2910-595 Setúbal, Portugal
Tel.: +351 265 520 184
Fax: +351 265 520 186
e-mail: ic3k.secretariat@insticc.org
website: <http://www.ic3k.org/>

Rooms Layout

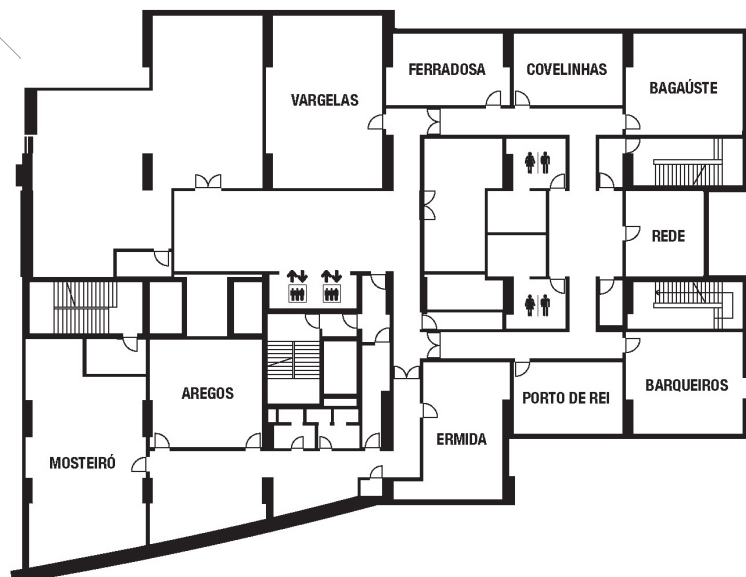
FLOOR 0



FLOOR 1



FLOOR 3



Program Layout

	Tuesday, November 8	Wednesday, November 9	Thursday, November 10	Friday, November 11
8:00				
8:30			Welcome Desk & Registration	
9:00	Welcome Desk & Registration	SKY Session 1	European Project Space - Session	Tutorial
9:30			IC3K Session 3	Tutorial
10:00			DC3K Session	
10:30	Opening Session		Welcome Desk & Registration	
11:00	Panel	Invited Talk	Coffee-Break	
11:30		KITA Session 1	IC3K Session 4	IC3K Session 6
12:00			Tutorial	
12:30	IC3K Session 1	SKY Session 2	Keynote Lecture Marijn Janssen	
13:00		KITA Session 2	Keynote Lecture Dieter A. Fensel	
13:30	Lunch		Lunch	
14:00	Lunch		Lunch	
14:30				
15:00		Invited Talk	European Project Space - Panel	
15:30	IC3K Session 2			IC3K Session 7
16:00		SKY Session 3	European Project Space - Session	
16:30			IC3K Session 5	
17:00	Welcome Desk & Registration	Coffee-Break	Coffee-Break	Coffee-Break
17:30		IC3K Poster Session 1	IC3K Poster Session 2	IC3K Poster Session 3
18:00		Keynote Lecture Una-May O'Reilly	Invited Talk	
18:30		Welcome Drink		Keynote Lecture Frans Coenen
19:00			Buses to Banquet	Closing Session
19:30			Social Event	Farewell Drink
20:00				
20:30				
23:00				
23:30				

Final Program and Book of Abstracts

Contents

Wednesday Sessions: November 9

7th International Workshop on Software Knowledge (SKY) - Session 1 (09:00 - 10:15)	
Room Bagaúste - SKY	21
1: Development and Evaluation of a Software Requirements Ontology, <i>by He Tan, Muhammad Ismail, Vladimir Tarasov, Anders Adlemo and Mats Johansson</i>	21
6: Syntactic-Semantic Extraction of Patterns Applied to the US and European Patents Domain, <i>by Anabel Fraga, Juan Llorens, Eugenio Parra, Leticia Arroyo and Valentín Moreno</i>	21
Opening Session (10:15 - 10:30)	
Room Pinhão - IC3K	21
Panel (10:30 - 11:30)	
Room Pinhão - IC3K	21
7th International Workshop on Software Knowledge (SKY) (10:30 - 11:30)	
Room Bagaúste - SKY Invited Talk	21
Concept Lattices as a Versatile Tool to Discover Knowledge in Software , <i>by Daniel Speicher</i>	21
2nd International Workshop on the design, development and use of Knowledge IT Artifacts in professional communities and aggregations. Knowledge Artifacts as resources in the maker and DIY communities (KITA) - Session 1 (10:30 - 11:30)	
Room Barqueiros - KITA	21
2: Integrating Knowledge Artifacts and Inertial Measurement Unit Sensors for Decision Support, <i>by Stefano Pinardi, Fabio Sartori and Riccardo Melen</i>	21
4: Making and Understanding - A Vision for IoT Makerspaces, <i>by Julian Dax and Volkmar Pipek</i>	21
Parallel Session 1 (11:45 - 13:15)	
Room Mosteiró - KDIR	22
9: StorylineViz: A [Space, Time, Actors, Motion] Segmentation Method for Visual Text Exploration, <i>by Iwona Dudek and Jean-Yves Blaise</i>	22
32: A Multi-Layer System for Semantic Textual Similarity, <i>by Ngoc Phuoc An Vo and Octavian Popescu</i>	22
52: Automatic Text Summarization by Non-topic Relevance Estimation, <i>by Ignacio Arroyo-Fernández, Juan-Manuel Torres-Moreno, Gerardo Sierra and Luis Adrián Cabrera-Diego</i>	22
Room Aregos - KEOD	22
19: Verifying and Mapping the Mereotopology of Upper-Level Ontologies, <i>by Lydía Silva Muñoz and Michael Grüninger</i>	22
31: Developing a Model of Agreement Negotiation Dialogue, <i>by Mare Koit</i>	23
32: Data Integration and Visualization for Knowledge Mapping in Strasbourg University, <i>by Amira Essaid, Quynh Nguyen Thi and Cecilia Zanni-Merk</i>	23
39: Empowering the Model-driven Engineering of Robotic Applications using Ontological Semantics and Reasoning, <i>by Stefan Zander, Nadia Ahmed and Yingbing Hua</i>	23
Room Vargelas - KMIS	23
19: Enabling Centralised Management of Local Sensor Data Refinement in Machine Fleets, <i>by Petri Kannisto and David Hästbacka</i>	23
30: Dual-Priority Congestion Control Mechanism for Video Services - Real Network Tests of CVIHIS, <i>by Juha Vihervaara, Pekka Loula and Teemu Alapaholuoma</i>	23
34: Coordination Problems in Knowledge Transfer: A Case Study of Inter-Organizational Projects, <i>by Nestor A. Nova and Rafael A. Gonzalez</i>	24
2nd International Workshop on the design, development and use of Knowledge IT Artifacts in professional communities and aggregations. Knowledge Artifacts as resources in the maker and DIY communities (KITA) - Session 2 (11:45 - 13:15)	
Room Barqueiros - KITA	24
1: Knowledge-Oriented Technologies & Network Marketing Direct Selling Organizations (NMDSO) - Some Preliminary Insights into the Nature and the Goals of Shared Knowledge, <i>by Carolina Guerini and Eliana Alessandra Minelli</i>	24

3: Makers in the Plant? Exploring the Impact of Knowledge IT Artifacts on DIY Practices in Manufacturing Firms, <i>by Luca Cremona and Aurelio Ravarini</i>	24
5: More Time for the Doing, Having <i>Made</i> the Thinking - 3D Printing for Knowledge Circulation in Healthcare, <i>by Federico Cabitza, Angela Locoro, Aurelio Ravarini and Vittorio Satta</i>	24
7th International Workshop on Software Knowledge (SKY) - Session 2 (11:45 - 13:15)	
Room Bagaúste - SKY	25
5: The Modularity Matrix as a Source of Software Conceptual Integrity, <i>by Iaakov Exman</i>	25
9: Requirements Engineering: More Is Not Necessarily Better, <i>by Gonzalo Génova</i>	25
8: Modeling Semantics <i>sans</i> Mathematical Formalism, <i>by Reuven Gallant</i>	25
7th International Workshop on Software Knowledge (SKY) (14:30 - 15:30)	
Room Bagaúste - SKY Invited Talk	25
Some Blends are Better than Others: Blend Quality Assessment in Computational Models of Concept Blending, <i>by Pedro Martins</i>	25
Parallel Session 2 (14:30 - 16:30)	
Room Mosteiró - KDIR	25
20: Keyword-based Approach for Lyrics Emotion Variation Detection, <i>by Ricardo Malheiro, Hugo Gonçalo Oliveira, Paulo Gomes and Rui Pedro Paiva</i>	25
21: Classification and Regression of Music Lyrics: Emotionally-Significant Features, <i>by Ricardo Malheiro, Renato Panda, Paulo Gomes and Rui Pedro Paiva</i>	26
47: A Machine Learning Approach for Layout Inference in Spreadsheets, <i>by Elvis Koci, Maik Thiele, Oscar Romero and Wolfgang Lehner</i>	26
68: Estimating Sentiment via Probability and Information Theory, <i>by Kevin Labille, Sultan Alfarhood and Susan Gauch</i>	26
Room Aregos - KEOD	26
9: A Smart System for Haptic Quality Control - Introducing an Ontological Representation of Sensory Perception Knowledge, <i>by Bruno Albert, Cecilia Zanni-Merk, François de Bertrand de Beuvron, Jean-Luc Maire, Maurice Pillet, Julien Charrier and Christophe Knecht</i>	26
24: Proactive Prevention of False-Positive Conflicts in Distributed Ontology Development, <i>by Lavdim Hallaj, Irlán Grangel-González, María-Esther Vidal, Steffen Lohmann and Sören Auer</i>	27
26: Model-Intersection Problems with Existentially Quantified Function Variables: Formalization and a Solution Schema, <i>by Kiyoshi Akama and Ekawit Nantajeewarawat</i>	27
25: CDM-Core: A Manufacturing Domain Ontology in OWL2 for Production and Maintenance, <i>by Luca Mazzola, Patrick Kapahnke, Marko Vujic and Matthias Klusch</i>	27
Room Vargelas - KMIS	27
9: Interaction Patterns in Web-based Knowledge Communities: Two-Mode Network Approach, <i>by Wouter Vollenbroek and Sjoerd de Vries</i>	27
13: Analyzing a Knowledge Country - How is Sweden Managing its Innovation Process?, <i>by Maria Obeso and Maria Jesus Luengo-Valderrey</i>	28
35: Differences between Knowledge and Information Management Practices: Empirical Investigation, <i>by Michal Krčál and Michal Kubiš</i>	28
50: Model-based Strategic Knowledge Elicitation, <i>by J. Pedro Mendes</i>	28
52: Big Data and Knowledge Management: How to Implement Conceptual Models in NoSQL Systems?, <i>by Fatma Abdelhedi, Amal Ait Brahim, Faten Atigui and Gilles Zurfluh</i>	28
7th International Workshop on Software Knowledge (SKY) - Session 3 (15:30 - 16:30)	
Room Bagaúste - SKY	29
3: Fast and Reliable Software Translation of Programming Languages to Natural Language, <i>by Iaakov Exman and Olesya Shapira</i>	29
2: Modulaser: A Tool for Conceptual Analysis of Software Systems, <i>by Iaakov Exman and Phillip Katz</i>	29
Poster Session 1 (16:30 - 17:30)	
Foyer - KDIR	29
3: AToMRS: A Tool to Monitor Recommender Systems, <i>by André Costa, Tiago Cunha and Carlos Soares</i>	29
6: Grammar and Dictionary based Named-entity Linking for Knowledge Extraction of Evidence-based Dietary Recommendations, <i>by Tome Eftimov, Barbara Koroušič Seljak and Peter Korošec</i>	29
12: Skip Search Approach for Mining Probabilistic Frequent Itemsets from Uncertain Data, <i>by Takahiko Shintani, Tadashi Ohmori and Hideyuki Fujita</i>	30
22: Detecting User Emotions in Twitter through Collective Classification, <i>by İbrahim İleri and Pinar Karagoz</i>	30
45: An Analysis of Online Twitter Sentiment Surrounding the European Refugee Crisis, <i>by David Pope and Josephine Griffith</i>	30
66: Using Deep Learning in Arabic-English Cross Language Information Retrieval, <i>by Omar Attia, Michael Azmy, Ahmed Abu Emeira, Karim El Azzouni, Omar Hussein, Nagwa M. El-Makky and Khaled Nagi</i>	30
71: Development of Domains and Keyphrases along Years, <i>by Yaakov HaCohen-Kerner and Meir Hassan</i>	30
72: Gender Clustering of Blog Posts using Distinguishable Features, <i>by Yaakov HaCohen-Kerner, Yarden Tzach and Ori Asis</i>	31

77: The Flightschedule Profiler: An Attempt to Synthetise Visually an Airport's Flight Offer in Time and Space, by Jean-Yves Blaise and Iwona Dudek	31
Foyer - KEOD	31
7: Toward User Profile Representation in Adapted Mediation Systems, by Sara Ouafrouh, Ahmed Zellou and Ali Idri	31
37: Concept-based versus Realism-based Approach to Represent Neuroimaging Observations, by Emna Amdouni and Bernard Gibaud	31
Foyer - KMIS	31
10: Development Results of the Intelligent Device for Storage of the Transfusion Environments Containing Platelets, by A. G. Gudkov, V. Yu. Leushin, A. F. Bobrikhin, V. N. Lemondjava and E. N. Gorlacheva	31
55: A Study on the Public Construction Project Cost-Saving Practice Information Provision Service Method, by Hyun Ok and Tae-Hak Kim	32
Keynote Lecture (17:30 - 18:30)	
Room Pinhão - IC3K	32
STEALTH: Modeling Coevolutionary Dynamics of Tax Evasion and Auditing, by Una-May O'Reilly	32

Thursday Sessions: November 10

7th International Workshop on Software Knowledge (SKY) (09:00 - 09:45)

Room Bagaúste - SKY Intensive School	35
Tutorial: AMG and Applications to Software Knowledge and Software Engineering, by Radel Ben-Av	35

Session 3 (09:00 - 10:30)

Room Mosteiro - KDIR	35
17: Bootstrapping a Semantic Lexicon on Verb Similarities, by Shaheen Syed, Marco Spruit and Melania Borit	35
37: Sentiment Analysis of Breast Cancer Screening in the United States using Twitter, by Kai O. Wong, Faith G. Davis, Osmar R. Zaiane and Yutaka Yasui	35
39: An Active Learning Approach for Ensemble-based Data Stream Mining, by Rabaa Alabdulrahman, Herna Viktor and Eric Paquet	35
40: Prediction of Company's Trend based on Publication Statistics and Sentiment Analysis, by Fumiyo Fukumoto, Yoshimi Suzuki, Akihiro Nonaka and Karman Chan	35
90: Lexicon Expansion System for Domain and Time Oriented Sentiment Analysis, by Nuno Guimaraes, Luis Torgo and Alvaro Figueira	36

Doctoral Consortium on Knowledge Discovery, Knowledge Engineering and Knowledge Management (DC3K) (09:00 - 10:30)

Room Barqueiros - DC3K	36
1: Tacit Knowledge Management in Public Sector - A New Perspective for Organizational Knowledge Management, by Mauro Araújo Câmara	36
2: Barriers and Perceived Benefits to IT Service Management Adoption in SMEs – A Literature Review, by David Molamphy	36
4: Semantic Web Technologies to Enhance the Knowledge Discovery Process in Predictive Analytics, by Iker Esnaola-Gonzalez	36
5: Semantic Coverage Measures: Analytic Operators for Ontologists, by Pallavi Karanth and Kavi Mahesh	37

European Project Space - Session (09:00 - 10:30)

Room Pinhão - IC3K	37
CLAFIS Project - Intelligent integrated solution for communication between automation systems and IT systems in farms process, by Shalom Broyer	37
CLAFIS Project - Knowledge Processing Framework and Security Implementations, by Josef Küng	37
CLAFIS Project - Data Fusion Framework for Agricultural Services, by Bijjeet Kaur	37
System Self-Awareness Towards Deep Learning and Discovering High-Value Information, by Ying Zhao	37

7th International Workshop on Software Knowledge (SKY) (09:45 - 10:30)

Room Bagaúste - SKY Intensive School	37
Mining Source Code Regularities with FCA and Association Rules, by Daniel Speicher	37

7th International Workshop on Software Knowledge (SKY) (10:45 - 11:30)

Room Bagaúste - SKY Intensive School	37
Systems Engineering in the industrial environment: Requirements Engineering Processes, by Anabel Fraga	37

Parallel Session 4 (10:45 - 12:15)

Room Mosteiro - KDIR	37
46: Unsupervised Irony Detection: A Probabilistic Model with Word Embeddings, by Debora Nozza, Elisabetta Fersini and Enza Messina	37
55: Discovering Data Lineage from Data Warehouse Procedures, by Kalle Tomingas, Priit Järv and Tanel Tammet	38

59: A Linear-dependence-based Approach to Design Proactive Credit Scoring Models, <i>by Roberto Saia and Salvatore Carta</i>	38
Room Vargelas - KMIS	38
28: Knowledge Management Framework for Early Phases in TOGAF-based Enterprise Architecture, <i>by Juan Pablo Meneses-Ortegón and Rafael A. Gonzalez</i>	38
29: Business Intelligence Solution for an SME: A Case Study, <i>by Raghavendra Raj, Shun Ha Sylvia Wong and Anthony J. Beaumont</i>	38
45: Organizational Records Systems - An Alternative View to (Enterprise) Information Systems, <i>by Sherry Li Xie and Guanyan (Amelie) Fan</i>	39
7th International Workshop on Software Knowledge (SKY) (11:30 - 12:15)	
Room Bagaúste - SKY Intensive School	39
Conceptual Integrity of Software Systems, An Overview, <i>by laakov Exman</i>	39
Keynote Lecture (12:15 - 13:15)	
Room Pinhão - IC3K	39
From Open Data to Knowledge: Capitalizing Experiences, <i>by Marijn Janssen</i>	39
European Project Space - Panel (14:30 - 15:30)	
Room Pinhão - IC3K: H2020 Knowledge and Intelligent Systems: the Good, the Bad and the Disruptive	39
Parallel Session 5 (15:30 - 16:30)	
Room Mosteiró - KDIR	39
23: Enriching What-If Scenarios with OLAP Usage Preferences, <i>by Mariana Carvalho and Orlando Belo</i>	39
42: Explanation Retrieval in Semantic Networks - Understanding Spreading Activation based Recommendations, <i>by Vanessa N. Michalke and Kerstin Hartig</i>	39
65: Unsupervised Classification of Opinions, <i>by Itu Vlad Vasile, Rodica Potolea and Mihaela Dinsoreanu</i>	40
Room Vargelas - KMIS	40
20: Challenges and Practices for Effective Knowledge Transfer in Globally Distributed Teams - A Systematic Literature Review, <i>by Josiane Kroll, Juho Mäkiö and Manal Assaad</i>	40
42: How Do Young Researchers Take the Steps toward Startup Activities? - A Case Study of a One-day Workshop for Entrepreneur Education, <i>by Miki Saijo, Makiko Watanabe, Takumi Ohashi, Haruna Kusu, Hikaru Tsukagoshi and Ryuta Takeda</i>	40
56: Towards Comprehensive Security Related Pedagogy - An Approach to Learning and Resilience, <i>by Rauno Pirinen, Juha Mäkinen and Arto Salonen</i>	40
European Project Space - Session (15:30 - 16:30)	
Room Régua - IC3K	40
Context-Dependent Associative Learning, <i>by Oussama H. Hamid</i>	40
OpenSkiMr - Open Skill Match Maker, <i>by Andreas Kofler and Marianne Prast</i>	41
Robust Risk-Based Strategy Design, <i>by J.Pedro Mendes</i>	41
Poster Session 2 (16:30 - 17:30)	
Foyer - KDIR	41
5: Discovering the Deep Web through XML Schema Extraction, <i>by Yasser Saissi, Ahmed Zellou and Ali Idri</i>	41
29: Graph-based Rating Prediction using Eigenvector Centrality, <i>by Dmitry Dolgikh and Ivan Jelínek</i>	41
50: Efficient Social Network Multilingual Classification using Character, POS <i>n</i> -grams and Dynamic Normalization, <i>by Carlos-Emiliano González-Gallardo, Juan-Manuel Torres-Moreno, Azucena Montes Rendón and Gerardo Sierra</i>	41
61: The Longest Common Subsequence Distance using a Complexity Factor, <i>by Octavian Lucian Hasna and Rodica Potolea</i>	41
64: Syllabification with Frequent Sequence Patterns - A Language Independent Approach, <i>by Adrian Bona, Camelia Lemnaru and Rodica Potolea</i>	41
84: Exploring Urban Tree Site Planting Selection in Mexico City through Association Rules, <i>by Héctor Javier Vázquez and Mihaela Juganaru-Mathieu</i>	42
86: Computing Ideal Number of Test Subjects - Sensorial Map Parametrization, <i>by Jessica Dacleu Ndengue, Mihaela Juganaru-Mathieu and Jenny Faucheu</i>	42
Foyer - KEOD	42
2: The Federated Ontology of the PAL Project - Interfacing Ontologies and Integrating Time-dependent Data, <i>by Hans-Ulrich Krieger, Riica Peters, Bernd Kiefer, Michael A. van Bekkum, Frank Kaptein and Mark A. Neerincx</i>	42
20: Towards a Semantic Approach for the Design of Social Network Users' Geographical Trajectories, <i>by Hadhami Ounissi, Marwa Manaa and Jalel Akaichi</i>	42
Foyer - KMIS	43
14: Pro-Innovative Orientation of Polish Social Enterprises: The Empirical Perspective, <i>by Agata Sudolska and Monika Chodorek</i>	43
27: A Survey of Open Government Data in Russian Federation, <i>by Dmitriy Koznov, Olga Andreeva, Uolevi Nikula, Andrey Maglyas, Dmitry Muromtsev and Irina Radchenko</i>	43

36: A Tracing System for User Interactions towards Knowledge Extraction of Power Users in Business Intelligence Systems, by <i>Safwan Sulaiman, Tariq Mahmoud, Stephan Robbers, Jorge Marx Gómez and Joachim Kurzhöfer</i>	43
54: A Plan for Constructing an Information Service System for Construction Project Management using GIS Information, by <i>SeongJin Kim and Hyun Ok</i>	43
58: KPIs 4 Workplace Learning, by <i>Sandro Emmenegger, Knut Hinkelmann, Barbara Thönssen and Frieder Witschel</i>	44
Foyer - EPS	44
CLAFIS Project - Implementing Agricultural Machine Remote Diagnostics over CLAFIS platform, by <i>Ari Ronkainen</i>	44
CLAFIS Project - Integrated System for Farm Operations - EU FP7 Funded Project, by <i>Tomasz Blaszczyk</i>	44
OpenSkiMr-Open Skill Match Maker, by <i>Marianne Prast</i>	44
OSTAG - Ontology-based Software Test Case Generation, by <i>Anders Adlemo</i>	44
Invited Talk (17:30 - 18:30)	
Room Pinhão - IC3K	44
Horizon 2020's SME Instrument, by <i>Daniel Chirtes</i>	44

Friday Sessions: November 11

Parallel Session 6 (10:45 - 12:15)

Room Mosteiró - KDIR	47
18: Heterogeneous Ensemble for Imaginary Scene Classification, by <i>Saleh Alyahyan, Majed Farrash and Wenjia Wang</i>	47
34: Result Diversity for RDF Search, by <i>Hiba Arnaout and Shady Elbassuoni</i>	47
60: Efficient Distraction Detection in Surveillance Video using Approximate Counting over Decomposed Micro-streams, by <i>Avi Bleiweiss</i>	47
79: Evaluation of Statistical Text Normalisation Techniques for Twitter, by <i>Phavanh Sosamphan, Veronica Liesaputra, Sira Yongchareon and Mahsa Mohaghegh</i>	47
Room Aregos - KEOD	47
12: An Ontology-enabled Context-aware Learning Record Store Compatible with the Experience API, by <i>Jonas Anseeuw, Stijn Verstichel, Femke Ongenaes, Ruben Lagatie, Sylvie Venant and Filip De Turck</i>	47
22: Information Search in Ontology Visualization – An Eyetracking User Study of Indented List on Desktop and Tablet Computers, by <i>Anh Huynh and Bo Fu</i>	48
28: Qualitative Reasoning for Understanding the Behaviour of Complex Biomolecular Networks, by <i>Ali Ayadi, Cecilia Zanni-Merk and François de Beuvron de Bertrand</i>	48
44: Avionics Maintenance Ontology Building for Failure Diagnosis Support, by <i>Luis Palacios, Gaëlle Lortal, Claire Laudy, Christian Sannino, Ludovic Simon, Giuseppe Fusco, Yue Ma and Chantal Reynaud</i>	48
Room Vargelas - KMIS	48
44: How Can We Facilitate the Diffusion of Electric Vehicles in Japan? - Consumer Characteristics and Perceived Innovation Attributes, by <i>Taiyu Sato and Miki Saijo</i>	48
31: The STAPS Method - Process-tailored Introduction of Knowledge Management Solutions, by <i>Christoph Sigmanek and Birger Lantow</i>	49
60: A State Transition Model for Business Processes - Towards Object-oriented Business Process Automation, by <i>Anantharamaiah Prabhakar</i>	49
Room Vargelas - Open Communications	49
4: Is Knowledge a strategic and unavoidable resource? : from the VRIN framework to organizational performance, by <i>Abdeslam Jajaf</i>	49
5: Using tablet devices to foster KM processes within working contexts in France: towards a new approach, by <i>Mehdi EL ABED, Abdeslam Jajaf and Jean-François LEMOINE</i>	49

Keynote Lecture (12:15 - 13:15)

Room Pinhão - IC3K	50
No Smartness Without Data, by <i>Dieter A. Fensel</i>	50

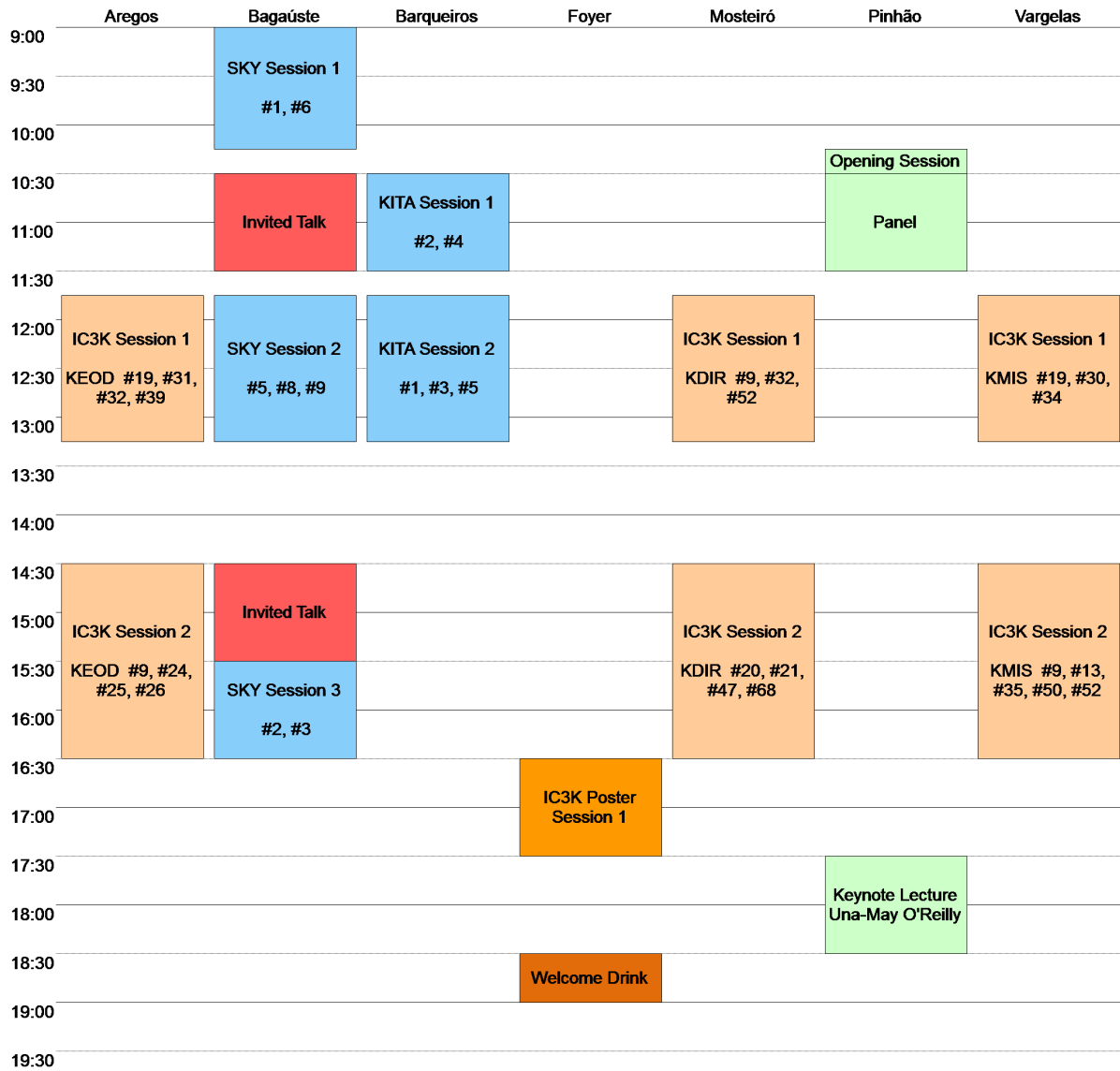
Parallel Session 7 (14:30 - 16:30)

Room Mosteiró - KDIR	50
10: Putting Web Tables into Context, by <i>Katrin Braunschweig, Maik Thiele, Elvis Koci and Wolfgang Lehner</i>	50
11: DTATG: An Automatic Title Generator based on Dependency Trees, by <i>Liqun Shao and Jie Wang</i>	50
28: Summarization of Spontaneous Speech using Automatic Speech Recognition and a Speech Prosody based Tokenizer, by <i>György Szaszák, Máté Ákos Tündik and András Beke</i>	50
58: Introducing the Key Stages for Addressing Multi-perspective Summarization, by <i>Elena Lloret</i>	51
63: Learning Diagnosis from Electronic Health Records, by <i>Ioana Barbantan and Rodica Potolea</i>	51
76: Power to the People! - Meta-Algorithmic Modelling in Applied Data Science, by <i>Marco Spruit and Raj Jagesar</i>	51
Room Aregos - KEOD	51
6: Reconstruction of Implied Semantic Relations in Russian Wiktionary, by <i>Serge Klimenkov, Evgenij Tsopa, Alexey Pismak and Alexander Yarkeev</i>	51
17: Unfolding Existentially Quantified Sets of Extended Clauses, by <i>Kiyoshi Akama and Ekawit Nantajeewarawat</i>	52
21: SOCUDO-SCACLO: Ontologies for Socio-cultural Aware Collaborative Learning, by <i>Fadoua Ouamani, Narjès Bellamine Ben Saoud and Henda Hajjami Ben Ghézala</i>	52

35: A Bayesian Approach for Weighted Ontologies and Semantic Search, by Anna Formica, Michele Missikoff, Elaheh Pourabbas and Francesco Taglino	52
38: OntoMetrics: Putting Metrics into Use for Ontology Evaluation, by Birger Lantow	52
Room Vargelas - KMIS	53
5: Mobile Data Store Platforms: Test Case based Performance Evaluation, by Karim Kussainov and Bolatzhan Kumalakov	53
11: Big Data Knowledge Service Framework based on Knowledge Fusion, by Fei Wang, Hao Fan and Gang Liu	53
15: Ensuring Action: Identifying Unclear Actor Specifications in Textual Business Process Descriptions, by Ulf Sanne, Hans Friedrich Witschel, Alessio Ferrari and Stefania Gnesi	53
26: Managing Heterogeneous Information in a System of Information Systems, by Majd Saleh, Véronique Misséri and Marie-Hélène Abel	53
48: Categorization and Matching for Drone-based Services, by Simona Ibbi, Filippo Eros Pani and Alberto buschetti	54
Poster Session 3 (16:30 - 17:30)	
Foyer - KDIR	54
16: Success Prediction System for Student Counseling using Data Mining, by Jörg Frochte and Irina Bernst	54
30: Adopting Privacy Regulations in a Data Warehouse - A Case of the Anonymity versus Utility Dilemma, by Chaïm van Toledo and Marco Spruit	54
33: A Conceptual Model of Actors and Interactions for the Knowledge Discovery Process, by Lauri Tuovinen	54
36: Exploration and Visualization of Big Graphs - The DBpedia Case Study, by Enrico G. Calderola, Antonio Picariello, Antonio M. Rinaldi and Marco Sacco	55
53: Recommending Groups to Users based Both on Their Textual and Image Posts, by Elias Oliveira, Howard Roatti, Gustavo Ramos Lima and Patrick Marques Ciarelli	55
81: Rating Prediction with Contextual Conditional Preferences, by Aleksandra Karpus, Tommaso Di Noia, Paolo Tomeo and Krzysztof Goczyla	55
85: Anti-cancer Drug Activity Prediction by Ensemble Learning, by Ertan Tolan and Mehmet Tan	55
87: Big Data and Deep Analytics Applied to the Common Tactical Air Picture (CTAP) and Combat Identification (CID), by Ying Zhao, Tony Kendall and Bonnie Johnson	55
88: SPARQL Query Generation based on RDF Graph, by Mohamed Kharrat, Anis Jedidi and Faiez Gargouri	56
89: A Novel Clustering Algorithm to Capture Utility Information in Transactional Data, by Piyush Lakhawat, Mayank Mishra and Arun Somani	56
Foyer - KEOD	56
18: Ontological Approach to Share Product Design Semantics for an Assembly, by Baha Hasan, Jan Wikander and Mauro Onori	56
29: Enhancing Community Detection in Social Network using Ontology, by Salma Khattab, Abeer ElKorany and Akram Salah	56
40: Temporal Interval Modeling for UML Activity Diagrams, by Joanna Isabelle Olszewska	57
Foyer - KMIS	57
6: Everyday Lives and Practices of a Joint Venture in the Iranian Construction Sector, by Laurent Scaringella and François Burtschell	57
17: Exploration Participants Engagement in Organisational Knowledge Sharing, by Mohammadbashir Sedighi, Stephan Lukosch, Sander van Splunter, Frances Brazier and Cees van Beers	57
38: Model of Organizational Readiness to Implement Mentoring, by Malgorzata Baran	57
Keynote Lecture (17:30 - 18:30)	
Room Pinhão - IC3K	57
Mining Satellite Images for Census Data Collection: A Study Using the Google Static Maps Service, by Frans Coenen	57
Closing Session (18:30 - 18:45)	
Room Pinhão - IC3K	58

Wednesday Sessions: November 9

Wednesday Sessions: November 9 Program Layout



Workshop - Session 1
09:00 - 10:15
SKY

SKY
Room Bagaúste

Paper #1

Development and Evaluation of a Software Requirements Ontology

He Tan¹, Muhammad Ismail¹, Vladimir Tarasov¹, Anders Adlemo¹ and Mats Johansson²

¹ Jönköping University, Jönköping, Sweden

² Saab Avionics Systems, Jönköping, Sweden

Keywords: Ontology, Software Requirements, Ontology Development, Ontology Evaluation, Avionics Software Development.

Abstract: This paper presents an ontology which has been developed to represent the requirements of a software component pertaining to an embedded system in the avionics industry. The ontology was built based on the software requirements documents and was used to support advanced methods in the subsequent stages of the software development process. In this paper it is described the process that was used to build the ontology. Two pertinent quality measures that were applied to the ontology, i.e. usability and applicability, are also described, as well as the methods used to evaluate the quality measures and the result of these evaluations.

Paper #6

Syntactic-Semantic Extraction of Patterns Applied to the US and European Patents Domain

Anabel Fraga¹, Juan Llorens², Eugenio Parra², Leticia Arroyo² and Valentín Moreno²

¹ Carlos III University of Madrid, Leganes, Spain

² Carlos III of Madrid University, Madrid, Spain

Keywords: Indexing, Ontologies, Knowledge, Patterns, Reuse, Retrieval, Patents, US Patents, European Patents.

Abstract: Nowadays, there are many scientific inventions referring to any topic like medicine, technology, economics, finance, banking, computer science, and so on. These inventions are suggested as patents to the agencies working in US and Europe for the registration and revision of the patent applications. But, the job of reviewing the patents might be complicated because every day the quantity of it is bigger and bigger. And also, the amount of work dedicated writing a proper application might be intricate and needs several revisions from investor and examiners. This revision job might have costs for the inventor because they don't know the proper language for writing the application in the formal mode used. As part of a solution, one approach to minimize the impact of this fact and increase the success of the reviewing process is aid the human reviewer and also inventors with a set of patterns created using Natural Language Processing techniques that accelerate the review just looking in the massive set of registration any similar one already patented and on the other hand aid the inventor writing in the formal manner the application.

Opening Session
10:15 - 10:30

IC3K
Room Pinhão

Panel
10:30 - 11:30

IC3K
Room Pinhão

10:30 - 11:30

SKY Invited Talk
Room Bagaúste

Concept Lattices as a Versatile Tool to Discover Knowledge in Software

Daniel Speicher

University of Bonn, Germany

Abstract: Concept Lattices in the sense of Formal Concept Analysis have been used in many different areas of Software Engineering, but most prominently in Reverse Engineering in an attempt to rediscover concepts of different kinds from existing systems. Once we extract a meaningful binary relation from a system we have a rich and pragmatic theory at hand to explore the inner structure of this relation. Interesting relations range from compile-time or run-time dependencies over direct or indirect usage relations to co-occurrence of terms in identifiers and message invocations. More advanced relations take a few program elements simultaneously into account. Formal Concept Analysis offers us then a unifying perspective to explore concepts as clusters or classes, to explore implications or associations rules as implementation regularities, where the regularities might be useful implementation idioms while the outliers might be refactoring opportunities. Finally Attribute Exploration gives us a tool to review whether the implications found in the system under consideration are generic.

Workshop - Session 1
10:30 - 11:30

KITA
Room Barqueiros

Paper #2

Integrating Knowledge Artifacts and Inertial Measurement Unit Sensors for Decision Support

Stefano Pinardi, Fabio Sartori and Riccardo Melen

University of Milano-Bicocca, Milan, Italy

Keywords: Inertial Measurement Unit, Android, Wearable Devices, Recommendation Systems, Knowledge Artifacts.

Abstract: Modern wearable devices provide new opportunities for the development of knowledge artifacts and decision support systems. In this paper we present a recent development of KAFKA, a knowledge engineering methodology based on knowledge artifact notion, that make it able to manage real-time data detected and analyzed by means of Inertial Measurement Units sensors, mounted on Android wearables. This improvement makes KAFKA suitable to deal with many domains where real-time data are necessary, in particular the health-care and rehabilitation ones.

Paper #4

Making and Understanding A Vision for IoT Makerspaces

Julian Dax and Volkmar Pipek

University of Siegen, Siegen, Germany

Keywords: Making, IoT, Embedded Devices, End-User Development, End-User Software Engineering, Infrastructuring.

Abstract: We present a vision for the IoT makerspace of the future. Currently, makers design their spaces with a focus on building (or making), but the core challenge they face in the IoT era is understanding. In our vision this is archived by gathering data about the IoT devices and their environment, storing that data in a central repository, consolidating it and making it easily accessible. We also describe the first steps we took towards this vision.

Parallel Session 1
11:45 - 13:15

KDIR
Room Mosteiró

Paper #9

StorylineViz: A [Space, Time, Actors, Motion] Segmentation Method for Visual Text Exploration

Iwona Dudek and Jean-Yves Blaise
UMR CNRS/MCC 3495 MAP, Marseille, France

Keywords: Visualisation, Knowledge Modelling, Sensemaking, Spatio-Temporal Data, Textual Content, Narratives.

Abstract: Supporting knowledge discovery through visual means is a hot research topic in the field of visual analytics in general, and a key issue in the analysis of textual data sets. In that context, the StorylineViz study aims at developing a generic approach to narrative analysis, supporting the identification of significant patterns inside textual data, and ultimately knowledge discovery and sensemaking. It builds on a text segmentation procedure through which sequences of situations are extracted. A situation is defined by a quadruplet of components: actors, space, time and motion. The approach aims at facilitating visual reasoning on the structure, rhythm, patterns and variations of heterogeneous texts in order to enable comparative analysis, and to summarise how the space/time/actors/motion components are organised inside a given narrative. It encompasses issues that are rooted in Information Sciences - visual analytics, knowledge representation – and issues that more closely relate to Digital Humanities – comparative methods and analytical reasoning on textual content, support in teaching and learning, cultural mediation.

Paper #32

A Multi-Layer System for Semantic Textual Similarity

Ngoc Phuoc An Vo¹ and Octavian Popescu²

¹ Xerox Research Centre Europe, Meylan, France

² IBM T.J.Watson Research, YorkTown, U.S.A.

Keywords: Machine Learning, Natural Language Processing (NLP), Semantic Textual Similarity (STS).

Abstract: Building a system able to cope with various phenomena which falls under the umbrella of semantic similarity is far from trivial. It is almost always the case that the performances of a system do not vary consistently or predictably from corpora to corpora. We analyzed the source of this variance and found that it is related to the word-pair similarity distribution among the topics in the various corpora. Then we used this insight to construct a 4-module system that would take into consideration not only string and semantic word similarity, but also word alignment and sentence structure. The system consistently achieves an accuracy which is very close to the state of the art, or reaching a new state of the art. The system is based on a multi-layer architecture and is able to deal with heterogeneous corpora which may not have been generated by the same distribution.

Paper #52

Automatic Text Summarization by Non-topic Relevance Estimation

Ignacio Arroyo-Fernández^{1,2}, Juan-Manuel Torres-Moreno², Gerardo Sierra¹ and Luis Adrián Cabrera-Diego²

¹ UNAM, Mexico City, Mexico

² UAPV, Avignon, France

Keywords: Automatic Text Summarization, Machine Learning, Generalization Ability, Regression Estimation.

Abstract: We investigate a novel framework for Automatic Text Summarization. In this framework underlying language-use features are learned from a minimal sample corpus. We argue the low complexity of this kind of features allows relying in generalization ability of a learning machine, rather than in diverse human-abstracted summaries. In this way, our method reliably estimates a relevance measure for predicting summary candidature scores, regardless topics in unseen documents. Our output summaries are comparable to the state-of-the-art. Thus we show that in order to extract meaning summaries, it is not crucial what is being said; but rather how it is being said.

Parallel Session 1
11:45 - 13:15

KEOD
Room Aregos

Paper #19

Verifying and Mapping the Mereotopology of Upper-Level Ontologies

Lydia Silva Muñoz and Michael Grüninger

University of Toronto, Toronto, Canada

Keywords: DOLCE, DOLCE-CORE, SUMO, Ontology Mapping, Ontology Verification, Upper-Level Ontology, Mereology, Topology, Mereotopology.

Abstract: Upper-level ontologies provide an account of the most basic, domain-independent, existing entities, such as time, space, objects, and processes. Ontology verification is the process by which a theory is checked to rule out unintended models, and possibly characterize missing intended ones. In this paper, we verify the core characterization of mereotopology of the Suggested Upper Merged Ontology (SUMO), and the mereology of the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), while relating their axiomatizations via ontology mapping. As a result, we propose the correction and addition of some axioms to the analyzed theories which eliminate unintended models and characterize missing ones. In addition, we show by formal means which is the relation existing between the axiomatization of mereology in both upper-level ontologies, and make available a modular representation in first-order logic of the SUMO characterization of mereotopology.

Paper #31

Developing a Model of Agreement Negotiation Dialogue

Mare Koit

University of Tartu, Tartu, Estonia

Keywords: Dialogue Structure, Dialogue Act, Dialogue Corpus, Knowledge Representation.

Abstract: We are investigating human-human dialogues in the Estonian dialogue corpus with the further aim to develop a dialogue system which carries out negotiations with a user in a natural language. Two sub-corpora are analysed and compared: (1) MSN conversations, and (2) everyday dialogues, both phone calls and face-to-face conversations. In the dialogues, the participants are trying to achieve an agreement about doing an action. The structure of negotiations is represented as a sequence of dialogue acts. A special case of negotiation – debate where the participants have contradictory communicative goals – has been implemented as an experimental dialogue system.

Paper #32

Data Integration and Visualization for Knowledge Mapping in Strasbourg University

Amira Essaid¹, Quynh Nguyen Thi² and Cecilia Zanni-Merk¹¹ *INSA de Strasbourg, ICUBE laboratory, Illkirch, France*² *Strasbourg University, Strasbourg, France*

Keywords: Ontology, Data Integration, Data Visualization, Intelligibility.

Abstract: The work described in this paper is part of the IDEX (excellence initiative) project "Complex Identities" launched by Strasbourg University in 2015. The main goal is to map available knowledge in Strasbourg university in order to provide a comprehensive and structured view of its different components. Our approach consists, first, in building an ontology able to represent available knowledge in the university, making it understandable by users. Then, we are interested in visualizing the ontology to help users explore easily the represented knowledge.

Paper #39

Empowering the Model-driven Engineering of Robotic Applications using Ontological Semantics and Reasoning

Stefan Zander¹, Nadia Ahmed¹ and Yingbing Hua²¹ *FZI Research Center for Information Technology, Karlsruhe, Germany*² *Karlsruhe Institute of Technology, Karlsruhe, Germany*

Keywords: Knowledge Representation, Semantic Technologies, Ontologies, Cyber-physical Systems, Robotics.

Abstract: This work discusses two scenarios in which the model-driven engineering of robotic applications can be improved using ontological semantics and reasoning. The objective of the presented approach is to facilitate reuse and interoperability between cooperating software and hardware components. Central to the presented approach is the usage of ontologies and

description logics as knowledge representation frameworks for the axiomatic description of component metadata models. In the first scenario, we show how application templates can be created using the concept of placeholders in which requirements for integrating external components can be axiomatically specified and eligible components can be computed using subsumption reasoning. The second scenario extends this idea for the inference of compatibilities between cooperating components. The practical applicability of the approach is demonstrated by a concrete use case from the ReApp project.

Parallel Session 1
11:45 - 13:15

KMIS
Room Vargelas

Paper #19

Enabling Centralised Management of Local Sensor Data Refinement in Machine Fleets

Petri Kannisto and David Hästbacka

Tampere University of Technology, Tampere, Finland

Keywords: Distributed Knowledge Management, Mobile Machinery.

Abstract: In modern mobile machines, a lot of measurement data is available to generate information about machine performance. Exploiting it locally in machines would enable optimising their operation and, thus, yield competitive advantage and reduce environmental load due to reduced emissions. However, optimisation requires extensive knowledge about machine performance and characteristics in various conditions. As physical machines may be located geographically far from each other, the management of ever evolving knowledge is challenging. This study introduces a software concept to enable centralised management of data refinement performed locally in the machines of a geographically distributed fleet. It facilitates data utilisation in end user applications that provide useful information for operators in the field. Whatever the further data analysis requirements are, multiple preprocessing tasks are performed: it enables outlier limit configuration, the calculation of derived variables, data set categorisation and context recognition. A functional prototype has been implemented for the refinement of real operational data collected from forestry machines. The results show that the concept has considerable potential to bring added value for enterprises due to improved possibilities in managing data utilisation.

Paper #30

Dual-Priority Congestion Control Mechanism for Video Services

Real Network Tests of CVIHIS

Juha Vihervaara, Pekka Loula and Teemu Alapahuoma

Tampere University of Technology, Pori, Finland

Keywords: Congestion Control, Video Traffic, Priority.

Abstract: Information can be shared effectively by using of videos. Therefore, it is no wonder that videos form most of the Internet traffic. For the efficient operation of the Internet, it is essential that these videos are shared with a proper and effective way. In our previous paper, we have already presented a congestion control mechanism for the proper transmission of videos. This mechanism can offer two priority levels for video. There is a low priority service where the bandwidth is given away to other connections after the load level of a network exceeds a

certain level. Instead, the other priority level, a real-time mode, always wants its fair share of the bandwidth. In this study, we have tested the operation of this mechanism by doing real network tests.

Paper #34

Coordination Problems in Knowledge Transfer: A Case Study of Inter-Organizational Projects

Nestor A. Nova and Rafael A. Gonzalez
Pontificia Universidad Javeriana, Bogotá, Colombia

Keywords: Coordination, Knowledge Transfer, Knowledge Management, Information Processing, Case Study.

Abstract: When multiple organizations are involved in a heritage management project, the coordination of actions is complex and can affect the knowledge transfer process. This paper contributes a systematic and empirical study of the dynamics of coordination activities inside a knowledge transfer process in heritage management activities. Using the information-processing view of coordination, we explore the following question: what kinds of coordination issues affect effective coordination of knowledge transfer in inter-organizational projects? The discussion is supported by a case study in the architectural heritage domain. We reveal that there are many coordination issues that affect the mutual understanding between actors, limiting information exchange and knowledge transfer. These issues uncover a gap between the conception and use of ICTs that support coordination, and a lack of understanding about how ICT usage affects the knowledge transfer process. Thus, a socio-material perspective about relationship between people and coordination technologies could improve knowledge transfer performance.

Workshop - Session 2
11:45 - 13:15

KITA
Room Barqueiros

Paper #1

Knowledge-Oriented Technologies & Network Marketing Direct Selling Organizations (NMDSO) *Some Preliminary Insights into the Nature and the Goals of Shared Knowledge*

Carolina Guerini and Eliana Alessandra Minelli
Università Carlo Cattaneo – LIUC, Castellanza, Italy

Keywords: Business Models, Knowledge Artifact, Knowledge Sharing in Network Marketing Direct Selling Organization.

Abstract: The work extends the sales and organization literature by analyzing the nature and the goals of knowledge sharing within networkers' downlines in Network Marketing Direct Selling Organizations (NMDSO). The main results of the research, based on a qualitative methodology and referred to Lyoness network in Italy, acknowledge the relevance of knowledge sharing & creation via digital technology, distinguish the nature of knowledge sharing and identify the main goals.

Paper #3

Makers in the Plant? Exploring the Impact of Knowledge IT Artifacts on DIY Practices in Manufacturing Firms

Luca Cremona and Aurelio Ravarini
Università Carlo Cattaneo – LIUC, Castellanza (VA), Italy

Keywords: Makers, DIY, Digital Fabrication, Digital DIY.

Abstract: In this study we investigate the impact of digital technologies on fabrication activities carried out by a worker leading her organizational role to be critically reshaped. We assume that the characteristics of the Makers (individual and environmental characteristics) could be applied to workers in a manufacturing plant, bringing benefits in terms of higher achievements deriving by the digitization of fabrication. We propose to interpret the digital technologies enabling digitization through the lens of the KITA construct. Two case studies have been carried out in order to explore these assumptions and providing preliminary insights of the effects of Digital DIY practices on manufacturing firms.

Paper #5

More Time for the Doing, Having *Made* the Thinking *3D Printing for Knowledge Circulation in Healthcare*

Federico Cabitza^{1,2}, Angela Locoro³, Aurelio Ravarini⁴
and Vittorio Satta⁴

¹ *Università degli Studi di Milano-Bicocca, Milano, Italy*

² *IRCCS Istituto Ortopedico Galeazzi, Milano, Italy*

³ *Università degli Studi Milano-Bicocca, Milano, Italy*

⁴ *Università Carlo Cattaneo, Castellanza, Italy*

Keywords: DiDIY, Medical 3D Printing, Diagnostic, Training, and Surgical Tools, Knowledge Circulation in Healthcare.

Abstract: This paper investigates the phenomenon of the Digital Do-It-Yourself (Di-DIY) in the medical domain. In particular, the main contribution of the paper is a conceptual framework based on the notion of DiDIY in healthcare. To help focus on the main actors and assets composing the 3D printing innovation roles in healthcare we model: the DiDIY-er as the main initiator of the practice innovation; the available technology allowing the envisioning of new practices; the specific activities gaining benefits from the innovative techniques introduced; and the knowledge community continuously supporting and evolving knowledge practices. A general introduction on the notion of Knowledge Artifacts (KAs) and on the use of 3D printing (3DP) in medicine will be followed by our research questions and by a more detailed analysis of diagnostic, training and surgical planning activities for clinicians and patients. Observations carried out in a hospital in Italy are reported to exemplify activities based on 3DP bone models in the radiological and orthopaedic fields. These observations can be considered a second contribution of the paper, although secondary with respect to the conceptual framework. They also help proof how knowledge sharing and circulation in the community of healthcare professionals may be improved by the introduction of tangible and intangible KAs around the practice of DiDIY. Our framework is then presented in the end.

Workshop - Session 2
11:45 - 13:15
SKY

SKY
Room Bagaúste

Paper #5

The Modularity Matrix as a Source of Software Conceptual Integrity

Iaakov Exman

The Jerusalem College of Engineering - JCE - Azrieli, Jerusalem, Israel

Keywords: Conceptual Integrity, Modularity Matrix, Conceptual Lattice, Linear Software Models, Liskov Substitution Principle, Abstract Mathematical Concepts, Standard Modularity Matrix, Software System Design.

Abstract: Conceptual Integrity has been declared the most important consideration for software system design. However, the very concept of Conceptual Integrity remained quite vague, lacking a precise formal definition. This paper offers a path to a novel definition of Conceptual Integrity in terms of the Modularity Matrix, the basic structure of Linear Software Models. We provide arguments for the plausibility of the Modularity Matrix as the suggested source of software system Conceptual Integrity, viz. the orthogonality and propriety of the Matrix modules. Furthermore, the paper also reveals some new characteristic properties of Software Conceptual Integrity.

Paper #9

Requirements Engineering: More Is Not Necessarily Better

Gonzalo Génova

Universidad Carlos III de Madrid, Leganés, Spain

Keywords: Requirements Engineering, Software Development Process, Software Economics.

Abstract: We show that documenting requirements (and, in general, requirements engineering) is profitable for the project, but not as profitable as to consider that "the more, the better".

Paper #8

Modeling Semantics sans Mathematical Formalism

Reuven Gallant

JCT Lev Academic Center, Jerusalem, Israel

Keywords: Executable Modeling, Formal Methods, Graphical Modeling Languages, Production Quality Code, Syntax, Semantics, Semantic Mapping, Simulation, UML.

Abstract: Much ink has been spilled regarding the trials and tribulations of adapting formal methods to the needs of software engineering practitioners. With the exception of computer scientists with a passion for algorithm design and optimization, a plethora of Greek letters and symbols can be an anathema to those whose first love is writing code. The advent of graphical modeling languages such as UML, and supporting tools that generate production quality code, executable modeling behavioral simulations for bridging the gap between formalism and coding. This paper proposes, with illustrative examples, an exploratory

learning modality, by which the practicing engineer can investigate and empirically learn the semantic mapping of UML syntax to the semantic domains of system instantiation and reactive behavior.

14:30 - 15:30

SKY Invited Talk
Room Bagaúste

Some Blends are Better than Others: Blend Quality Assessment in Computational Models of Concept Blending

Pedro Martins

University of Coimbra, Portugal

Abstract: Concept blending (CB) is a cognitive mechanism by which two or more mental spaces are integrated to produce new concepts. The theory built around CB provides not only a set of sound principles but also a consistent terminology that can be used in creativity modeling. This has been a major motivation to design artificial creative systems based on computational approaches to CB. In this talk, I will overview some computational models of CB theory and present the current challenges in the design of such models. A special emphasis will be given to our research on the quality assessment of blends, including a recent study on the relevance of the optimality principles.

Parallel Session 2
14:30 - 16:30

KDIR
Room Mosteiro

Paper #20

Keyword-based Approach for Lyrics Emotion Variation Detection

Ricardo Malheiro^{1,2}, Hugo Gonçalo Oliveira¹, Paulo Gomes¹ and Rui Pedro Paiva¹

¹ *Center for Informatics and Systems of the University of Coimbra (CISUC), Coimbra, Portugal*

² *Miguel Torga Higher Institute, Coimbra, Portugal*

Keywords: Music Information Retrieval, Lyrics Music Emotion Recognition, Lyrics Music Emotion Variation Detection, Keyword based Approach.

Abstract: This research addresses the role of the lyrics in the context of music emotion variation detection. To accomplish this task we create a system to detect the predominant emotion expressed by each sentence (verse) of the lyrics. The system employs Russell's emotion model and contains 4 sets of emotions associated to each quadrant. To detect the predominant emotion in each verse, we propose a novel keyword-based approach, which receives a sentence (verse) and classifies it in the appropriate quadrant. To tune the system parameters, we created a 129-sentence training dataset from 68 songs. To validate our system, we created a separate ground-truth containing 239 sentences (verses) from 44 songs annotated manually with an average of 7 annotations per sentence. The system attains 67.4% F-Measure score.

Paper #21

Classification and Regression of Music Lyrics: Emotionally-Significant Features

Ricardo Malheiro^{1,2}, Renato Panda¹, Paulo Gomes¹ and Rui Pedro Paiva¹

¹ Center for Informatics and Systems of the University of Coimbra (CISUC), Coimbra, Portugal

² Miguel Torga Higher Institute, Coimbra, Portugal

Keywords: Music Information Retrieval, Lyrics Music Emotion Recognition, Lyrics Music Classification, Lyrics Music Regression, Lyrics Feature Extraction.

Abstract: This research addresses the role of lyrics in the music emotion recognition process. Our approach is based on several state of the art features complemented by novel stylistic, structural and semantic features. To evaluate our approach, we created a ground truth dataset containing 180 song lyrics, according to Russell's emotion model. We conduct four types of experiments: regression and classification by quadrant, arousal and valence categories. Comparing to the state of the art features (ngrams - baseline), adding other features, including novel features, improved the F-measure from 68.2%, 79.6% and 84.2% to 77.1%, 86.3% and 89.2%, respectively for the three classification experiments. To study the relation between features and emotions (quadrants) we performed experiments to identify the best features that allow to describe and discriminate between arousal hemispheres and valence meridians. To further validate these experiments, we built a validation set comprising 771 lyrics extracted from the AllMusic platform, having achieved 73.6% F-measure in the classification by quadrants. Regarding regression, results show that, comparing to similar studies for audio, we achieve a similar performance for arousal and a much better performance for valence.

Paper #68

Estimating Sentiment via Probability and Information Theory

Kevin Labille, Sultan Alfarhood and Susan Gauch

University of Arkansas, Fayetteville, U.S.A.

Keywords: Lexicons, Sentiment Analysis, Data Mining, Text Mining, Opinion Mining.

Abstract: Opinion detection and opinion analysis is a challenging but important task. Such sentiment analysis can be done using traditional supervised learning methods such as naive Bayes classification and support vector machines (SVM) or unsupervised approaches based on a lexicon may be employed. Because lexicon-based sentiment analysis methods make use of an opinion dictionary that is a list of opinion-bearing or sentiment words, sentiment lexicons play a key role. Our work focuses on the task of generating such a lexicon. We propose several novel methods to automatically generate a general-purpose sentiment lexicon using a corpus-based approach. While most existing methods generate a lexicon using a list of seed sentiment words and a domain corpus, our work differs from these by generating a lexicon from scratch using probabilistic techniques and information theoretical text mining techniques on a large diverse corpus. We conclude by presenting an ensemble method that combines the two approaches. We evaluate and demonstrate the effectiveness of our methods by utilizing the various automatically-generated lexicons during sentiment analysis. When used for sentiment analysis, our best single lexicon achieves an accuracy of 87.60% and the ensemble approach achieves 88.75% accuracy, both statistically significant improvements over 81.60% with a widely-used sentiment lexicon.

Parallel Session 2
14:30 - 16:30

KEOD
Room Argos

Paper #47

A Machine Learning Approach for Layout Inference in Spreadsheets

Elvis Koci¹, Maik Thiele¹, Oscar Romero² and Wolfgang Lehner¹

¹ Technische Universität Dresden, Dresden, Germany

² Universitat Politècnica de Catalunya (UPC-BarcelonaTech), Barcelona, Spain

Keywords: Spreadsheets, Tabular, Layout, Structure, Machine Learning, Knowledge Discovery.

Abstract: Spreadsheet applications are one of the most used tools for content generation and presentation in industry and the Web. In spite of this success, there does not exist a comprehensive approach to automatically extract and reuse the richness of data maintained in this format. The biggest obstacle is the lack of awareness about the structure of the data in spreadsheets, which otherwise could provide the means to automatically understand and extract knowledge from these files. In this paper, we propose a classification approach to discover the layout of tables in spreadsheets. Therefore, we focus on the cell level, considering a wide range of features not covered before by related work. We evaluated the performance of our classifiers on a large dataset covering three different corpora from various domains. Finally, our work includes a novel technique for detecting and repairing incorrectly classified cells in a post-processing step. The experimental results show that our approach delivers very high accuracy bringing us a crucial step closer towards automatic table extraction.

Paper #9

A Smart System for Haptic Quality Control

Introducing an Ontological Representation of Sensory Perception Knowledge

Bruno Albert^{1,2,3}, Cecilia Zanni-Merk¹, François de Bertrand de Beuvron¹, Jean-Luc Maire², Maurice Pillet², Julien Charrier³ and Christophe Knecht³

¹ Insa de Strasbourg, Strasbourg, France

² Université Savoie Mont-Blanc, Annecy-le-vieux, France

³ INEVA, Illkirch, France

Keywords: Sensory Perception, Haptics, Smart System, Semantic Analysis, Quality Control, Perceived Quality.

Abstract: Perceived quality has become an important factor in the choice of products by customers. The human perception process involves complex phenomena at a physical and psychological level that enable us to sense the world and extract information about it. Because of the qualitative way humans represent and communicate sensations, the field of sensory perceptions makes extensive use of semantics. The use of knowledge-based systems in the field of perceived quality is hence natural. This project focuses on haptics in quality control in industry. In particular, the aim is to develop a smart system which will enable to make decisions about the haptic quality of a product. This paper introduces the framework used for the development of this smart system, based on the KREM model. An ontological structure is

proposed in order to represent knowledge related to the measure of sensory perceptions in general, and of haptic ones in particular. The proposed domain ontologies about haptic control, that were elicited using semantic analysis, are aligned with the SSN ontology.

Paper #24

Proactive Prevention of False-Positive Conflicts in Distributed Ontology Development

Lavdim Halilaj, Irlán Grangel-González, Maria-Esther Vidal, Steffen Lohmann and Sören Auer

Fraunhofer IAIS and University of Bonn, Bonn, Germany

Keywords: Ontology Development, Unique Serialization, Version Control System, Editor Agnostic, RDF, OWL, Turtle.

Abstract: A Version Control System (VCS) is usually required for successful ontology development in distributed settings. VCSs enable the tracking and propagation of ontology changes, as well as collecting metadata to describe changes, e.g., who made a change at which point in time. Modern VCSs implement an optimistic approach that allows for simultaneous changes of the same artifact and provides mechanisms for automatic as well as manual conflict resolution. However, different ontology development tools serialize the ontology artifacts in different ways. As a consequence, existing VCSs may identify a huge number of false-positive conflicts during the merging process, i.e., conflicts that do not result from ontology changes but the fact that two ontology versions are differently serialized. Following the principle of prevention is better than cure, we designed SerVCS, an approach that enhances VCSs to cope with different serializations of the same ontology. SerVCS is based on a unique serialization of ontologies to reduce the number of false-positive conflicts produced whenever different serializations of the same ontology are compared. We implemented SerVCS on top of Git, utilizing tools such as Rapper and Rdf-toolkit for syntax validation and unique serialization, respectively. We have conducted an empirical evaluation to determine the conflict detection accuracy of SerVCS whenever simultaneous changes to an ontology are performed using different ontology editors. The evaluation results suggest that SerVCS empowers VCSs by preventing them from wrongly identifying serialization related conflicts.

Paper #26

Model-Intersection Problems with Existentially Quantified Function Variables: Formalization and a Solution Schema

Kiyoshi Akama¹ and Ekawit Nantajeewarawat²

¹ *Information Initiative Center, Hokkaido University, Sapporo, Japan*

² *Sirindhorn International Institute of Technology, Thammasat University, Pathumthani, Thailand*

Keywords: Model-Intersection Problem, Extended Clause, Function Variable, Equivalent Transformation.

Abstract: Built-in constraint atoms play a very important role in knowledge representation and are indispensable for practical applications. It is very natural to use built-in constraint atoms together with user-defined atoms when formalizing logical problems using first-order formulas. In the presence of built-in constraint atoms, however, the conventional Skolemization in general preserves neither the satisfiability nor the logical meaning of a given first-order formula, motivating us to step outside the conventional Skolemization and the usual space of first-order formulas. We propose general solutions for proof problems and query-answering (QA) problems on first-order formulas possibly

with built-in constraint atoms. We map, by using new meaning-preserving Skolemization, all proof problems and all QA problems, preserving their answers, into a new class of model-intersection (MI) problems on an extended clause space, where clauses are in a sense “higher-order” since they may contain not only built-in constraint atoms but also function variables. We propose a general schema for solving this class of MI problems by equivalent transformation (ET), where problems are solved by repeated simplification using ET rules. The correctness of this solution schema is shown. Since MI problems in this paper form a very large class of logical problems, this theory is also useful for inventing solutions for many classes of logical problems.

Paper #25

CDM-Core: A Manufacturing Domain Ontology in OWL2 for Production and Maintenance

Luca Mazzola, Patrick Kapahnke, Marko Vujic and Matthias Klusch

German Research Center for Artificial Intelligence (DFKI), Saarbruecken, Germany

Keywords: CDM-Core, Applied Ontology, Semantic Annotation, Knowledge Engineering, CREMA H2020 RIA Project, Ontology Quality Measurement.

Abstract: Ontology engineering is known to be a complex, time-consuming, and costly process, in particular, if an ontology has to be developed from scratch, and respective domain knowledge has to be formally encoded. This paper presents the largest publicly available manufacturing ontology CDM-Core in the standard formal ontology language OWL2¹. The CDM-Core ontology has been developed within the European research project CREMA in close collaboration with the user partners in order to sufficiently cover the CREMA use case domains of metal press maintenance and automotive exhaust production. CDM-Core makes use of many relevant standard vocabularies and ontologies, with only about one fifth of its size being CREMA use case specific. The practical applicability of CDM-Core for semantic annotation of domain-related process models, sensor data and services has been approved by the user partners, and its quality according to selected common criteria of verification and validation was successfully evaluated. From the public release of the CDM-Core, we expect to cover the lack of a base common ontology for the manufacturing domain, thanks to feedbacks from industrial reuse and improvements from the community.

Parallel Session 2
14:30 - 16:30

KMIS
Room Vargelas

Paper #9

Interaction Patterns in Web-based Knowledge Communities: Two-Mode Network Approach

Wouter Vollenbroek and Sjoerd de Vries

University of Twente, Enschede, The Netherlands

Keywords: Knowledge Management, Lifelong Learning, Social Network Analysis, Web-based Knowledge Communities.

Abstract: The importance of web-based knowledge communities (WKC) in the ‘network society’ is growing. This trend is seen in many disciplines, like education, government, finance and other profit- and non-profit organisations. There is a need for understanding the development of these online communities in order to steer it and to affect the impact it has. In this research,

we aimed to identify interaction patterns in these communities to visualize and understand community developments, and show the relevance of WKC for the development of learning education. We conducted a content analysis and a network analysis on big social data to identify the patterns in two Facebook-groups which were focused on educational development. Analysis of interaction patterns enabled us to identify three interaction stages within WKC in educational settings: introduction, evolution and maturity. In the first stage, participants mainly introduce themselves. In the second stage, one shares information and in the final stage, participants are more open to share their opinions. The study shows that our network analysis approach is appropriate to analyse and visualize the development of interaction patterns and the results could help us to steer communities effectively and efficiently.

Paper #13

Analyzing a Knowledge Country

How is Sweden Managing its Innovation Process?

Maria Obeso¹ and Maria Jesus Luengo-Valderrey²

¹ University of Cantabria, Santander, Spain

² University of the Basque Country, Bilbao, Spain

Keywords: Knowledge Country, Sweden, Innovation, Knowledge Management, Innovation Strategy.

Abstract: In a scenario where knowledge is considered an essential tool for firms and organizations, the paper presented an analysis about how Sweden, considered as a knowledge country, manages its innovation process. First, the paper presents the measures that Sweden government are doing, exposing the essential points in its innovation strategy. Then, the behaviour of Swedish enterprises related to innovation is analysed. Using an ANOVA technique, the paper shows that there are not differences between Swedish regions managing innovation, but there are differences between industries. After that, a comparison between industries in order to analyse which are the differences has been realized. Results show that machinery and retail industries are the most innovative, and fabricated metal industry is considered the least innovative. Although there are differences between industries, results show that in general Swedish enterprises are innovative because only a 9 per cent of enterprises have not realized something related to innovation process.

Paper #35

Differences between Knowledge and Information Management Practices: Empirical Investigation

Michal Krčál and Michal Kubiš

Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic

Keywords: Knowledge Management, Information Management, Comparison, Information Systems.

Abstract: In Knowledge Management (KM) discipline, the nature of KM itself has long been discussed and sometimes even its existence and meaning have been questioned. At the same time, research focusing on the difference between KM and Information Management (IM) was scarce. Therefore we tried through empirical investigation of differences between KM and IM to distinguish the KM from IM and to try to draw a distinct line between both approaches. To fulfil our goal we employed exploratory inductive qualitative research design as not many studies have tried to empirically distinguish KM from IM. For data gathering we used expert semi-structured interviews. The interviews and also

results were structured according to 8 perspectives: conceptual, process, technological, organisational, implementation, human resources, economical, and administration. For each perspective, we examined the context of IM and KM and analysed, described and interpreted the differences.

Paper #50

Model-based Strategic Knowledge Elicitation

J. Pedro Mendes

Instituto Superior Tecnico, Lisboa, Portugal

Keywords: Knowledge Elicitation, Mental Models, Strategic Problems, Problem Symptoms, Problem Dynamics, Strategy Implementation.

Abstract: Strategic problems are difficult. They often only exist in the mental models of some top managers. They are typically too vague to be given a precise meaning, yet at the same time concrete enough to cause discomfort. They cannot be discarded but they cannot be tackled either because the means for diagnostic and solution are lacking. The body of knowledge of strategy offers little help, in the sense that a set of tools for strategic problem solving does not exist, in practice. In science and engineering, problem solving is model-based. In the past, the social sciences and management have discarded model building due to its inherent difficulties. Today, the means are available to elicit knowledge about the symptoms of strategic problems, create a model to obtain a solution, and produce an implementation plan.

Paper #52

Big Data and Knowledge Management: How to Implement Conceptual Models in NoSQL Systems?

Fatma Abdelhedi¹, Amal Ait Brahim², Faten Atigui³ and Gilles Zurfluh²

¹ CBI2 – TRIMANE, Paris, France

² Toulouse Institute of Computer Science Research (IRIT), Toulouse Capitole University, Toulouse, France

³ CEDRIC-CNAM, Paris, France

Keywords: Big Data, NoSQL, Knowledge, MDA, QVT Transformation.

Abstract: In 2014, Big Data has passed the top of the Gartner Hype Cycle, proving that Big Data technologies and application start to be mature, becoming more realistic about how Big Data can be useful for organizations. NoSQL data stores are becoming widely used to handle Big Data; these databases operate on schema-less data model enabling users to incorporate new data into their applications without using a predefined schema. But, there is still a need for a conceptual model to define how data will be structured in the database. In this paper, we show how to store Big Data within NoSQL systems. For this, we use the Model Driven Architecture (MDA) that provides a framework for models automatic transformation. Starting from a conceptual model that describes a set of complex objects, we propose transformation rules formalized with QVT to generate a column-oriented NoSQL model. To ensure efficient automatic transformation, we use a logical model that limits the impacts related to technical aspects of column-oriented platforms. We provide experiments of our approach using a case study example taken from the health care domain. The results of our experiments show that the proposed logical model can be effectively implemented in different column-oriented systems independently of their specific technical details.

Workshop - Session 3
15:30 - 16:30

SKY
Room Bagaúste

Paper #3

Fast and Reliable Software Translation of Programming Languages to Natural Language

laakov Exman and Olesya Shapira

The Jerusalem College of Engineering - JCE - Azrieli, Jerusalem, Israel

Keywords: Software Translation, Source Code, Natural Language, Programming Languages, Software Tool, Fast Translation, Generality, Relevance and Reliability.

Abstract: An experienced software professional with several years of programming in some languages is usually expected to read or write with proficiency in a new programming language. However, if severe time constraints are involved, and given the current availability of internet sources, there is no reason to avoid shortcuts supporting fast translation of source code keywords into Natural Language. This work describes our tool coined PL-to-NL Translator, the main ideas behind it, and its extensions. One basic assumption that was clear from the beginning of this work is the need to keep as far as possible a clear separation between generic infra-structure and the specifics of particular programming languages. Moreover, the tool keeps its generality relative to programming languages, enabling through its contributor engine, addition of any desired current or future programming language. The ideas and the software tool characteristics are illustrated by some case studies involving a few sufficiently different programming languages.

Paper #2

Modulaser: A Tool for Conceptual Analysis of Software Systems

laakov Exman and Phillip Katz

The Jerusalem College of Engineering - JCE - Azrieli, Jerusalem, Israel

Keywords: Modulaser, Software Tools, Conceptual Analysis, Conceptual Integrity, Software System Design, Modularity Matrix, Standard.

Abstract: Modulaser is a software tool which produces a Modularity Matrix, to analyse the design of a software system given by its executable code. However, besides the concrete practical purposes of Modulaser, it is important to understand its techniques in a deeper sense. It is immediately clear that it describes the system in a higher abstraction level than the executable code, as the Modularity Matrix follows an implicit class diagram. But behind classes there are concepts. Thus, the ultimate purpose of Modulaser is conceptual analysis. This paper explains the ideas, describe Modulaser in these terms, and illustrate it by a series of case studies.

Poster Session 1
16:30 - 17:30

KDIR
Foyer

Paper #3

AToMRS: A Tool to Monitor Recommender Systems

André Costa, Tiago Cunha and Carlos Soares

Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

Keywords: Recommender Systems, Collaborative Filtering, Evaluation.

Abstract: Recommender systems arose in response to the excess of available online information. These systems assign, to a given individual, suggestions of items that may be relevant. These system's monitoring and evaluation are fundamental to the proper functioning of many business related services. It is the goal of this paper to create a tool capable of collecting, aggregating and supervising the results obtained from the recommendation systems' evaluation. To achieve this goal, a multi-granularity approach is developed and implemented in order to organize the different levels of the problem. This tool also aims to tackle the lack of mechanisms to enable visually assessment of the performance of a recommender systems' algorithm. A functional prototype of the application is presented, with the purpose of validating the solution's concept.

Paper #6

Grammar and Dictionary based Named-entity Linking for Knowledge Extraction of Evidence-based Dietary Recommendations

Tome Eftimov^{1,2}, Barbara Koroušič Seljak¹ and Peter Korošec^{1,3}

¹ *Jožef Stefan Institute, Ljubljana, Slovenia*

² *Jožef Stefan International Postgraduate School, Ljubljana, Slovenia*

³ *Faculty of Mathematics, Natural Sciences and Information Technologies, Koper, Slovenia*

Keywords: Named-entity Linking, Knowledge Extraction, Dietary Recommendations, Computational Linguistics, Public Health.

Abstract: In order to help people to follow the new knowledge about healthy diet that comes rapidly each day with the new published scientific reports, a grammar and dictionary based named-entity linking method is presented that can be used for knowledge extraction of evidence-based dietary recommendations. The method consists of two phases. The first one is a mix of entity detection and determination of a set of candidates for each entity, and the second one is a candidate selection. We evaluate our method using a corpus from dietary recommendations presented in one sentence provided by the World Health Organization and the U.S. National Library of Medicine. The corpus consists of 50 dietary recommendations and 10 sentences that are not related with dietary recommendations. For 47 out of 50 dietary recommendations the proposed method extract all the useful knowledge, and for remaining 3 only the information for one entity is missing. Due to the 10 sentences that are not dietary recommendation the method does not extract any entities, as expected.

Paper #12

Skip Search Approach for Mining Probabilistic Frequent Itemsets from Uncertain Data

Takahiko Shintani, Tadashi Ohmori and Hideyuki Fujita
The University of Electro-Communications, Tokyo, Japan

Keywords: Frequent Itemset, Probabilistic Data, Uncertain Data.

Abstract: Due to wider applications of data mining, data uncertainty came to be considered. In this paper, we study mining probabilistic frequent itemsets from uncertain data under the Possible World Semantics. For each tuple has existential probability in probabilistic data, the support of an itemset is a probability mass function (pmf). In this paper, we propose skip search approach to reduce evaluating support pmf for redundant itemsets. Our skip search approach starts evaluating support pmf from the average length of candidate itemsets. When an evaluated itemset is not probabilistic frequent, all its superset of itemsets are deleted from candidate itemsets and its subset of itemset is selected as a candidate itemset to evaluate next. When an evaluated itemset is probabilistic frequent, its superset of itemset is selected as a candidate itemset to evaluate next. Furthermore, our approach evaluates the support pmf by difference calculus using evaluated itemsets. Thus, our approach can reduce the number of candidate itemsets to evaluate their support pmf and the cost of evaluating support pmf. Finally, we show the effectiveness of our approach through experiments.

Paper #22

Detecting User Emotions in Twitter through Collective Classification

İbrahim İleri and Pinar Karagoz
Middle East Technical University, Ankara, Turkey

Keywords: Social Networks, Emotion Analysis, Sentiment Analysis, Collective Classification.

Abstract: The explosion in the use of social networks has generated a big amount of data including user opinions about varying subjects. For classifying the sentiment of user postings, many text-based techniques have been proposed in the literature. As a continuation of sentiment analysis, there are also studies on the emotion analysis. Due to the fact that many different emotions are needed to be dealt with at this point, the problem gets more complicated as the number of emotions to be detected increases. In this study, a different user-centric approach for emotion detection is considered such that connected users may be more likely to hold similar emotions; therefore, leveraging relationship information can complement emotion inference task in social networks. Employing Twitter as a source for experimental data and working with the proposed collective classification algorithm, emotions of the users are predicted in a collaborative setting.

Paper #45

An Analysis of Online Twitter Sentiment Surrounding the European Refugee Crisis

David Pope and Josephine Griffith
National University of Ireland Galway, Galway, Ireland

Keywords: Social Media Analysis, Sentiment Analysis, Refugee Crisis.

Abstract: Using existing natural language and sentiment analysis techniques, this study explores different dimensions of mood states of tweet content relating to the refugee crisis in Europe. The study has two main goals. The first goal is to compare the mood states of negative emotion, positive emotion, anger and anxiety across two populations (English and German speaking). The second goal is to discover if a link exists between significant real-world events relating to the refugee crisis and online sentiment on Twitter. Gaining an insight into this comparison and relationship can help us firstly, to better understand how these events shape public attitudes towards refugees and secondly, how online expressions of emotion are affected by significant events.

Paper #66

Using Deep Learning in Arabic-English Cross Language Information Retrieval

Omar Attia, Michael Azmy, Ahmed Abu Emeira, Karim El Azzouni, Omar Hussein, Nagwa M. El-Makky and Khaled Nagi

Faculty of Engineering, Alexandria University, Alexandria, Egypt

Keywords: Cross-language Information Retrieval, Deep Learning, Deep Belief Network, Canonical Correlation Analysis, Wikipedia, Arabic.

Abstract: In this paper, we apply a combination of keyword-based information retrieval with a latent semantic-based model in Arabic-English Cross-Language Information Retrieval (CLIR). We aim at enabling Arabic-speaking users to access English content using their native language. Due to the complex morphology of Arabic, we use deep learning for both Arabic and English text to extract the deep semantics in the given languages and then use canonical correlation analysis (CCA) to project the semantic representations into a shared space in which retrieval can be done easily. We evaluate our system on selected articles from Wikipedia and show that the proposed combination can improve the state-of-the-art keyword-based CLIR performance.

Paper #71

Development of Domains and Keyphrases along Years

Yaakov HaCohen-Kerner and Meir Hassan
Jerusalem College of Technology (Machon Lev), Jerusalem, Israel

Keywords: Development Concepts, Development Model, NLP Domains, Keyphrases, Trends, Word Bigrams and Trigrams.

Abstract: This paper presents a methodology (including a detailed algorithm, various development concepts and measures, and stopword lists) for measuring the development of domains and keyphrases along years. The examined corpus contains 1020 articles that were accepted for full presentation in PACLIC along the last 18 years. The experimental results for 5 chosen domains (digital humanities, language resources, machine translation, sentiment analysis and opinion mining, and social media) suggest that development trends of domains and keyphrases can be efficiently measured. Top bigrams and trigrams were found as efficient to identify general trends in NLP domains.

Paper #72

Gender Clustering of Blog Posts using Distinguishable Features

Yaakov HaCohen-Kerner, Yarden Tzach and Ori Asis
Jerusalem College of Technology (Machon Lev), Jerusalem, Israel

Keywords: Blog Posts, Distinguishable Features, Gender Clustering.

Abstract: The aim of this research is to find out how to perform effective clustering of unlabeled personal blog posts written in English by gender. Given a gender-labeled blog corpus and a blog corpus that is not gender-labeled, we extracted from the labeled corpus distinguishable unigrams for both males and females. Then, we defined two general features that represent the relative frequencies of the distinguishable males' unigrams and females' unigrams, (males' frequency and females' frequency). The best distinguishable feature was found to be the males' frequency feature with a ratio factor at least 1.4 times that of females. This feature leads to accuracy rate of 83.7% for gender clustering of the unlabeled blog corpus. To the best of our knowledge, this study presents two novelties: (1) this is the first study to cluster blog posts by gender, and (2) clustering of an unlabeled corpus using distinguishable features that were extracted from a labeled corpus.

Paper #77

The Flightschedule Profiler: An Attempt to Synthetise Visually an Airport's Flight Offer in Time and Space

Jean-Yves Blaise and Iwona Dudek
UMR CNRS/MCC 3495 MAP, Marseille, France

Keywords: Information Visualisation, Time-oriented Data, Transport.

Abstract: Online route planners and travel reservations systems have become in the past years part of our everyday lives. Such sites, originating from the airlines themselves or oriented on "search and compare" tasks, do provide valuable services. But the very nature of the queries users formulate (ultimate result: one flight) limits the type of information one can expect to retrieve, and in particular does not allow to get an overall view of an airport's flight offer over time and in space. In this contribution we introduce a proof-of-concept visualisation that sums up in a synthetic way the [where to, when to] profile of an airport, its realm of possibilities. The visualisation acts as an upstream service, independently of any actual reservation loop: its main role is to help unveiling significant spatio-temporal patterns (densities and continuity over time for instance). The prototype is implemented on a real life data set: the winter 2013/2014 schedule of the airport in Nice. Ultimately, beyond a discussion on the issue, on the pluses and minuses of the prototype, this position paper questions the way travel data is presented, and as such can promote debates over the potential impact of information visualisation solutions in that context.

Poster Session 1
16:30 - 17:30

KEOD
Foyer

Paper #7

Toward User Profile Representation in Adapted Mediation Systems

Sara Ouaftouh, Ahmed Zellou and Ali Idrri
Mohammed V University In Rabat, Rabat, Morocco

Keywords: User Modeling, Mediation Systems, User Model Representation.

Abstract: The amount of information offered by different software systems is growing exponentially and the need of personalized approaches for information access increases. This personalization aims to offer the user the pertinent information corresponding to his needs basing on his profile. For the same purpose, mediation systems have to identify user preferences in order to offer him the most relevant information. In this work we discuss different representations of user profile models designed for providing personalized information access in order to make a comparison and identify the most appropriate for our context in mediation systems.

Paper #37

Concept-based versus Realism-based Approach to Represent Neuroimaging Observations

Emna Amdouni¹ and Bernard Gibaud²

¹ *B-com Institute of Research and Technology, Rennes, France*

² *Université de Rennes 1, Rennes, France*

Keywords: Knowledge Representation, Domain Analysis and Modeling, Concept-based and Realism-based Ontologies.

Abstract: The aim of this paper is to argue why we should adopt a realism-based approach to describe neuroimaging features that are involved in clinical assessments rather than a concept-based approach. This work is a part of a proposal aiming at making explicit the meaning of neuroimaging observations via realism-based ontologies.

Poster Session 1
16:30 - 17:30

KMIS
Foyer

Paper #10

Development Results of the Intelligent Device for Storage of the Transfusion Environments Containing Platelets

A. G. Gudkov, V. Yu. Leushin, A. F. Bobrikhin, V. N. Lemondjava and E. N. Gorlacheva

Moscow State Technical University n.a. N.E. Bauman, Moscow, Russian Federation

Keywords: Blood, Transfusion, Platelet Concentrates, Storage.

Abstract: Research results directed on creation of the device for safe storage of transfusion environments containing platelets with the use of thermal stabilization system based on semiconductor

thermoelectric elements, information support of storage process monitoring and radio frequency identification (RFID) technologies are considered. Calculation results of the storage modes that differ the initial conditions of the process are given. Three-dimensional models of a device based on multi-layer walls are elaborated. Computational experiments with subsequent analysis of the temperature distribution on the computational domain are described.

Paper #55

A Study on the Public Construction Project Cost-Saving Practice Information Provision Service Method

Hyun Ok and Tae-Hak Kim

Korea Institute of Civil Engineering and Building Technology, Goyang-Si, Korea, Republic of

Keywords: CODIL(Construction technology Digital Library), CPIPS(Construction Project Information Portal System), Design Value Engineering, Cost-Saving Practices, Public Construction Projects.

Abstract: Cost saving in construction projects consists in building structures with the most economical method while satisfying the user requirements with regard to the features and quality. Cost saving through systematic cost management is required to improve the competitiveness in line with the changing construction environment. Organizations affiliated with the Ministry of Land, Infrastructure, and Transport are striving to save on construction costs by reviewing the economic efficiency of their designs, process management, and project management when implementing public construction projects. They are sharing the design value engineering (VE) results information among the public ordering agencies during the design stage, but the sharing of the cost-saving practices in the construction stage is still insufficient. This study presented step-by-step information systems development and service methods so that people can refer to and utilize cost-saving practices in similar construction projects while carrying out public construction projects. The results of this study are expected to contribute to more rational decision-making in and the higher investment efficiency of public construction projects through the analysis and provision of cost-saving cases information for each stage of the construction project life cycle, such as the planning, design, construction, and maintenance stages.

Keynote Lecture
17:30 - 18:30

IC3K
Room Pinhão

STEALTH: Modeling Coevolutionary Dynamics of Tax Evasion and Auditing

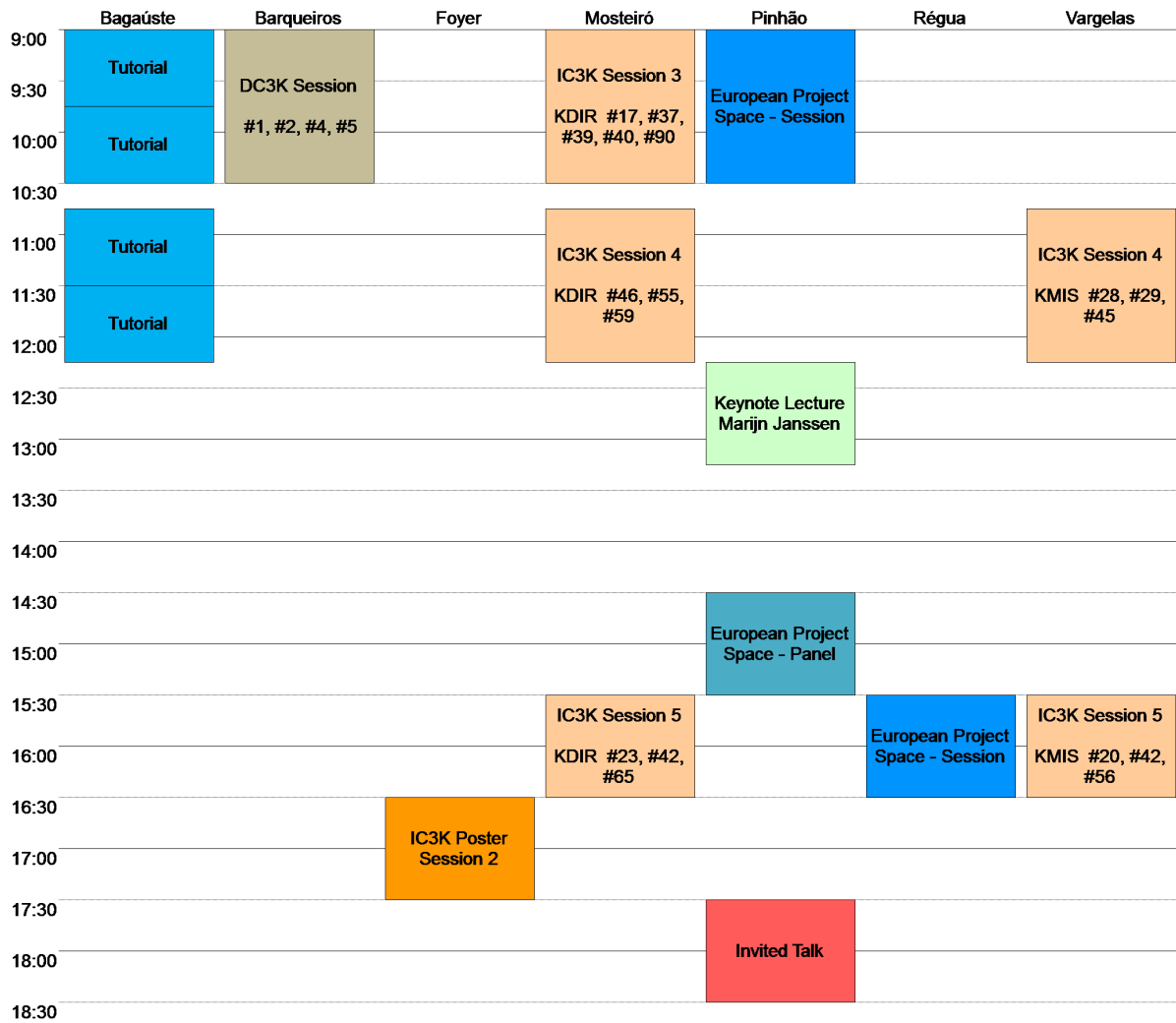
Una-May O'Reilly

MIT Computer Science and Artificial Intelligence Laboratory, MA, U.S.A.

Abstract: STEALTH is an AI system that detects tax law non-compliance by modeling the co-evolution of tax evasion schemes and their discovery through abstracted audits. Tax evasion accounts for billions of lost income each year. When the government pursues a tax evasion scheme and changes the tax law or audit procedures, the tax evasion schemes evolve and change into an undetectable form. The arms race between tax evasion schemes with tax authority actions presents a significant challenge to guide and plan enforcement efforts. Acknowledgement: Work done with Jacob Rosen, Erik Hemberg of ALFA (<http://groups.csail.mit.edu/ALFA>), plus Geoff Warner and Sanith Wijesinghe of MITRE Corporation (<http://csail.mit.edu/>).

Thursday Sessions: November 10

Thursday Sessions: November 10 Program Layout



Tutorial
09:00 - 09:45

SKY Intensive School
Room Bagaúste

Tutorial: AMG and Applications to Software Knowledge and Software Engineering

Radel Ben-Av

Jerusalem College of Engineering, Israel

Session 3
09:00 - 10:30

KDIR
Room Mosteiró

Paper #17

Bootstrapping a Semantic Lexicon on Verb Similarities

Shaheen Syed¹, Marco Spruit¹ and Melania Borit²

¹ *Utrecht University, Utrecht, The Netherlands*

² *University of Tromsø, Tromsø, Norway*

Keywords: Semantic Lexicon, Bootstrapping, Extraction Patterns, Web Mining.

Abstract: We present a bootstrapping algorithm to create a semantic lexicon from a list of seed words and a corpus that was mined from the web. We exploit extraction patterns to bootstrap the lexicon and use collocation statistics to dynamically score new lexicon entries. Extraction patterns are subsequently scored by calculating the conditional probability in relation to a non-related text corpus. We find that verbs that are highly domain related achieved the highest accuracy and collocation statistics affect the accuracy positively and negatively during the bootstrapping runs.

Paper #37

Sentiment Analysis of Breast Cancer Screening in the United States using Twitter

Kai O. Wong¹, Faith G. Davis², Osmar R. Zaiane² and Yutaka Yasui^{3,2}

¹ *University of Alberta, Alberta, Canada*

² *University of Alberta, Edmonton, Canada*

³ *St. Jude Children's Research Hospital, Memphis, U.S.A.*

Keywords: Cancer Screening, Social Media, Data Visualization, Sentiment Analysis, Spatial Analysis, Twitter.

Abstract: Whether or not U.S. women follow the recommended breast cancer screening guidelines is related to the perceived benefits and harms of the procedure. Twitter is a rich source of subjective information containing individuals' sentiment towards public health interventions/technologies. Using our modified version of Hutto and Gilbert (2014) sentiment classifier, we described the temporal, geospatial, and thematic patterns of public sentiment towards breast cancer screening with 8 months of tweets (n=64,524) in the U.S. To examine how sentiment was related to screening uptake behaviour, we investigated and identified significant associations between breast cancer screening sentiment (via Twitter) and breast cancer screening uptake (via BRFS) at the state level.

Paper #39

An Active Learning Approach for Ensemble-based Data Stream Mining

Rabaa Alabdulrahman¹, Herna Viktor¹ and Eric Paquet^{2,1}

¹ *University of Ottawa, Ottawa, Canada*

² *National Research Council of Canada, Ottawa, Canada*

Keywords: Online Learning, Data Streams, Active Ensemble Learning, Oracle.

Abstract: Data streams, where an instance is only seen once and where a limited amount of data can be buffered for processing at a later time, are omnipresent in today's real-world applications. In this context, adaptive online ensembles that are able to learn incrementally have been developed. However, the issue of handling data that arrives asynchronously has not received enough attention. Often, the true class label arrives after with a time-lag, which is problematic for existing adaptive learning techniques. It is not realistic to require that all class labels be made available at training time. This issue is further complicated by the presence of late-arriving, slowly changing dimensions (i.e., late-arriving descriptive attributes). The aim of active learning is to construct accurate models when few labels are available. Thus, active learning has been proposed as a way to obtain such missing labels in a data stream classification setting. To this end, this paper introduces an active online ensemble (AOE) algorithm that extends online ensembles with an active learning component. Our experimental results demonstrate that our AOE algorithm builds accurate models against much smaller ensemble sizes, when compared to traditional ensemble learning algorithms. Further, our models are constructed against small, incremental data sets, thus reducing the number of examples that are required to build accurate ensembles.

Paper #40

Prediction of Company's Trend based on Publication Statistics and Sentiment Analysis

Fumiyo Fukumoto¹, Yoshimi Suzuki¹, Akihiro Nonaka¹ and Karman Chan²

¹ *Univ. of Yamanashi, Kofu, Japan*

² *IJ Innovation Institute Inc., Tokyo, Japan*

Keywords: Publication Statistics, Sentiment Analysis, Prediction, Company's Trend.

Abstract: This paper presents a method for predicting company's trend on research and technological innovation/development(R&D) in business area. We used three types of data collections, i.e, scientific papers, open patents, and newspaper articles to estimate temporal changes of trends on company's business area. We used frequency counts on scientific papers and open patents to be published in time series. For news articles, we applied sentiment analysis to extract positive news reports related to the company's business areas, and count their frequencies. For each company, we then created temporal changes based on these frequency statistics. For each business area, we clustered these temporal changes. Finally, we estimated prediction models for each cluster. The results show that the the model obtained by combining three data is effective to predict company's future trends, especially the results show that SP clustering contributes overall performance

Paper #90

Lexicon Expansion System for Domain and Time Oriented Sentiment Analysis

Nuno Guimaraes¹, Luis Torgo² and Alvaro Figueira¹¹ CRACS - INESC TEC, Porto, Portugal² LIADD - INESC TEC, Porto, Portugal

Keywords: Sentiment Lexicon, Lexicon Expansion, Twitter Sentiment Analysis.

Abstract: In sentiment analysis the polarity of a text is often assessed recurring to sentiment lexicons, which usually consist of verbs and adjectives with an associated positive or negative value. However, in short informal texts like tweets or web comments, the absence of such words does not necessarily indicates that the text lacks opinion. Tweets like "First Paris, now Brussels... What can we do?" imply opinion in spite of not using words present in sentiment lexicons, but rather due to the general sentiment or public opinion associated with terms in a specific time and domain. In order to complement general sentiment dictionaries with those domain and time specific terms, we propose a novel system for lexicon expansion that automatically extracts the more relevant and up to date terms on several different domains and then assesses their sentiment through Twitter. Experimental results on our system show an 82% accuracy on extracting domain and time specific terms and 80% on correct polarity assessment. The achieved results provide evidence that our lexicon expansion system can extract and determined the sentiment of terms for domain and time specific corpora in a fully automatic form.

Doctoral Consortium - Session
09:00 - 10:30
Doctoral Consortium

DC3K
Room Barqueiros

Paper #1

Tacit Knowledge Management in Public Sector *A New Perspective for Organizational Knowledge Management*

Mauro Araújo Câmara

Federal University of Minas Gerais, Belo Horizonte, Brazil

Keywords: Tacit Knowledge, Tacit Knowledge Management, Public Sector, Expertise, Levels of Expertise.

Abstract: The Knowledge Management has been used as a tool to organize the knowledge produced by social actors in order to create the organizational knowledge itself. A wide variety of practices has helped to organize and formalize the knowledge that flows within organizations. Usually, it is made with the support of technologies through databases, repositories, intranet, SharePoint, and others. These tools have been implemented and described in different knowledge management models. These mechanisms have proven effective to deal with explicit knowledge allowing their setup and dissemination. However, the gap that is identified is precisely the superficiality when it comes to tacit knowledge. The main point of this research is to explore the importance of this kind of knowledge, i.e., the individual experience in the tasks that can contribute to positive outcomes, and needs to be preserved. The problem that concerns this research and which it intends to investigate is how to identify and transfer of tacit knowledge of experienced employees in order to protect the organizational knowledge. Organizing and disseminating information by simply using technology support does not guarantee that the knowledge developed over the years

will remain in the organization. It is necessary to do more than that.

Paper #2

Barriers and Perceived Benefits to IT Service Management Adoption in SMEs – A Literature Review

David Molamphy

Ulster University, Belfast, U.K.

Keywords: IT Service Management, ITSM, ITIL, SME

Abstract: The continuous development of technology and its increasing criticality in organisations across all sectors of the economy has led to the creation of models and frameworks for ensuring delivery of high quality IT services. Models such as the Information Technology Infrastructure Library (ITIL), Microsoft Operating Framework (MOF) and HP ITSM amongst others have been adopted by many firms to achieve this goal - ultimately to reduce costs and increase organisational productivity. Such models and frameworks however require substantial commitment, investment and human capital to both implement and operate. Whilst many large firms have the resources available to implement existing popular IT Service Management (ITSM) frameworks, SMEs most often do not, nor do they necessarily stand to benefit from these existing frameworks. As a result, there is a general consensus that all major ITSM frameworks currently published are designed to cater for the needs of the large firm. The aim of this doctoral consortium contribution is to present the current state of the art relating to the barriers and perceived benefits to ITSM adoption in smaller organisations. The paper follows a systematic review process, and concludes by presenting the expected outcome of the research project of the author.

Paper #4

Semantic Web Technologies to Enhance the Knowledge Discovery Process in Predictive Analytics

Iker Esnaola-Gonzalez

IK4-Tekniker, Eibar, Spain

Keywords: Semantic Web Technologies, Predictive Analytics, Energy Efficiency in Buildings.

Abstract: Knowledge Discovery in Databases (KDD) is the process which leads from raw data to actionable knowledge. The knowledge to discover can be contained in the primary data itself, from where it is discovered using appropriate algorithms and tools. However, this is not the only knowledge source: it can be contained in external data sources, or even only in the data analyst's mind. It has been pointed out that Semantic Web Technologies (SWT) can be used to improve each step of the KDD process. The approached research problem is how the SWT can be used to improve a Predictive Analytics process in the domain of the Energy Efficiency in Buildings, and our approach is expected accomplish a high level of abstraction of the problem to solve, which in turn should allow to replicate the process in similar use cases of the same domain with very little effort. Besides, the obtained predictions are also expected to improve.

Paper #5

Semantic Coverage Measures: Analytic Operators for Ontologists

Pallavi Karanth¹ and Kavi Mahesh²¹ Center for Knowledge Analytics and Ontological Engineering – KanOE, PES University, Bangalore, India² Center for Knowledge Analytics and Ontological Engineering – KanOE, PES University, Bangalore, India

Keywords: Semantic Web, Knowledge Analytics, Metrics, Semantic Coverage, Ontology.

Abstract: The field of Analytics has grown over the past decade with the promise of delivering insights from data in all its forms ranging from well structured to highly unstructured forms such as documents and web pages. With open data and open science initiatives and the development of Semantic Web, semantic data which includes ontological concepts and relations is becoming available in increasing amounts from diverse applications. Such semantic data is made up of knowledge structures typically in the form of graphs with well defined hierarchies of concepts and relations specified in an ontology of the domain. Analytics is still in its nascent stage in taking advantage of such semantically rich ontological data to deliver better insights. Ability to analyze semantically rich data enriched by various semantic constraints and knowledge structures can generate insights which can take the field of Analytics to a new level. Ontological instances can be analyzed for coverage which leads to new analytic metrics which we call Semantic Coverage measures. In this paper, we present different types of Semantic Coverage measures to analyze the A-Box against the T-Box. Such analytic metrics help ontologists fine tune ontologies and improve data collection and sampling techniques for better semantic coverage.

European Project Space - Session
09:00 - 10:30

IC3K
Room Pinhão

CLAFIS Project - Intelligent integrated solution for communication between automation systems and IT systems in farms process

Shalom Broyer

Technical University of Denmark (DTU), Denmark

CLAFIS Project - Knowledge Processing Framework and Security Implementations

Josef Küng

Technical University of Denmark (DTU), Denmark

CLAFIS Project - Data Fusion Framework for Agricultural Services

Bipjeet Kaur

Technical University of Denmark (DTU), Denmark

System Self-Awareness Towards Deep Learning and Discovering High-Value Information

Ying Zhao

Naval Postgraduate School, U.S.A.

Tutorial
09:45 - 10:30

SKY Intensive School
Room Bagaúste

Mining Source Code Regularities with FCA and Association Rules

Daniel Speicher

University of Bonn, Germany

Abstract: Software encodes knowledge about the application domain as well as about the solution domain, i.e. about the used technologies and how they are combined. The correct implementation at this level is very often characterized by regularities, which are seldom expressed explicitly. Therefore it is useful to identify these regularities automatically. To name two obvious examples: "95% of the classes that implement equals() implement as well hash Code()" or "80% of the methods that open a transaction close it as well". The deviations from these rules are at least worth a review if not implementation errors. Formal Concept Analysis offers a formal framework to mine Association Rules automatically and Concept Lattices are helpful in gaining an intuitive understanding.

Tutorial
10:45 - 11:30

SKY Intensive School
Room Bagaúste

Systems Engineering in the industrial environment: Requirements Engineering Processes

Anabel Fraga

Carlos III of Madrid University, Spain

Parallel Session 4
10:45 - 12:15

KDIR
Room Mosteiró

Paper #46

Unsupervised Irony Detection: A Probabilistic Model with Word Embeddings

Debora Nozza, Elisabetta Fersini and Enza Messina

University of Milano-Bicocca, Milan, Italy

Keywords: Irony Detection, Unsupervised Learning, Probabilistic Model, Word Embeddings.

Abstract: The automatic detection of figurative language, such as irony and sarcasm, is one of the most challenging tasks of Natural Language Processing (NLP). This is because machine learning methods can be easily misled by the presence of words that have a strong polarity but are used ironically, which means that the opposite polarity was intended. In this paper, we propose an unsupervised framework for domain-independent irony detection. In particular, to derive an unsupervised Topic-Irony Model (TIM), we built upon an existing probabilistic topic model initially introduced for sentiment analysis purposes. Moreover, in order to improve its generalization abilities, we took advantage of Word Embeddings to obtain domain-aware ironic orientation of words. This is the first work that addresses this task in unsupervised settings and the first study on the topic-irony distribution. Experimental results have shown that TIM is comparable, and sometimes even better with respect to supervised state of the art approaches for irony detection. Moreover, when integrating the probabilistic model

with word embeddings (TIM+WE), promising results have been obtained in a more complex and real world scenario.

Paper #55

Discovering Data Lineage from Data Warehouse Procedures

Kalle Tomingas, Priit Järv and Tanel Tammet

Tallinn University of Technology, Tallinn, Estonia

Keywords: Data Warehouse, Data Lineage, Dependency Analysis, Data Flow Visualization.

Abstract: We present a method to calculate component dependencies and data lineage from the database structure and a large set of associated procedures and queries, independently of actual data in the data warehouse. The method relies on the probabilistic estimation of the impact of data in queries. We present a rule system supporting the efficient calculation of the transitive closure. The dependencies are categorized, aggregated and visualized to address various planning and decision support problems. System performance is evaluated and analysed over several real-life datasets.

Paper #59

A Linear-dependence-based Approach to Design Proactive Credit Scoring Models

Roberto Saia and Salvatore Carta

Università di Cagliari, Cagliari, Italy

Keywords: Business Intelligence, Credit Scoring, Fraud Detection, Data Mining, Metrics.

Abstract: The main aim of a credit scoring model is the classification of the loan customers into two classes, reliable and unreliable customers, on the basis of their potential capability to keep up with their repayments. Nowadays, credit scoring models are increasingly in demand, due to the consumer credit growth. Such models are usually designed on the basis of the past loan applications and used to evaluate the new ones. Their definition represents a hard challenge for different reasons, the most important of which is the imbalanced class distribution of data (i.e., the number of default cases is much smaller than that of the non-default cases), and this reduces the effectiveness of the most widely used approaches (e.g., neural network, random forests, and so on). The Linear Dependence Based (LDB) approach proposed in this paper offers a twofold advantage: it evaluates a new loan application on the basis of the linear dependence of its vector representation in the context of a matrix composed by the vector representation of the non-default applications history, thus by using only a class of data, overcoming the imbalanced class distribution issue; furthermore, it does not exploit the defaulting loans, allowing us to operate in a proactive manner, by addressing also the cold-start problem. We validate our approach on two real-world datasets characterized by a strong unbalanced distribution of data, by comparing its performance with that of one of the best state-of-the-art approach: random forests.

Parallel Session 4
10:45 - 12:15

KMIS
Room Vargelas

Paper #28

Knowledge Management Framework for Early Phases in TOGAF-based Enterprise Architecture

Juan Pablo Meneses-Ortegón and Rafael A. Gonzalez

Pontificia Universidad Javeriana, Bogotá, Colombia

Keywords: Knowledge Management, Enterprise Architecture, TOGAF.

Abstract: Consulting firms in enterprise architecture that develop projects through the TOGAF framework may generate valuable knowledge from project to project. However, for this knowledge to create value, it must be supported by an effective ability to capture, store and reuse it. This paper proposes a knowledge management framework focused on TOGAF initial phases to enable reusing lessons from previous projects. Through a specific meta-model, it offers “ways of” thinking, working, supporting, controlling and modelling this process. As a result, we present some steps to develop knowledge management in TOGAF-based enterprise architecture projects through a case study in a consulting firm.

Paper #29

Business Intelligence Solution for an SME: A Case Study

Raghavendra Raj, Shun Ha Sylvia Wong and Anthony J. Beaumont

School of Engineering and Applied Science, Aston University, Birmingham, U.K.

Keywords: Business Intelligence, Data Warehouse, Microsoft BI, SME.

Abstract: Business Intelligence (BI) leverages the usefulness of existing information. It equips business users with relevant information to perform various analyses to make key business decisions. Over the last two decades, BI has become a core strategy for the growth of many companies, in particular large corporations. However, studies show that small and medium-sized enterprises (SMEs) lag behind in implementation and exploitation of BI solutions. To stay ahead of the competition, SMEs must be able to monitor and effectively use all of their resources, in particular information resources, to assist them in making important business decisions. In this paper, we examine the challenges such as lack of technical expertise and limited budget when implementing a BI solution within an SME in the UK. In light of our experiences in tackling these issues, we discuss how these challenges can be overcome through applying various tools and strategies and the potential benefits.

Paper #45

Organizational Records Systems

An Alternative View to (Enterprise) Information Systems

Sherry Li Xie^{1,2,3} and Guanyan (Amelie) Fan^{2,3}

¹ Key Laboratory of Data Engineering and Knowledge Engineering, Beijing, China

² Center for Electronic Records Management Research, Beijing, China

³ Renmin University of China, Beijing, China

Keywords: Information Systems, Enterprise Information System, Digital Records Management, InterPARES Project.

Abstract: The field of Information Systems (ISs) has been around since the 1960s (Hirschheim and Kleinand, 2011) and the notion of Enterprise Information Systems has been fully acknowledged for close to 30 years (Xu, 2007). Long existing in organizations or enterprises is also the field of records management, now predominantly, digital records management, which shares the many goals of the fields of ISs and EIS in terms of supporting enterprise operations and advancements. The DRM field recognized rather early in its battle to combat digital records challenges that the need to work closely with the ICT profession for devising information system functional requirements and for developing long term preservation strategies for valuable digital records. It is still rare, however, to spot discussions regarding the relationships among the fields in the ICT literature today. It is the intention of this communication piece to introduce one of the major developments of the international DRM field, i.e., the Chain of Preservation model, in particular the types of records systems that it encompasses, to the ISs and EIS professions, for the purpose of invoking further discussions and future collaborations.

Tutorial
11:30 - 12:15

SKY Intensive School
Room Bagaúste

Conceptual Integrity of Software Systems, An Overview

Iaakov Exman

The Jerusalem College of Engineering - JCE - Azrieli, Israel

Keynote Lecture
12:15 - 13:15

IC3K
Room Pinhão

From Open Data to Knowledge: Capitalizing Experiences

Marijn Janssen

Delft University of Technology, Delft, The Netherlands

Abstract: More and more data is opened and shared among public and private organizations. Despite the data deluge the creation of knowledge from open data proves to be more cumbersome. Indiscriminately opening and sharing of data has often limited value, as the gap between open data and the creation of knowledge is not bridged. New tools and instruments are necessary to share valuable insights created from the processing and use of the data. Various initiatives at the technical and organizational level are attempting to bridge this gap. In this keynote the open data vision is revisited, insight from experiences are capitalized and challenges and research directions are presented.

European Project Space - Panel

14:30 - 15:30

IC3K

Room Pinhão

H2020 Knowledge and Intelligent Systems: the Good, the Bad and the Disruptive

Parallel Session 5

15:30 - 16:30

KDIR

Room Mosteiró

Paper #23

Enriching What-If Scenarios with OLAP Usage Preferences

Mariana Carvalho and Orlando Belo

University of Minho, Braga, Portugal

Keywords: Business Intelligence, What-If Analysis, on-Line Analytical Processing, Usage Preferences, Multidimensional Databases.

Abstract: Nowadays, enterprise managers involved in decision-making processes struggle with numerous problems related to market position or business reputation of their companies. Owning the right and high quality set of information is a crucial factor for developing business activities and consequently gaining competitive advantages on business arenas. However, retrieving information is not enough. The possibility to simulate hypothetical scenarios without harming the business using What-If analysis tools and to retrieve highly refined information is an interesting way of achieving such advantages. In this paper, we propose an approach for helping to optimize enterprise decision processes using What-If analysis scenarios combined with OLAP usage preferences. We designed and developed a specific piece of software, which aims to discover the best recommendations for What-If analysis scenarios' parameters using OLAP usage preferences, which incorporates user experience in the definition and analysis of a target decision-making scenario.

Paper #42

Explanation Retrieval in Semantic Networks

Understanding Spreading Activation based Recommendations

Vanessa N. Michalke and Kerstin Hartig

TU Berlin, Berlin, Germany

Keywords: Explanation Retrieval, Spreading Activation, Pattern Recognition, Information Retrieval.

Abstract: Spreading Activation is a well-known semantic search technique to determine the relevance of nodes in a semantic network. When used for decision support, meaningful explanations of semantic search results are crucial for the user's acceptance and trust. Usually, explanations are generated based on the original network. Indeed, the data accumulated during the spreading activation process contains semantically extremely valuable information. Therefore, our approach exploits the so-called spread graph, a specific data structure that comprises the spreading progress data. In this paper, we present a three-step explanation retrieval method based on spread graphs. We show how to retrieve the most relevant parts of a network by minimization and extraction techniques and formulate meaningful explanations. The evaluation of the approach is then performed with a prototypical decision support system for automotive safety analyses.

Paper #65

Unsupervised Classification of Opinions

Itu Vlad Vasile, Rodica Potolea and Mihaela Dinsoreanu
Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Keywords: Unsupervised Learning, Opinion Mining, NLP, Domain Independent Learning, Implementation.

Abstract: Opinion mining is gaining more interest thanks to the ever growing data available on the internet. This work proposes an unsupervised approach that clusters opinions in fine grain ranges. The approach is able to generate its own seed words for better applicability to the context and eliminating user input. Furthermore, we devise a computation strategy for the influence of valence shifters and negations on opinion words. The method is general enough to perform well while reducing subjectivity to a minimum.

Parallel Session 5
 15:30 - 16:30

KMIS
 Room Vargelas

Paper #20

Challenges and Practices for Effective Knowledge Transfer in Globally Distributed Teams

A Systematic Literature Review

Josiane Kroll¹, Juho Mäkiö² and Manal Assaad²

¹ PUCRS, Porto Alegre, Brazil

² Hochschule Emden/Leer, Emden, Germany

Keywords: Global Software Development, Globally Distributed Team, Knowledge Management, Knowledge Transfer, Challenge, Practice.

Abstract: In the context of Global Software Development (GSD), team members face a number of challenges that needs to be solved. One of them relates to the transfer of knowledge needed to fulfil the required tasks. The knowledge transfer process may be organized and implemented in various ways, as companies use varying strategies to transfer knowledge from onsite to offshore sites. However, without effective knowledge management practices, success in GSD will be difficult. In this paper, we aims to identify the process activities of knowledge transfer in globally distributed teams, challenges and suitable solutions for effective knowledge transfer. In order to achieve this objective, a systematic literature review (SLR) of the existing knowledge transfer literature is conducted. Our findings describing the process of knowledge transfer in globally distributed teams, a set of challenges and recommended practices for effective knowledge transfer. Finally, we conclude this study with a discussion of the directions for further and future research.

Paper #42

How Do Young Researchers Take the Steps toward Startup Activities?

A Case Study of a One-day Workshop for Entrepreneur Education

Miki Saijo¹, Makiko Watanabe¹, Takumi Ohashi¹, Haruna Kusu², Hikaru Tsukagoshi² and Ryuta Takeda²

¹ Tokyo Institute of Technology, Tokyo, Japan

² Leave a Nest Co., Ltd., Tokyo, Japan

Keywords: Business Startup, Creativity, KEYS Scale, Knowledge Creation, Business Plan Refinement Workshop.

Abstract: This is an exploratory study of how young researchers with specific scientific knowledge, through deep conversations with mentors from industry in a Business Refinement Workshop (BRWS), are likely to change their original business plans, and what the factors are that will stimulate them to take action for business startup. It was found that the BRWS did lead to changes in business plan issues and solutions, but these changes did not necessarily lead to specification of the business plans. It was also found that a positive perception to the discrepancy of the mentors' comments was a factor that could stimulate startup activity after the workshop.

Paper #56

Towards Comprehensive Security Related Pedagogy

An Approach to Learning and Resilience

Rauno Pirinen¹, Juha Mäkinen² and Arto Salonen³

¹ Laurea University of Applied Sciences, Espoo, Finland

² The National Defense University, Helsinki, Finland

³ Metropolia University of Applied Sciences, Helsinki, Finland

Keywords: Action Competence, Adaptive Change, Comprehensive Security, Security Related Pedagogy, Resilience.

Abstract: The indent of this study is in progress of comprehensive security related pedagogy in the forms of national-international information sharing and knowledge management with the shared policy developing, collaboration in externally funded research consortiums, structures of security and safety organisations, and integration of strategic research and development (R&D) agenda with higher education functions. The study includes multiple case study analysis of integration of R&D projects and higher education functions, revised viewpoints to comprehensive security pedagogy and R&D related learning, and an approach to adaptive change process and resilience. The main contribution of study addresses to the progress of emergent educational aspects for the security related interactions, pedagogy, integration of higher education R&D, and collective research with national and European Commission research programmes.

European Project Space - Session
 15:30 - 16:30

IC3K
 Room Régua

Context-Dependent Associative Learning

Oussama H. Hamid

University of Kurdistan Hewler, Iraq

OpenSkiMr - Open Skill Match Maker

Andreas Kofler and Marianne Prast
Management Center Innsbruck, Austria

Robust Risk-Based Strategy Design

J.Pedro Mendes
CENTEC - Instituto Superior Tecnico, Lisboa, Portugal

Poster Session 2
16:30 - 17:30

KDIR
Foyer

Paper #5

Discovering the Deep Web through XML Schema Extraction

Yasser Saissi, Ahmed Zellou and Ali Idri
Mohammed V University in Rabat, Rabat, Morocco

Keywords: Deep Web, Schema Extraction, Web Integration.

Abstract: The web accessible by the search engines contains a vast amount of information. However, there is another part of the web called the deep web accessible only through its associated HTML forms, and containing much more information. The integration of the deep web content presents many challenges that are not fully addressed by the actual deep web access approaches. The integration of the deep web data requires knowing the schema describing each deep web source. This paper presents our approach to extract the XML schema describing a selected deep web source. The XML schema extracted will be used to integrate the associated deep web source into a mediation system. The principle of our approach is to apply a static and a dynamic analysis to the HTML forms giving access to the selected deep web source. We describe the algorithms of our approach and compare it to the other existing approaches.

Paper #29

Graph-based Rating Prediction using Eigenvector Centrality

Dmitry Dolgikh and Ivan Jelínek
Czech Technical University in Prague, Prague, Czech Republic

Keywords: Recommendation Systems, Graph-based Recommendations, User Preference, Social Network Analysis, Eigenvector Centrality.

Abstract: The most of recommendation systems rely on the statistical correlations of the past explicitly given user rating for items (e.g. collaborative filtering). However, in conditions of insufficient data of past rating activities, these systems are facing difficulties in rating prediction, this situation is commonly known as the cold-start problem. This paper describes how graph-based representation and Social Network Analysis can be used to help dealing with cold-start problem. We proposed a method to predict user rating based on the hypothesis that the rating of the node in the network corresponded to the rating of the most important nodes which are connected to it. The proposed method has been particularly applied to three MovieLens datasets to evaluate rating prediction performance. Obtained results showed competitiveness of our method.

Paper #50

Efficient Social Network Multilingual Classification using Character, POS n -grams and Dynamic Normalization

Carlos-Emiliano González-Gallardo¹, Juan-Manuel Torres-Moreno^{2,1}, Azucena Montes Rendón³ and Gerardo Sierra⁴

¹ *Université d'Avignon et des Pays de Vaucluse, Avignon, France*

² *École Polytechnique de Montréal, Montréal, Canada*

³ *Centro Nacional de Investigación y Desarrollo Tecnológico, Cuernavaca, Mexico*

⁴ *GIL-Instituto de Ingeniería, Universidad Nacional Autónoma de México, Ciudad de México, Mexico*

Keywords: Text Mining, Machine Learning, Classification, n -grams, POS, Blogs, Tweets, Social Network.

Abstract: In this paper we describe a dynamic normalization process applied to social network multilingual documents (Facebook and Twitter) to improve the performance of the Author profiling task for short texts. After the normalization process, n -grams of characters and n -grams of POS tags are obtained to extract all the possible stylistic information encoded in the documents (emoticons, character flooding, capital letters, references to other users, hyperlinks, hashtags, etc.). Experiments with SVM showed up to 90% of performance.

Paper #61

The Longest Common Subsequence Distance using a Complexity Factor

Octavian Lucian Hasna and Rodica Potolea
Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Keywords: Time Series, Classification, Longest Common Subsequence, Discretize, Complexity.

Abstract: In this paper we study the classic longest common subsequence problem and we use the length of the longest common subsequence as a similarity measure between two time series. We propose an original algorithm for computing the approximate length of the LCSS that uses a discretization step, a complexity invariant factor and a dynamic threshold used for skipping the computation.

Paper #64

Syllabification with Frequent Sequence Patterns *A Language Independent Approach*

Adrian Bona, Camelia Lemnaru and Rodica Potolea
Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Keywords: Syllabification, Frequent Pattern Sequence, Language Independence, Graphs, Data Mining.

Abstract: In this paper we show how words represented as sequences of syllables can provide valuable patterns for achieving language independent syllabification. We present a novel approach for word syllabification, based on frequent pattern mining, but also a more general framework for syllabification. Preliminary

evaluations on Romanian and English words indicated a word level accuracy around 77% for Romanian words and around 70% for English words. However, we believe the method can be refined in order to improve performance.

Paper #84

Exploring Urban Tree Site Planting Selection in Mexico City through Association Rules

Héctor Javier Vázquez¹ and Mihaela Juganaru-Mathieu²

¹ *Universidad Autonoma Metropolitana, Unidad Azcapotzalco, Mexico D.F, Mexico*

² *Ecole National Supérieure des Mines, Saint Etienne, France*

Keywords: Data Mining, Association Rules, Prediction, Rule Validation, Urban Trees, Planting Sites.

Abstract: In this paper we present an exploration of association rules determine planting sites considering urban tree's characteristics. In first step itemsets and rules are generated using the unsupervised algorithm Apriori. They are rapidly characterized in terms of tree planting sites. In a second step planting sites are fixed as target values to establish rules (a supervised version of the a priori algorithm). An original approach is also presented and validated for the prediction of the planting site of the species.

Paper #86

Computing Ideal Number of Test Subjects Sensorial Map Parametrization

Jessica Dacleu Ndengue, Mihaela Juganaru-Mathieu and Jenny Faucheu

Ecole Nationale Supérieure des Mines, Saint Etienne, France

Keywords: Napping®, Sensorial Map, RV-coefficient.

Abstract: A sensory analysis was carried using a special Napping® table on two different set of products in order to investigate on texture perception of material, the tests were done using a human panel. The data collected were analyzed through multiple factor analysis (MFA) which is a particular case of principal component analysis (PCA). The aim of this study is to know the minimum number of subjects in the human panel that can guarantee a meaningful statistical analysis of data, and so far allows a better understanding of the sensory results. We built a particular function that measures the similarity between two representations (two matrices) which are computed using the output of Napping® table. Based on this function and using the whole datasets an algorithm able to measure the robustness is implemented. We found on the two datasets that a minimum number of subjects between 10 and 12 seems to insure a stable and robust statistical analysis of the sensory results.

Poster Session 2
16:30 - 17:30

KEOD
Foyer

Paper #2

The Federated Ontology of the PAL Project Interfacing Ontologies and Integrating Time-dependent Data

Hans-Ulrich Krieger¹, Rifca Peters², Bernd Kiefer¹, Michael A. van Bekkum³, Frank Kaptein² and Mark A. Neerincx³

¹ *DFKI, The German Research Center for AI, Saarbrücken, Germany*

² *Interactive Intelligence Group, EEMCS, Delft University of Technology, Delft, The Netherlands*

³ *TNO, Soesterberg, The Netherlands*

Keywords: Knowledge Representation & Ontologies, Description Logics & OWL, Representation of & Reasoning with Time-dependent Data, Transaction Time, Semantic Interoperability, Integration of Upper & Domain Ontologies.

Abstract: This paper describes ongoing work carried out in the European project PAL which will support children in their diabetes self-management as well as assist health professionals and parents involved in the diabetes regimen of the child. Here, we will focus on the construction of the PAL ontology which has been assembled from several independently developed sub-ontologies and which are brought together by a set of hand-written interface axioms, expressed in OWL. We will describe in detail how the triple model of RDF has been extended towards transaction time in order to represent time-varying data. Examples of queries and rules involving temporal information will be presented as well. The approach is currently been in use in diabetes camps.

Paper #20

Towards a Semantic Approach for the Design of Social Network Users' Geographical Trajectories

Hadhami Ounissi¹, Marwa Mana¹ and Jalel Akaichi²

¹ *Université de Tunis, ISG, BESTMOD, Tunis, Tunisia*

² *College of Computer Science, King Khaled University, Abha, Saudi Arabia*

Keywords: Human Mobility, Ontology, Semantic Modeling, Social Networks, Spatial Data, Trajectory Data.

Abstract: The volume of data keeps growing rapidly, especially with the arrival and the frequent access to social networks. The spread of these networks provides users the opportunity to share their social, geographical and temporal information through geo-localized tweets and check-ins. The challenge is to exploit these data leads to a decision in favour of different situations encountered by these users. Thus, if we successfully analyze their trends according to the models of users' movements, we can then draw conclusions about the evolution of their instantaneous behavior and accomplished activities. But, the problem is that the use of such data decreases the provision of a representative formalism that combines spatial data and user information. In this paper, we propose an approach for a semantic modeling of social network users' trajectories. To do so, ontology seems to be a promising solution that allows us to annotate raw trajectories with semantic information to give birth to semantic trajectories. Such semantic trajectories are then analyzed in order to detect user behavior in a dynamic way.

Poster Session 2
16:30 - 17:30

KMIS
Foyer

Paper #14

Pro-Innovative Orientation of Polish Social Enterprises: The Empirical Perspective

Agata Sudolska and Monika Chodorek

Nicolaus Copernicus University, Torun, Poland

Keywords: Social Enterprise, Innovations, Pro-innovative Orientation.

Abstract: Contemporary times require an entrepreneurial approach to social problems. As the consequence social enterprises have emerged as a new type of organization which despite its non-profit character is treated as a fully entrepreneurial organization. As social enterprises adopt business principles while providing social goods or services, it has to develop an orientation that enables not only to survive, but also to compete in the contemporary environment. The paper is an attempt to contribute to the research in the field of social enterprises activities by examining their attitudes towards innovations. The main purpose of the paper is an attempt to assess pro-innovative orientation of Polish social enterprises. The paper addresses two tasks. The first part of the paper provides short theoretical overview of the idea and nature of social enterprise. As the theoretical analysis presented in the paper is exemplified with the empirical study, the second part of the paper presents the research findings which indicate the manifestations of social enterprises pro-innovative orientation. Based on the conducted research, we attempt to identify how social enterprises understand innovations, what kind of innovations they introduce. Finally, we attempt to connect the fact of creating innovations with pro-innovative behaviors of social enterprise employees.

Paper #27

A Survey of Open Government Data in Russian Federation

Dmitrij Koznov¹, Olga Andreeva¹, Uolevi Nikula², Andrey Maglyas², Dmitry Muromtsev³ and Irina Radchenko³

¹ *St. Petersburg State University, St. Petersburg, Russian Federation*

² *Lappeenranta University of Technology, Lappeenranta, Finland*

³ *ITMO University, St. Petersburg, Russian Federation*

Keywords: E-Government, Open Government Data, Public Sector, Open Government Data Ecosystem.

Abstract: Open data can increase transparency and accountability of a country government, leading to free information sharing and stimulation of new innovations. This paper analyses government open data policy as well as open data initiatives and trends in Russian Federation. The OECD analytical framework for national open government data portals and supporting initiatives is used as the bases for the study. The key issues of Russian open government data movement are summarized and aggregated. The paper argues the necessity of systematic development of the open data ecosystem, the leading role of the government in data release, a deeper business involvement, and a reduction of bureaucratic barriers.

Paper #36

A Tracing System for User Interactions towards Knowledge Extraction of Power Users in Business Intelligence Systems

Safwan Sulaiman¹, Tariq Mahmoud¹, Stephan Robbers¹, Jorge Marx Gómez¹ and Joachim Kurzhöfer²

¹ *Oldenburg University, Oldenburg, Germany*

² *Lufthansa Industry Solutions, Norderstedt, Germany*

Keywords: User Interactions, Knowledge Extraction, Power User, Self Service, Business Intelligence.

Abstract: Business intelligence has been widely integrated in enterprises to help their employees in their decision making process by delivering the needed information at the right time. Statistics from Gartner Group showed that the investment in the business intelligence domain has recently been very high. However, different studies and market researches showed that the pervasiveness and the usage percentage rate of business intelligence are still very low. The reason behind that is the complexity of the usage of business intelligence systems. Moreover, enterprise users lack analytical skills. To mitigate this problem, a new concept of self-service business intelligence has been developed. Within this system, the knowhow of power user is extracted and delivered to business users in form of recommendations. In this paper, we present the conception and development of the tracing module of this new system. This module has the goal of tracing the interactions of power users as the first step to extract their procedural knowledge in form of analysis paths. This is done by creating a user interaction catalogue in which the interactions are defined based on their relevance to the knowledge extraction process. Finally, this paper presents the internal architecture of this tracing module and its components.

Paper #54

A Plan for Constructing an Information Service System for Construction Project Management using GIS Information

SeongJin Kim and Hyun Ok

Korea Institute of Civil Engineering and Building Technology, Goyang-si, Korea, Republic of

Keywords: CALS, GIS, PMIS, Construction Project Information System

Abstract: Construction Management System provides assistance in the online exchange and sharing of various documents related to projects, contracts, and processes between government agencies and construction companies that perform national road and river construction projects. The system is being used for the road and river construction projects at approximately 500 locations. The recent development of various information technologies, including big data, GIS(Geographic Information System), and cloud, has enabled construction workers to utilize information devices for their work. A plan for improving Construction Management System was suggested in this study by applying spatial information so that various map-based information services could be provided to construction workers. The suggested plan is expected to provide workers with the visual construction status, including the cost, by integrating construction information and spatial information, and to reduce the onsite work hours and efforts by providing customized functions for each user.

Paper #58

Invited Talk
17:30 - 18:30IC3K
Room Pinhão**KPIs 4 Workplace Learning**

Sandro Emmenegger, Knut Hinkelmann, Barbara Thönssen and Frieder Witschel

*University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland***Keywords:** Model Driven Engineering, Workplace Learning, Learning Scorecard, Ontology, KPI.

Abstract: Enterprises and Public Administrations alike need to ensure that newly hired employees are able to learn the ropes fast. Employers also need to support continuous workplace learning. Workplace learning should be strongly related to business goals and thus, learning goals should directly add to business goals. To measure achievement of both learning and business goals we propose augmented Key Performance Indicators (KPI). In our research we applied model driven engineering. Hence we developed a model for a Learning Scorecard comprising of business and learning goals and their KPIs represented in an ontology. KPI performance values and scores are calculated with formal rules based on the SPARQL Inferencing Notation. Results are presented in a dashboard on an individual level as well as on a team/group level. Requirements, goals and KPIs as well as performance measurement were defined in close cooperation with Marche Region, business partner in Learn PAD.

Horizon 2020's SME Instrument

Daniel Chirtes

EASME - Executive Agency for SMEs - European Commission, Romania

Abstract: As part of the Horizon 2020 program, the European Commission is selecting potentially disruptive businesses to invest in and support as part of the SME Instrument. SMEs with a strong growth potential and the ambition to become world-market leaders could receive up to €2.5 million in funding, and world-class business support and mentoring.

Thursday, 10

Poster Session 2
16:30 - 17:30EPS
Foyer**CLAFIS Project - Implementing Agricultural Machine Remote Diagnostics over CLAFIS platform**

Ari Ronkainen

*Technical University of Denmark (DTU), Denmark***CLAFIS Project - Integrated System for Farm Operations - EU FP7 Funded Project**

Tomasz Blaszczyk

*Technical University of Denmark (DTU), Denmark***OpenSkiMr-Open Skill Match Maker**

Marianne Prast

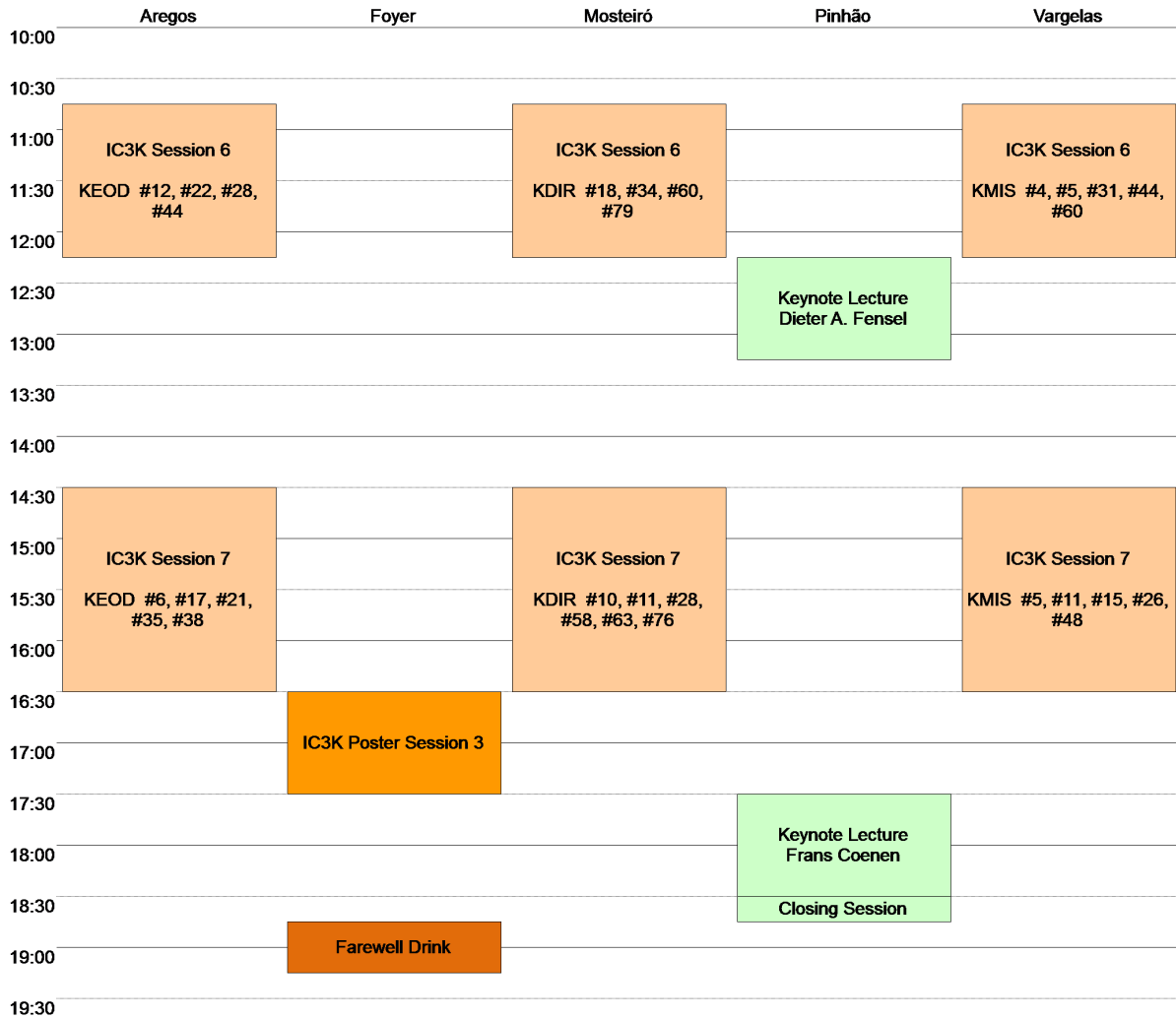
*Management Center Innsbruck, Austria***OSTAG - Ontology-based Software Test Case Generation**

Anders Adlemo

Jönköping University - School of Engineering, Sweden

Friday Sessions: November 11

Friday Sessions: November 11 Program Layout



Parallel Session 6
10:45 - 12:15

KDIR
Room Mosteiro

Paper #18

Heterogeneous Ensemble for Imaginary Scene Classification

Saleh Alyahyan, Majed Farrash and Wenjia Wang
University of East Anglia, Norwich, U.K.

Keywords: Heterogeneous Ensemble, Diversity, Big Data, Scene Classification.

Abstract: In data mining, identifying the best individual technique to achieve very reliable and accurate classification has always been considered as an important but non-trivial task. This paper presents a novel approach - heterogeneous ensemble technique, to avoid the task and also to increase the accuracy of classification. It combines the models that are generated by using methodologically different learning algorithms and selected with different rules of utilizing both accuracy of individual modules and also diversity among the models. The key strategy is to select the most accurate model among all the generated models as the core model, and then select a number of models that are more diverse from the most accurate model to build the heterogeneous ensemble. The framework of the proposed approach has been implemented and tested on a real-world data to classify imaginary scenes. The results show our approach outperforms other the state of the art methods, including Bayesian network, SVM and AdaBoost.

Paper #34

Result Diversity for RDF Search

Hiba Arnaout and Shady Elbassuoni
American University of Beirut, Beirut, Lebanon

Keywords: Diversity, Novelty, RDF-graphs, Evaluation.

Abstract: RDF repositories are typically searched using triple-pattern queries. Often, triple-pattern queries will return too many results, making it difficult for users to find the most relevant ones. To remedy this, some recent works have proposed relevance-based ranking-models for triple-pattern queries. However it is often the case that the top-ranked results are homogeneous. In this paper, we propose a framework to diversify the results of triple-pattern queries over RDF datasets. We first define different notions for result diversity in the setting of RDF. We then develop an approach for result diversity based on the Maximal Marginal Relevance. Finally, we develop a diversity-aware evaluation metric based on the Discounted Cumulative Gain and use it on a benchmark of 100 queries over DBpedia.

Paper #60

Efficient Distraction Detection in Surveillance Video using Approximate Counting over Decomposed Micro-streams

Avi Bleiweiss
BShalem Research, Sunnyvale, U.S.A.

Keywords: Stream Data, Approximate Counting, Sliding Window, Cosine Distance, Surveillance Video, Classification.

Abstract: Mining techniques of infinite data streams often store synoptic information about the most recently observed data elements. Motivated by space efficient solutions, our work exploits approximate counting over a fixed-size sliding window to detect distraction events in video. We propose a model that transforms inline the incoming video sequence to an orthogonal set of thousands of binary micro-streams, and for each of the bit streams we estimate at every timestamp the count of number-of-ones in a preceding sub-window interval. On window bound frames, we further extract a compact feature representation of a bag of count-of-1's occurrences to facilitate effective query of transitive similarity samples. Despite its simplicity, our prototype demonstrates robust knowledge discovery to support the intuition of a context-neutral window summary. To evaluate our system, we use real-world scenarios from a video surveillance online-repository.

Paper #79

Evaluation of Statistical Text Normalisation Techniques for Twitter

Phavanh Sosamphan¹, Veronica Liesaputra¹, Sira Yongchareon² and Mahsa Mohaghegh¹

¹ *Unitec Institute of Technology, Auckland, New Zealand*

² *AUT, Auckland, New Zealand*

Keywords: Text Mining, Social Media, Text Normalisation, Twitter, Statistical Language Models, Lexical Normalisation.

Abstract: One of the major challenges in the era of big data use is how to 'clean' the vast amount of data, particularly from micro-blog websites like Twitter. Twitter messages, called tweets, are commonly written in ill-forms, including abbreviations, repeated characters, and misspelled words. These 'noisy tweets' require text normalisation techniques to detect and convert them into more accurate English sentences. There are several existing techniques proposed to solve these issues, however each technique possess some limitations and therefore cannot achieve good overall results. This paper aims to evaluate individual existing statistical normalisation methods and their possible combinations in order to find the best combination that can efficiently clean noisy tweets at the character-level, which contains abbreviations, repeated letters and misspelled words. Tested on our Twitter sample dataset, the best combination can achieve 88% accuracy in the Bilingual Evaluation Understudy (BLEU) score and 7% Word Error Rate (WER) score, both of which are considered better than the baseline model.

Parallel Session 6
10:45 - 12:15

KEOD
Room Aregos

Paper #12

An Ontology-enabled Context-aware Learning Record Store Compatible with the Experience API

Jonas Anseeuw¹, Stijn Verstichel¹, Femke Ongenaes¹, Ruben Lagatie², Sylvie Venant² and Filip De Turck¹

¹ *Ghent University – iMinds, Ghent, Belgium*

² *Televic NV, Izegem, Belgium*

Keywords: Learning Analytics, Ontologies, Experience API, Linked Data, Learning Management System, Learning Record Store, Context-aware.

Abstract: In education, learners no longer perform learning activities in a well-defined and static environment like a physical

classroom. Digital learning environments promote learners anytime, anywhere and anyhow learning. As such, the context in which learners undertake these learning activities can be very diverse. To optimize learning and the environment in which it occurs, learning analytics measure data about learners and their context. Unfortunately, current state of the art standards and systems are limited in capturing the context of the learner. In this paper we present a Learning Record Store (LRS), compatible with the Experience API, that is able to capture the learners' context, more concretely his location and used device. We use ontologies to model the xAPI and context information. The data is stored in a RDF triple store to give access to different services. The services will show the advantages of capturing context information. We tested our system by sending statements from 100 learners completing 20 questions to the LRS.

Paper #22

Information Search in Ontology Visualization – An Eyetracking User Study of Indented List on Desktop and Tablet Computers

Anh Huynh and Bo Fu

California State University, Long Beach, U.S.A.

Keywords: Eye-tracking, Indented List Visualization, Usability, Biomedical Ontologies.

Abstract: A large amount of research efforts have focused on designing and developing ontology visualization methods over the years, but less effort in comparison has been put on evaluating usability support of these existing ontology visualization techniques particularly in rising interaction mediums such as touchscreen devices. This paper investigates the visual support of indented list visualization - traditionally designed for desktop computers - in the context of class search activities using traditional desktop computers as well as tablet computers. Using task-based user studies conducted on desktop and tablet computers, we analysed the difference between task success, task speed, eye gaze, as well as qualitative data collected from usability questionnaires, we found that the indented list visualization is not as effective on tablet computers with increased gaze activities, where many users preferred using it on desktop computers.

Paper #28

Qualitative Reasoning for Understanding the Behaviour of Complex Biomolecular Networks

Ali Ayadi^{1,2}, Cecilia Zanni-Merk¹ and François de Beuvron de Bertrand¹¹ ICUBE/SDC Team (UMR CNRS 7357), Illkirch, France² Institut Supérieur de Gestion de Tunis, University of Tunis, Bardo, Tunisia

Keywords: Biomolecular Networks, Dynamical Modelling, Qualitative Reasoning, Qualitative Simulation.

Abstract: Understanding the dynamical behaviour of cellular systems requires the development of effective modelling techniques. The modeling aims to facilitate the study and understanding of the dynamic behaviour of these systems, by the simulation of their designed models. Complex biomolecular networks are the basis of these models. In this paper, we propose a method of qualitative reasoning, based on a formal logical modeling, to qualitatively simulate the biomolecular network and interpret its behaviour over time. The power of our approach is illustrated by applying it to the case study of the autoregulation of the bacteriophage T4 gene 32.

Paper #44

Avionics Maintenance Ontology Building for Failure Diagnosis Support

Luis Palacios^{1,2}, Gaëlle Lortal¹, Claire Laudy¹, Christian Sannino³, Ludovic Simon³, Giuseppe Fusco², Yue Ma² and Chantal Reynaud²¹ Thales Research & Technologies, Palaiseau, France² Laboratoire de Recherche en Informatique (LRI), Orsay Cedex, France³ Thales Avionics, Toulouse, France

Keywords: Ontology Building, Ontology Alignment, Evaluation Prototye, Aeronautics, Predictive Maintenance.

Abstract: In the aviation industry, the delay in maintaining or recovering aircrafts heavily impacts the profit of an airline company. Consequently the maintenance actions identification and planning of aircrafts is crucial. However, due to the complexity of the domain in terms of data sources, distributed systems and information availability, it is hard to provide automatic maintenance support. We propose to use semantic technologies to model the domain at a conceptual level through ontology, thus abstracting from the data sources and the maintainers' uses and jobs. In this manner the information relevant for characterizing failures and maintenance events is encapsulated and provided to end users via an easier access, which otherwise would be inaccessible or would require expert analysis to obtain. Such a formal model of the domain can furthermore enable automated reasoning for maintenance discovery and failure causes detection by integrating a large amount of background contextual information scattering in different resources. In this paper we provide the rationale of the Avionics Maintenance ontology i.e. how we built it through expert knowledge and alignment of different sources and an ontology alignment evaluation tool.

Parallel Session 6
10:45 - 12:15

KMIS
Room Vargelas

Paper #44

How Can We Facilitate the Diffusion of Electric Vehicles in Japan?

Consumer Characteristics and Perceived Innovation Attributes

Taiyu Sato and Miki Saijo

Tokyo Institute of Technology, Tokyo, Japan

Keywords: Innovation Facilitation, Electric Vehicles, Consumer Characteristics, Perceived Innovation Attributes.

Abstract: Electric vehicles (EVs) have significant potential to substantially reduce CO2 emissions from transportation. Researchers have been working around the world to find ways to diffuse the use of this innovation in markets, but only a few such studies have been made in Japan. The aim of this study is to pinpoint potential adopters of EVs and the factors driving or hindering the diffusion of EV use in Japan. An online questionnaire was used to identify consumer characteristics and perceived innovation attributes of EVs among 208 car owners. Two groups of car owners divided by intentions to buy EVs were compared. We found that potential adopters perceive EVs positively and have positive environmental attitudes and are scientifically literate. Our findings also indicate social environment norms can drive the diffusion of EVs while consumers' lack of compatibility with EVs hinders the widespread diffusion of this innovation. Finally we discuss the limitations and

implications of this study.

Paper #31

The STAPS Method

Process-taylored Introduction of Knowledge Management Solutions

Christoph Sigmanek and Birger Lantow
University of Rostock, Rostock, Germany

Keywords: Knowledge Management, Business Process Oriented Knowledge Management, Knowledge Management System, Best Practices, Method.

Abstract: Nowadays, knowledge is recognized as an important enterprise resource. Thus, knowledge management is perceived as a necessary management task. Process oriented knowledge management is an approach that aligns knowledge management with the requirements of knowledge intensive processes. However, existing approaches to the implementation of process oriented knowledge management either operate on a very high abstraction level, incorporating much effort for operationalization in practice or on a very detailed level concentrating on process modelling. This paper introduces the STAPS method for the process oriented analysis and implementation of knowledge management solutions. It allows the assessment of already existing knowledge management solutions, the adoption of new solutions based on best practises, and a tailoring to organizational needs. In a case study, the applicability of STAPS is proved.

Paper #60

A State Transition Model for Business Processes

Towards Object-oriented Business Process Automation

Anantharamaiah Prabhakar
Enterprise Flow Technosoft Services Pvt. Ltd., Bangalore, India

Keywords: Business Process Modelling, Business Process Design, Human Centric BPM.

Abstract: This paper develops a new model for business processes, called State Transition Model. According to this model, a business process instance starts in some state, then flows, not necessarily sequentially, through zero or more intermediate states, until it is terminated in one of the states. The State Transition Model isolates the workflow part of a process from the non-workflow parts. The State Transition Model facilitates object-oriented design of business processes. It results in a top-down approach to human-centric business process automation. This is in contrast to the popular BPMN model, which results in a bottom-up approach.

Parallel Session 6
10:45 - 12:15

Open Communications
Room Vargelas

Paper #4

Is Knowledge a strategic and unavoidable resource? : from the VRIN framework to organizational performance

Abdeslam Jajaf
Université Paris 1, Panthéon Sorbonne, Paris, France

Keywords: Knowledge, VRIN, RBV, KBV, Organizational performance.

Abstract: Organizations of all sizes and sectors realized that an important change happened: the passage to a new "knowledge society" (Drucker, 1993; p.6). Indeed, this new society (Drucker, 1993) had several important effects on organizations and how they act and perform (Davenport and Prusak, 1998) and notably their environment that became overly moving and instable, with several and various changes coming to affect it (Davenport and Prusak, 1998). So to say, they had and have to adapt to it, by becoming more flexible and responsive (Davenport and Prusak, 1998). In order to do so, they had to focalize on their intangible or nonphysical resources (Halawi et al., 2005), and efficiently manage them. From these resources, knowledge is widely and broadly considered as one of, or even the most important one (Grant, 1996 ; Booto Ekionea et al., 2010), even seen as an unavoidable way of getting competitive advantages (Nonaka, 1991; Booto-Ekionea et al., 2010) in this new society (Drucker, 1993). More precisely and according to the Resource based-view (Barney, 1991), resources with certain characteristics might be able to yield a competitive advantage (Barney, 1991). Barney (1991) developed a kind of model or framework, which allows the assessment and analysis of the different resources, in order to find out if they can or not grant a competitive advantage to their owning organizations, and if the latter is sustainable or not. Sub approaches have also been developed, with the knowledge based-view underlining the important role/part played by knowledge (Grant, 1996). So, the objective of our conceptual research is first to analyze knowledge using the VRIN framework, to see if this unavoidable resource can indeed fulfill the four VRIN characteristics developed by Barney (1991). This analyze and discussion will help us see if the theoretical views are justified or not. Then, we will see if these four characteristics might be related to organizational performance, by trying to shed light on the theoretical complexities of their relations, and interactions. Finally, our research will discuss theoretical implications, in order to propose several future research tracks.

Paper #5

Using tablet devices to foster KM processes within working contexts in France: towards a new approach

Mehdi EL ABED¹, Abdeslam Jajaf² and Jean-François LEMOINE³

¹ Sorbonne University – Paris 1, Paris, France

² Université Paris 1, Panthéon Sorbonne, Paris, France

³ Sorbonne University – Paris 1 / ESSCA School of management, France

Keywords: Tablet, Knowledge management, infrastructures, adoption.

Abstract: Technologies are used to perform and fulfill a large va-

riety of tasks and became unavoidable. This focalization on using tablets is more or less true since the apparition of a new age called "digital age" by Berman (2012; p.21). More precisely, concerning tablets or "tablettes" in French, those tools or devices are used by customers and workers to realize a large variety and scope of tasks.

In the more specific scope of tablet use within the particular context of work and by the workers/employees, it is important to analyze their role played in the crucial execution of KM processes, notably as an important and new KM infrastructure (Gold et al., 2001). Besides and firstly, the factors affecting or not the adoption and use of those "tablettes", still by the workers, are evenly more important and need to be studied.

So to say, the objective of this open communication is then double:
- What are the key and important factors in the adoption of tablets or "tablettes" by French workers? (Notably by using the great model developed by Venkatesh et al., 2003).
- Can we consider tablets as key technological infrastructure supporting KM processes?

This open communication will then allow us to discuss these two key axes of analysis with researchers and practitioners, in order to enhance our draft model and spawn some research's trends, by crossing different disciplinary views.

References

Berman, S.J. "Digital transformation: opportunities to create new business models", *Strategy & Leadership*, (40:2), 2012, 16-24.

Gold, A.H., Malhotra A., and Segars A.H. "Knowledge Management: An Organizational Capabilities Perspective," *Journal of Management Information Systems* (18:1), 2001, 185-214.

Venkatesh, V., Morris, M.G., Davis, F.D., and Davis, G.B. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* (27:3), 2003, 425-478

Keynote Lecture
12:15 - 13:15

IC3K
Room Pinhão

No Smartness Without Data

Dieter A. Fensel

University Innsbruck, Innsbruck, Austria

Abstract: AI started with strong expectations 60 years ago. However, the knowledge acquisition bottleneck started an ice time called the AI winter. Meanwhile the situation drastically changed. Large volumes of data and their smart integration provide completely new possibilities. In the talk, we show how assistant systems of Amazon, Apple, Bing, Google and others make usage of semantic technologies to implement a new service layer on top of the current Web. We take the touristic industry as a vertical for illustrating that e-marketing and e-commerce can no longer be done successfully if ignoring these recent trends. Integration semantic technologies in their service provisioning turns from a "nice-to-have" into a "must-have" for the future of such business fields.

Parallel Session 7
14:30 - 16:30

KDIR
Room Mosteiró

Paper #10

Putting Web Tables into Context

Katrin Braunschweig, Maik Thiele, Elvis Koci and
Wolfgang Lehner

Technische Universität Dresden, Dresden, Germany

Keywords: Information Extraction, Web Tables, Text Tiling, Similarity Measures.

Abstract: Web tables are a valuable source of information used in many application areas. However, to exploit Web tables it is necessary to understand their content and intention which is impeded by their ambiguous semantics and inconsistencies. Therefore, additional context information, e.g. text in which the tables are embedded, is needed to support the table understanding process. In this paper, we propose a novel contextualization approach that 1) splits the table context in topically coherent paragraphs, 2) provides a similarity measure that is able to match each paragraph to the table in question and 3) ranks these paragraphs according to their relevance. Each step is accompanied by an experimental evaluation on real-world data showing that our approach is feasible and effectively identifies the most relevant context for a given Web table.

Paper #11

DTATG: An Automatic Title Generator based on Dependency Trees

Liqun Shao and Jie Wang

University of Massachusetts, Lowell, U.S.A.

Keywords: Title Generator, Central Sentence, Sentence Compression, Dependency Tree Pruning, TF-IDF.

Abstract: We study automatic title generation for a given block of text and present a method called DTATG to generate titles. DTATG first extracts a small number of central sentences that convey the main meanings of the text and are in a suitable structure for conversion into a title. DTATG then constructs a dependency tree for each of these sentences and removes certain branches using a Dependency Tree Compression Model we devise. We also devise a title test to determine if a sentence can be used as a title. If a trimmed sentence passes the title test, then it becomes a title candidate. DTATG selects the title candidate with the highest ranking score as the final title. Our experiments showed that DTATG can generate adequate titles. We also showed that DTATG-generated titles have higher F1 scores than those generated by the previous methods.

Paper #28

Summarization of Spontaneous Speech using Automatic Speech Recognition and a Speech Prosody based Tokenizer

György Szaszák¹, Máté Ákos Tündik¹ and András Beke²

¹ *Budapest University of Technology and Economics, Budapest, Hungary*

² *Research Institute for Linguistics of the Hungarian Academy of Sciences, Budapest, Hungary*

Keywords: Audio, Speech, Summarization, Tokenization, Speech Recognition, Latent Semantic Indexing.

Abstract: This paper addresses speech summarization of highly spontaneous speech. The audio signal is transcribed using an Automatic Speech Recognizer, which operates at relatively high word error rates due to the complexity of the recognition task and high spontaneity of speech. An analysis is carried out to assess the propagation of speech recognition errors into syntactic parsing. We also propose an automatic, speech prosody based audio tokenization approach and compare it to human performance. The so obtained sentence-like tokens are analysed by the syntactic parser to help ranking based on thematic terms and sentence position. The thematic term is expressed in two ways: TF-IDF and Latent Semantic Indexing. The sentence scores are calculated as a linear combination of the thematic term score

and a positional score. The summary is generated from the top 10 candidates. Results show that prosody based tokenization reaches human average performance and that speech recognition errors propagate moderately into syntactic parsing (POS tagging and dependency parsing). Nouns prove to be quite error resistant. Audio summarization shows 0.62 recall and 0.79 precision by an F-measure of 0.68, compared to human reference. A subjective test is also carried out on a Likert-scale. All results apply to spontaneous Hungarian.

Paper #58

Introducing the Key Stages for Addressing Multi-perspective Summarization

Elena Lloret

University of Alicante, Alicante, Spain

Keywords: Natural Language Processing, Text Summarization, Multi-perspective Summarization, Opinion Mining, Topic Identification, Knowledge Discovery.

Abstract: Generating summaries from evaluative text (e.g., reviews) is a challenging task, in which available metadata is hardly exploited, thus leading to the creation of very generic and biased summaries. In this paper, the novel task of multi-perspective summarization is introduced. The key stages for generating this type of summaries are defined, and a preliminary analysis of their feasibility is conducted. The main novelties of this study include: i) the linguistic treatment of the text at the level of basic information units, instead of sentences; and ii) the analysis carried out over the distribution of opinions and topics.

Paper #63

Learning Diagnosis from Electronic Health Records

Ioana Barbantan and Rodica Potolea

Technical University of Cluj-Napoca, Cluj-Napoca, Romania

Keywords: Data Mining, Classification, Value Mapping, Concept Extraction, Semantic Medical Data Alignment.

Abstract: In the attempt to build a complete solution for a medical assistive decision support system we proposed a complex flow that integrates a sequence of modules which target the different data engineering tasks. This solution can analyse any type of unstructured medical documents which are processed by applying specific NLP steps followed by semantic analysis which leads to the medical concepts identification, thus imposing a structure on the input documents. The data collection, document pre-processing, concept extraction, and correlation are modules that have been researched by us in our previous works and for which we proposed original solutions. Using the collected and structured representation of the medical records, informed decisions regarding the health status of the patients can be made. The current paper focuses on the prediction module that joins all the components in a logical flow and is completed with the suggested diagnosis classification for the patient. The accuracy rate of 81.25%, obtained on the medical documents supports the strength of our proposed strategy.

Paper #76

Power to the People!

Meta-Algorithmic Modelling in Applied Data Science

Marco Spruit and Raj Jagesar

Utrecht University, Utrecht, The Netherlands

Keywords: Applied Data Science, Meta-algorithmic Modelling, Machine Learning, Big Data.

Abstract: This position paper first defines the research field of applied data science at the intersection of domain expertise, data mining, and engineering capabilities, with particular attention to analytical applications. We then propose a meta-algorithmic approach for applied data science with societal impact based on activity recipes. Our people-centred motto from an applied data science perspective translates to design science research which focuses on empowering domain experts to sensibly apply data mining techniques through prototypical software implementations supported by meta-algorithmic recipes.

Parallel Session 7
14:30 - 16:30

KEOD
Room Aregos

Paper #6

Reconstruction of Implied Semantic Relations in Russian Wiktionary

Serge Klimenkov, Evgenij Tsopa, Alexey Pismak and Alexander Yarkeev

ITMO University, St. Petersburg, Russian Federation

Keywords: Semantic Analysis, Semantic Network, Semantic Web, Natural Language Processing, Wiktionary, Russian Language.

Abstract: There were several attempts to retrieve semantic relations from free, online Wiktionary for Russian language. Previous works combine automatic parsing of wiki snapshot with experts' assistance. Our main goal is to create machine readable lexical ontology from Russian Wiktionary, maximally close to its online state. This article provides approach to automatic creation of explicit and implicit semantic relations between words (lexemes) and meanings (senses) to provide exact relations from sense to sense. Explicit semantic relations are constructed comparatively easy. For example, if the lexeme contains single sense, then all relations that point to the lexeme will point to this single sense. Reconstruction of implicit relations relies on logical conclusions from already created explicit ones. Several algorithms for implicit semantic links were developed and tested on Russian Wiktionary. There were parsed more than 550000 online pages, containing about 250000 Russian lexemes with about 500000 senses in them, but only about 20% of these senses were linked with at least one external lexeme. About 47% of explicitly existing links were resolved as "sense-to-sense" relations and about 28% of new implicit "sense-to-sense" links were reconstructed. 53% of lexemes' references could not be resolved to exact sense.

Paper #17

Unfolding Existentially Quantified Sets of Extended Clauses

Kiyoshi Akama¹ and Ekawit Nantajeewarawat²¹ Information Initiative Center, Hokkaido University, Hokkaido, Japan² Sirinrhorn International Institute of Technology, Thammasat University, Pathumthani, Thailand

Keywords: Unfolding, Extended Clause, Function Variable, Model-intersection Problem, Equivalent Transformation.

Abstract: Conventional theories cannot solve many logical problems due to the limitations of the underlying clause space. In conventional clauses, all variables are universally quantified and no existential quantification is allowed. Conventional clauses are therefore not sufficiently expressive for representing first-order formulas. To extend clauses with the expressive power of existential quantification, variables of a new type, called function variables, have been introduced, resulting in a new space of extended clauses, called ECLS_F. This new space is necessary to overcome the limitations of the conventional clause space. To solve problems on ECLS_F, many equivalent transformation rules are used. We formally defined unfolding transformation on ECLS_F, which is applicable not only to definite clauses but also to multi-head clauses. The proposed unfolding transformation preserves the answers to model-intersection problems and is useful for solving many logical problems such as proof problems and query-answering problems on first-order logic with built-in constraint atoms.

Paper #21

SOCUDO-SCACLO: Ontologies for Socio-cultural Aware Collaborative Learning

Fadoua Ouamani, Narjès Bellamine Ben Saoud and Henda Hajjami Ben Ghézala

Ecole Nationale des Sciences de l'Informatique, Université de Manouba, Manouba, Tunisia

Keywords: Ontology, Culture, User Profile, Collaborative Learning, Adaptation, Personalization.

Abstract: At the time of a computer supported collaborative learning (CSCL) activity that brings together participants from different cultures, the need and the expectations of the latter from the collaborative system are not the same at many levels (Graphic user interface, activity content, pedagogical resources, functionalities). From this context, was born the need to adapt these environments to the socio-cultural specificities of learners. To do that, we need first to model these specificities, their impacts on learners and their learning. The proposed adaptation approach is then based on an ontology framework SOCUDO-SCACLO composed of two ontologies. SOCUDO is a generic core ontology that models the socio-cultural characteristics of any user of any application while SCACLO is a domain ontology that models their impacts on Collaborative Learning domain variables. The instantiation of SOCUDO triggers the instantiation of SCACLO based on association rules that infer SCACLO concepts values from SOCUDO concept values. The merging of the instances allows building a socio-cultural profile for each user which is used to personalize the layout, the content and the functionalities of such environments. In this paper, we report the results of our ontology engineering effort to build this framework, to operationalize it and evaluate its benefits to socio-cultural adaptation of CSCL environments.

Paper #35

A Bayesian Approach for Weighted Ontologies and Semantic Search

Anna Formica¹, Michele Missikoff², Elaheh Pourabbas¹ and Francesco Taglino¹¹ National Research Council, Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", Rome, Italy² National Research Council, Istituto di Scienze e Tecnologie della Cognizione, Rome, Italy

Keywords: Semantic Search, Similarity Reasoning, Weighted Reference Ontology, Bayesian Network.

Abstract: Semantic similarity search is one of the most promising methods for improving the performance of retrieval systems. This paper presents a new probabilistic method for ontology weighting based on a Bayesian approach. In particular, this work addresses the semantic search method *SemSim* for evaluating the similarity among a user request and semantically annotated resources. Each resource is annotated with a vector of features (annotation vector), i.e., a set of concepts defined in a reference ontology. Analogously, a user request is represented by a collection of desired features. The paper shows, on the bases of a comparative study, that the adoption of the Bayesian weighting method improves the performance of the *SemSim* method.

Paper #38

OntoMetrics: Putting Metrics into Use for Ontology Evaluation

Birger Lantow

Rostock University, Rostock, Germany

Keywords: Ontology, Ontology Evaluation, Ontology Metrics, Ontology Quality.

Abstract: Automatically calculated metrics are needed in order to evaluate ontology quality. Otherwise, major resources are required in order to manually assess certain Quality Criteria of an ontology. While there is rule based support for the detection modelling errors and the violation of ontology modelling guidelines, there is a lack of support for calculating ontology metrics. However, metrics can serve as indicators for possible quality problems that are not covered rule based ontology evaluation. Many metrics have been proposed that correlate for example with ontology characteristics like Readability, Adaptability, and Reusability. However, there is a lack of tool support. OntoMetrics provides free access to metric definition and calculation. Furthermore it fosters the development of knowledge regarding the application of Ontology Metrics. This paper provides theoretical background and usage scenarios for the OntoMetrics on-line platform.

Parallel Session 7
14:30 - 16:30

KMIS
Room Vargelas

Paper #5

Mobile Data Store Platforms: Test Case based Performance Evaluation

Karim Kussainov and Bolatzhan Kumalakov
Nazarbayev University, Astana, Kazakhstan

Keywords: Mobile Database, SQLite, Realm, SnappyDB, Performance Evaluation.

Abstract: Mobile applications are an important tool in knowledge management, as they collect and process massive amount of user data. Day-to-day usage of mobile services has rocketed by factors over the last decade. Average mobile device user installs multiple social network, messaging, professional and leisure applications. Saving and retrieving associated data becomes a challenging task in the light of the growing number of applications on a single device. While industry offers several well established platforms, such as BerkeleyDB and UnQLite, we examine comparatively poorly examined Realm and SnappyDB against industry standard - SQLite. In particular we are interested in performance and code maintainability, and use a test case in order to asses them. Results revile that SQLite shows the poorest performance, while Realm provides the most intuitive way of matching data to the application logic due to its object-oriented nature.

Paper #11

Big Data Knowledge Service Framework based on Knowledge Fusion

Fei Wang¹, Hao Fan¹ and Gang Liu²

¹ *Wuhan University, Wuhan, China*

² *Central China Normal University, Wuhan, China*

Keywords: Knowledge Fusion, Knowledge Service, Process Model, Implementation Pattern, System Framework.

Abstract: In big data environments, knowledge fusion is the necessary prerequisite and effective approach to implement knowledge service. This paper firstly analyses the requirements of big data knowledge service and the contents of knowledge fusion, constructs a multi-level architecture of knowledge service based on knowledge fusion. Then, this paper presents a design of a knowledge fusion process model and analyses its implementation patterns. Finally, a system framework of big data knowledge service is proposed based on knowledge fusion processes, in which processes of both knowledge fusion and knowledge service are organically combined together to provide an effective solution to achieve personalized, multi-level and innovative knowledge service.

Paper #15

Ensuring Action: Identifying Unclear Actor Specifications in Textual Business Process Descriptions

Ulf Sanne¹, Hans Friedrich Witschel¹, Alessio Ferrari² and Stefania Gnesi²

¹ *Fachhochschule Nordwestschweiz, Olten, Switzerland*

² *CNR, Pisa, Italy*

Keywords: Business Process Management, Quality Assessment, Natural Language Processing.

Abstract: In many organisations, business process (BP) descriptions are available in the form of written procedures, or operational manuals. These documents are expressed in informal natural language, which is inherently open to different interpretations. Hence, the content of these documents might be incorrectly interpreted by those who have to put the process into practice. It is therefore important to identify language defects in written BP descriptions, to ensure that BPs are properly carried out. Among the potential defects, one of the most relevant for BPs is the absence of clear actors in action-related sentences. Indeed, an unclear actor might lead to a missing responsibility, and, in turn, to activities that are never performed. This paper aims at identifying unclear actors in BP descriptions expressed in natural language. To this end, we define an algorithm named ABIDE, which leverages rule-based natural language processing (NLP) techniques. We evaluate the algorithm on a manually annotated data-set of 20 real-world BP descriptions (1,029 sentences). ABIDE achieves a recall of 87%, and a precision of 56%. We consider these results promising. Improvements of the algorithm are also discussed in the paper.

Paper #26

Managing Heterogeneous Information in a System of Information Systems

Majd Saleh, Véronique Misséri and Marie-Hélène Abel

Sorbonne Universités, Université de Technologie de Compiègne, Compiègne, France

Keywords: Information Systems, Knowledge Management, System of Systems, System of Information Systems.

Abstract: Organizations look for information coming from many heterogeneous resources, which are in turn produced by several users working on different Information Systems. Those heterogeneous resources might complement each other, hence producing useful information for the organization. Another possibility is that they might overlap or contradict, therefore producing redundant contradictory information. A common problem that faces organizations here is in the overwhelming amount of the heterogeneous resources produced, and the time required to access and manages those resources under the goal of finding useful information in them. Therefore, one of the most important aspects concerning Information Systems in organizations is the way to manage the knowledge produced from those systems. This issue becomes more significant when dealing with complex Information Systems that work in the digital environment forming the Digital Ecosystem (DE) of modern organizations. As organizations attempts to move forward in this complex Digital Ecosystem, They need to apply out-of-the-box ideas to complex systems such as System of Information Systems (SoIS). Thus, we aim to develop a solution for managing heterogeneous information in a System of Information Systems, and move forward from Digital Ecosystems to Collaborative System of Information Systems. In this paper

we present our definition of a Digital Ecosystem. Also, we will detail the model of the SoIS. This model is used to construct the prototype for managing heterogeneous information in a System of Information Systems.

Paper #48

Categorization and Matching for Drone-based Services

Simona Ibba¹, Filippo Eros Pani¹ and Alberto buschetti²

¹ University of Cagliari, Cagliari, Italy

² SulcisDrone, Cagliari, Italy

Keywords: Knowledge Management, Taxonomy, Folksonomy, Categorization of Resources, Matching, Drones.

Abstract: The exchange of supply and demand in drone-based services would benefit from the shared use of an online platform. Such a platform would need to offer two important opportunities. One is to share the service with other users, and the other is to receive offers from all providers interacting on the platform, who optimize their investments by sharing their own resources and technologies. The purpose of this position paper is to propose a categorization and matching algorithm on which to base this platform. The platform will aim to facilitate the sharing of services provided through the use of drones. The algorithm will match demand and offer, and will evolve through the use and the application of all participants (operators, users, lenders of shared resources). The platform, currently in development, could be the first web-based system in Europe to offer this model.

Poster Session 3
16:30 - 17:30

KDIR
Foyer

Paper #16

Success Prediction System for Student Counseling using Data Mining

Jörg Frochte and Irina Bernst

Bochum University of Applied Sciences, Heiligenhaus, Germany

Keywords: Data Mining, Classification, Supervised Learning, Information Privacy, Tertiary Education.

Abstract: A framework how to use data mining of central exam data for the prediction of student success in bachelor degree courses is presented. For the prediction a supervised learning approach is used based on successful and unsuccessful student biographies. We develop a traffic light rating system and present results for two different kinds of bachelor degree courses; one in economics and one in engineering. We discuss applications for students and student counseling institutions as well as the limitations dealing with information privacy aspects, especially under the conditions regarding data mining in Germany.

Paper #30

Adopting Privacy Regulations in a Data Warehouse

A Case of the Anonymity versus Utility Dilemma

Chaïm van Toledo and Marco Spruit

Utrecht University, Utrecht, The Netherlands

Keywords: Privacy, k-anonymity, p-sensitivity, Data Warehouse, Privacy Enhancing Technologies, ETL, Having Clause.

Abstract: This paper investigates how privacy can be protected in a data warehouse while, at the same time, an organisation tries to be as open as possible. First, we perform a literature review on relevant techniques and methods to preserve privacy and show that k-anonymity can be applied to comply with an organisation's requirements. Then, we propose two design strategies to adopt privacy regulations within a data warehouse. The first proposal shows that during the ETL process a data transformation can be performed to effectively realise anonymised records in a business intelligence environment. The second proposal shows that with views and a having clause, anonymisation can be arranged as well.

Paper #33

A Conceptual Model of Actors and Interactions for the Knowledge Discovery Process

Lauri Tuovinen

University of Oulu, Oulu, Finland

Keywords: Knowledge Discovery in Data, Process Model, Computer-supported Collaboration, Intelligent Systems.

Abstract: The knowledge discovery process is traditionally viewed as a sequence of operations to be applied to data; the human aspect of the process is seldom taken into account, and when it is, it is mainly the roles and actions of domain and technology experts that are considered. However, non-experts can also play an important role in knowledge discovery, and furthermore, the role of technology in the process may also be substantially expanded from what it traditionally has been, with special software facilitating interactions among human actors and even operating as an actor in its own right. This diversification of the knowledge discovery process is helpful in finding tenable solutions to the new problems presented by the current deluge of digital data, but only if the process model used to manage the process adequately represents the variety of forms that the process can take. The paper addresses this requirement by presenting a conceptual model that can be used to describe different types of knowledge discovery processes in terms of the actors involved and the interactions they have with one another. Additionally, the paper discusses how the interactions can be facilitated to provide effective support for each different type of process. As a future perspective, the paper considers the implications of intelligent software taking on responsibilities traditionally reserved for human actors.

Paper #36

Exploration and Visualization of Big Graphs *The DBpedia Case Study*

Enrico G. Caldarola^{1,2}, Antonio Picariello³, Antonio M. Rinaldi¹ and Marco Sacco⁴

¹ University of Naples Federico II, Napoli, Italy

² National Research Council, Bari, Italy

³ University of Napoli Federico II, Napoli, Italy

⁴ National Research Council, Milano, Italy

Keywords: Graph Database, Big Data, NoSQL, Data Visualization, DBpedia, Neo4J.

Abstract: Increasingly, the data and information visualization is becoming strategic for the exploration and explanation of large data sets. The Big Data paradigm pushes for new ways, new technological solutions to deal with the big volume and the big variety of data today. Not surprisingly, a plethora of new tools have emerged, each of them with pros and cons, but all espousing the cause of "Bigness of Data". In this paper, we take one of this emerging tools, namely Neo4J, and stress its capabilities in order to import, query and visualize data coming from a *big* case study: DBpedia. We will describe each step in this study focusing on the used strategies for overcoming the different problems mainly due to the intricate nature of the case study and its volume. We confront with both the intensional schema of DBpedia and its extensional part in order to obtain the best result in its visualization. Finally, an attempt to define some criteria to simplify the large-scale visualization of DBpedia will be made, providing some examples and considerations which have arisen. The ultimate goal of this work is to investigate techniques and approaches to get more insights from the visual representation and analytics of large graph databases.

Paper #53

Recommending Groups to Users based Both on Their Textual and Image Posts

Elias Oliveira, Howard Roatti, Gustavo Ramos Lima and Patrick Marques Ciarelli

Universidade Federal do Espírito Santo, Vitória, Brazil

Keywords: Text Classification, Image Classification, Social Network, Textmining.

Abstract: This article focuses on the recommendation of Facebook groups to users, based on the post profile of each user on Facebook. In order to accomplish this task, texts and images provided by users are used as source of information, and the experiments showed that the combination of these two types of information gives results better than or equal to the results obtained when using separately these data. The proposed approach in this paper is simple and promising to recommend Facebook groups.

Paper #81

Rating Prediction with Contextual Conditional Preferences

Aleksandra Karpus¹, Tommaso Di Noia², Paolo Tomeo² and Krzysztof Goczyla¹

¹ Gdańsk University of Technology, Gdańsk, Poland

² Polytechnic University of Bari, Bari, Italy

Keywords: Recommender Systems, Context Awareness, Conditional Preferences, Rating Prediction, Cold-start Problem.

Abstract: Exploiting contextual information is considered a good solution to improve the quality of recommendations, aiming at suggesting more relevant items for a specific context. On the other hand, recommender systems research still strive for solving the cold-start problem, namely where not enough information about users and their ratings is available. In this paper we propose a new rating prediction algorithm to face the cold-start system scenario, based on user interests model called contextual conditional preferences. We present results obtained with three publicly available data sets in comparison with several state-of-the-art baselines. We show that usage of contextual conditional preferences improves the prediction accuracy, even when all users have provided a few feedbacks, and hence small amount of data is available.

Paper #85

Anti-cancer Drug Activity Prediction by Ensemble Learning

Ertan Tolan and Mehmet Tan

TOBB University of Economics and Technology, Ankara, Turkey

Keywords: Cancer, Drug Activity, Ensemble Learning.

Abstract: Personalized cancer treatment is an ever-evolving approach due to complexity of cancer. As a part of personalized therapy, effectiveness of a drug on a cell line is measured. However, these experiments are backbreaking and money consuming. To surmount these difficulties, computational methods are used with the provided data sets. In the present study, we considered this as a regression problem and designed an ensemble model by combining three different regression models to reduce prediction error for each drug-cell line pair. Two major data sets were used to evaluate our method. Results of this evaluation show that predictions of ensemble method are significantly better than models *per se*. Furthermore, we report the cytotoxicity predictions of our model for the drug-cell line pairs that do not appear in the original data sets.

Paper #87

Big Data and Deep Analytics Applied to the Common Tactical Air Picture (CTAP) and Combat Identification (CID)

Ying Zhao, Tony Kendall and Bonnie Johnson

Naval Postgraduate School, Monterey, U.S.A.

Keywords: Big Data, Deep Analytics, Common Tactical Air Picture, Combat Identification, Machine Vision, Object Recognition, Pattern Recognition, Anomaly Detection, Lexical Link Analysis, Heterogeneous Data Sources, Unsupervised Learning.

Abstract: Accurate combat identification (CID) enables warfighters to locate and identify critical airborne objects as friendly, hostile or neutral with high precision. The current CID processes include processing and analysing data from a vast network of sensors, platforms, and decision makers. CID plays an important role in generating the Common Tactical Air Picture (CTAP) which provides situational awareness to air warfare decision-makers. The Big "CID" Data and complexity of the problem pose challenges as well as opportunities. In this paper, we discuss CTAP and CID challenges and some Big Data and Deep Analytics solutions to address these challenges. We present a use case using a unique deep learning method, Lexical Link Analysis (LLA), which is able to associate heterogeneous data sources for object recognition and anomaly detection, both of which are critical for CTAP and CID applications.

Paper #88

SPARQL Query Generation based on RDF Graph

Mohamed Kharrat, Anis Jedidi and Faiez Gargouri
University of Sfax, Sfax, Tunisia

Keywords: SPARQL, Queries, RDF.

Abstract: Data retrieval is becoming more difficult due to the heterogeneity and the huge amount of Data flowing in the Web. On the other hand, novice users could not handle querying languages (e.g., SPARQL) or knowledge based techniques. To simplify querying process, we introduce in this paper, a proposal of automatic SPARQL query generation based on user-supplied keywords. The construction of a SPARQL query is based on the top relevant RDF sub-graph, selected from our RDF Triplestore. This latter rely on our defined semantic network and on our Contextual Schema both published in two different papers of our previous studies. We evaluate 50 queries by using three measures. Results show an F-Score of about 50%. This proposal is already implemented as a web interface and the whole queries interpretation and processing is performed over this interface.

Paper #89

A Novel Clustering Algorithm to Capture Utility Information in Transactional Data

Piyush Lakhawat, Mayank Mishra and Arun Somani
Iowa State University, Ames, U.S.A.

Keywords: Clustering Algorithm, Transactional Data, High Utility Patterns.

Abstract: We develop and design a novel clustering algorithm to capture utility information in transactional data. Transactional data is a special type of categorical data where transactions can be of varying length. A key objective for all categorical data analysis is pattern recognition. Therefore, transactional clustering algorithms focus on capturing the information on high frequency patterns from the data in the clusters. In recent times, utility information for category types in the data has been added to the transactional data model for a more realistic representation of data. As a result, the key information of interest has become high utility patterns instead of high frequency patterns. To the best of our knowledge, no existing clustering algorithm for transactional data captures the utility information in the clusters found. Along with our new clustering rationale we also develop corresponding metrics for evaluating quality of clusters found. Experiments on real datasets show that the clusters found by our algorithm successfully capture the high utility patterns in the data. Comparative experiments with other clustering algorithms further illustrate the effectiveness of

our algorithm.

Poster Session 3
16:30 - 17:30

KEOD
Foyer

Paper #18

Ontological Approach to Share Product Design Semantics for an Assembly

Baha Hasan, Jan Wikander and Mauro Onori
KTH, Stockholm, Sweden

Keywords: Assembly Semantic, Assembly Feature, Feature-based Modelling, Ontology, Computer Aided Design (CAD).

Abstract: The aim of this paper is to facilitate the transfer of product data semantics from Computer Aided Design (CAD) program to assembly process planning (APP) in product life-cycle. In this paper, an approach to capture, share and transfer assembly design semantic data from SolidWorks (SW) CAD software to assembly device (robot Sony SRX series) is proposed. The proposed approach is based, on its first stage, on defining and extracting assembly design semantics from a CAD model using SolidWorks Application Programmable Interface (SW-API). The second stage of the proposed approach includes sharing and integrating the extracted assembly design semantics with assembly robot device by using three-layer ontology structure. In this layered ontology, different types of ontologies are proposed for each layer: general foundation ontology for the first, domain ontologies for the second and application ontology for the third. Each of these layers aids in defining concepts, relations and properties in assembly design domain and APP domain. Ultimately, the proposed ontology will be used to integrate both domains in product-life cycle.

Paper #29

Enhancing Community Detection in Social Network using Ontology

Salma Khattab, Abeer ElKorany and Akram Salah
Faculty of Computer and Information, Cairo University, Giza, Egypt

Keywords: Ontology, Semantic User Profile, Similarity, Modularity, Community Detection.

Abstract: In recent years, social networks have been spread widely. Within social network, people tend to form communities in order to have more chances to share opinions, experiences and expertise. Users in social networks belong to the same community according to their behaviour and common interest. This paper presents a semantic approach for community extraction based on identifying the interest of user in order to group them into communities. An ontological user profile is created indicating user interest that is associated with items domain ontology. A set of experiments was applied using real dataset (BookCrossing) to measure the accuracy of the proposed semantic-based framework.

Paper #40

Temporal Interval Modeling for UML Activity Diagrams

Joanna Isabelle Olszewska

University of Gloucestershire, Cheltenham, U.K.

Keywords: UML, Activity Diagram, DL, Temporal Intervals, Allen's Spatial Relations, Knowledge Engineering.

Abstract: The Unified Modeling Language (UML) is a conceptual language which is nowadays widely used to describe and model complex processes. At the origin, UML is focused on structural and behavioural modeling rather than on temporal one, whereas time notion is omnipresent in real-world processes. Hence, in this paper, we propose to introduce new temporal notations complementing standard UML activity diagrams notations in order to model the time aspect of processes in terms of temporal interval concepts. Experiments within real-world scenario have demonstrated the usefulness and efficiency of our developed approach.

Poster Session 3
16:30 - 17:30

KMIS
Foyer

Paper #6

Everyday Lives and Practices of a Joint Venture in the Iranian Construction Sector

Laurent Scaringella¹ and François BurtscHELL²¹ ESC Rennes School of Business, Rennes, France² Freyssinet Office, Mashhad, Iran, Islamic Republic of

Keywords: Radical Innovation, Joint Venture, Construction Sector.

Abstract: We investigate the collaboration between an Iranian and a French company in a joint venture aimed at developing radical innovation in the construction sector. We identify the challenges involved, the barriers to technological change, and the difficulties of transferring knowledge related to absorptive capacity. Our findings indicate that radical innovation is characterized by safety, quality, and planning challenges which engender delays, non-conformity to specifications, and additional costs. We contribute to strengthen the understanding of Iranian technology development by focusing on radical innovation standards, joint venture specific learning dyads, and complex knowledge transfer.

Paper #17

Exploration Participants Engagement in Organisational Knowledge Sharing

Mohammadbashir Sedighi¹, Stephan Lukosch¹, Sander van Splunter², Frances Brazier³ and Cees van Beers¹¹ Delft University of Technology, Delft, The Netherlands² Delft University of Technology, Hoorn, The Netherlands³ TU Delft, Delft, The Netherlands

Keywords: Knowledge Management System, Engagement, Participatory System.

Abstract: The importance of knowledge sharing within most

organisations is well recognised. While abundant KM systems have been matured to encourage individual engagement in knowledge sharing, practical evidences show a low success rate of KM systems. This paper reports on a qualitative exploratory multi-case study to explore level participants' engagement in knowledge sharing along the design principle for engagement of participatory systems. Results show that KM systems using a combined approach of supply- and demand side KM strongly influence participants' engagement for knowledge sharing.

Paper #38

Model of Organizational Readiness to Implement Mentoring

Malgorzata Baran

Collegium Civitas, Warsaw, Poland

Keywords: Formal Mentoring in Organizations, Model to Implement Mentoring, Mentoring Programs, Mentoring Implementation Process.

Abstract: The article presents the manners in which mentoring can be used by organizations. Prior to the implementation of mentoring programs, the organization's preparation level for implementing such programs should be evaluated. Evaluating the organization's readiness forms part of the pre-implementation analysis, which allows the determination of the organization's initial level of commitment to employee development and talent retention. The proposed model of preparing the organization for implementing mentoring programs enables management to determine the scope of human resource management changes that will be introduced. The tool used to carry out the research takes into consideration four areas of organization management: the organization's values and strategies; the development of the organization and its employees; setting objectives, conducting performance appraisals and motivating employees; cooperation and relationships within the organization. The article outlines the procedure of using the tool to carry out research, including both data gathering and analysis. The presented tool may prove useful during the pre-implementation analysis phase in all kinds of organizations that wish to introduce mentoring into their operations.

Keynote Lecture
17:30 - 18:30

IC3K
Room Pinhão

Mining Satellite Images for Census Data Collection: A Study Using the Google Static Maps Service

Frans Coenen

University of Liverpool, Liverpool, U.K.

Abstract: Census collection is a common practice throughout the world. However, the process is expensive and resource intensive. This is especially the case in areas which feature poor communication and transportation networks. A cost effective alternative is to use high-resolution satellite imagery to obtain a census approximation at a significantly reduced cost. This can be achieved by building a predictor that can label households, that feature in satellite image data, according to "family" size. The fundamental idea is to segment satellite images so as to obtain satellite sub-images describing individual households and to represent these segmentations in a manner conducive to household "family" size prediction. A number of representations are considered: graph-based, histogram based and texture based. By pairing each represented household with known census data,

namely family size, a predictor can be constructed to predict household size according to the nature of each representation. The generated predictor can then be used to provide a quick and easy mechanism for the approximate collection of census data that does not require significant resource. To generate the desired predictor training data was obtained by collecting "on ground" census data and matching this to satellite imagery. The test site for the work was a collection of villages lying in the Ethiopian hinterland. The operation of the proposed predictor was evaluated using test data collected in the same manner as the training data, and by utilizing the predictor in the context of a "large scale" study for an area of the Ethiopian hinterland for which the population had been previously recorded.

Closing Session
18:30 - 18:45

IC3K
Room Pinhão

Friday, 11

Author Index

- Abdelhedi, F. 28
 ABED, M. 49
 Abel, M. 53
 Adlemo, A. 21, 44
 Ahmed, N. 23
 Akaichi, J. 42
 Akama, K. 27, 52
 Alabdulrahman, R. 35
 Alapaholuoma, T. 23
 Albert, B. 26
 Alfahhood, S. 26
 Alyahyan, S. 47
 Amdouni, E. 31
 Andreeva, O. 43
 Anseeuw, J. 47
 Arnaout, H. 47
 Arroyo, L. 21
 Arroyo-Fernández, I. 22
 Asis, O. 31
 Assaad, M. 40
 Atigui, F. 28
 Attia, O. 30
 Auer, S. 27
 Ayadi, A. 48
 Azmy, M. 30
 Azzouni, K. 30
- Baran, M. 57
 Barbantan, I. 51
 Beaumont, A. 38
 Beers, C. 57
 Beke, A. 50
 Bekkum, M. 42
 Belo, O. 39
 Ben-Av, R. 35
 Bernst, I. 54
 Bertrand, F. 48
 Beuvron, F. 26
 Blaise, J. 22, 31
 Blaszczyk, T. 44
 Bleiweiss, A. 47
 Bobrikhin, A. 31
 Bona, A. 41
 Borit, M. 35
 Brahim, A. 28
 Braunschweig, K. 50
 Brazier, F. 57
 Broyer, S. 37
 Burtshell, F. 57
 buschetti, A. 54
- Cabitz, F. 24
 Cabrera-Diego, L. 22
 Caldarella, E. 55
 Câmara, M. 36
 Carta, S. 38
 Carvalho, M. 39
 Chan, K. 35
 Charrier, J. 26
 Chirtes, D. 44
 Chodorek, M. 43
 Ciarelli, P. 55
 Costa, A. 29
 Cremona, L. 24
 Cunha, T. 29
- Davis, F. 35
 Dax, J. 21
 Dinsoreanu, M. 40
 Dolgikh, D. 41
 Dudek, I. 22, 31
- Eftimov, T. 29
 El-Makky, N. 30
 Elbassuoni, S. 47
 ElKorany, A. 56
 Emeira, A. 30
 Emmenegger, S. 44
 Esnaola-Gonzalez, I. 36
 Essaid, A. 23
 Exman, I. 25, 29, 39
- Fan, G. 39
 Fan, H. 53
- Farrash, M. 47
 Faucheu, J. 42
 Ferrari, A. 53
 Fersini, E. 37
 Figueira, A. 36
 Formica, A. 52
 Fraga, A. 21, 37
 Frochte, J. 54
 Fu, B. 48
 Fujita, H. 30
 Fukumoto, F. 35
 Fusco, G. 48
- Gallant, R. 25
 Gargouri, F. 56
 Gauch, S. 26
 Génova, G. 25
 Ghézala, H. 52
 Gibaud, B. 31
 Gnesi, S. 53
 Goczyla, K. 55
 Gomes, P. 25, 26
 Gómez, J. 43
 Gonzalez, R. 24, 38
 González-Gallardo, C. 41
 Gorlacheva, E. 31
 Grangel-González, I. 27
 Griffith, J. 30
 Grüninger, M. 22
 Gudkov, A. 31
 Guerini, C. 24
 Guimaraes, N. 36
- HaCohen-Kerner, Y. 30, 31
 Hailaj, L. 27
 Hamid, O. 40
 Hartig, K. 39
 Hasan, B. 56
 Hasna, O. 41
 Hassan, M. 30
 Hästbacka, D. 23
 Hinkelmann, K. 44
 Hua, Y. 23
 Hussein, O. 30
 Huynh, A. 48
- lbba, S. 54
 Idri, A. 31, 41
 Ileri, I. 30
 Ismail, M. 21
- Jafjaf, A. 49
 Jagesar, R. 51
 Järp, P. 38
 Jeedi, A. 56
 Jelínek, I. 41
 Johansson, M. 21
 Johnson, B. 55
 Juganaru-Mathieu, M. 42
- Kannisto, P. 23
 Kapahnke, P. 27
 Kaptein, F. 42
 Karagoz, P. 30
 Karanth, P. 37
 Karpus, A. 55
 Katz, P. 29
 Kaur, B. 37
 Kendall, T. 55
 Kharrat, M. 56
 Khattab, S. 56
 Kiefer, B. 42
 Kim, S. 43
 Kim, T. 32
 Klimenkov, S. 51
 Klusch, M. 27
 Knecht, C. 26
 Koci, E. 26, 50
 Kofler, A. 41
 Koit, M. 23
 Korošec, P. 29
 Koznov, D. 43
 Krčál, M. 28
 Krieger, H. 42
- Kroll, J. 40
 Kubiš, M. 28
 Kumalakov, B. 53
 Küng, J. 37
 Kurzhöfer, J. 43
 Kussainov, K. 53
 Kusu, H. 40
- Labille, K. 26
 Lagatie, R. 47
 Lakhawat, P. 56
 Lantow, B. 49, 52
 Laudy, C. 48
 Lehner, W. 26, 50
 Lemnar, C. 41
 LEMOINE, J. 49
 Lemondjava, V. 31
 Leushin, V. 31
 Liesaputra, V. 47
 Lima, G. 55
 Liu, G. 53
 Llorens, J. 21
 Lloret, E. 51
 Locoro, A. 24
 Lohmann, S. 27
 Lortal, G. 48
 Loula, P. 23
 Luengo-Valderrey, M. 28
 Lukosch, S. 57
- Ma, Y. 48
 Maglyas, A. 43
 Mahesh, K. 37
 Mahmoud, T. 43
 Maire, J. 26
 Mäkinen, J. 40
 Mäkiö, J. 40
 Malheiro, R. 25, 26
 Manaa, M. 42
 Martins, P. 25
 Mazzola, L. 27
 Melen, R. 21
 Mendes, J. 28, 41
 Meneses-Ortegón, J. 38
 Messina, E. 37
 Michalke, V. 39
 Minelli, E. 24
 Mishra, M. 56
 Misséri, V. 53
 Missikoff, M. 52
 Mohaghegh, M. 47
 Molamphy, D. 36
 Moreno, V. 21
 Muñoz, L. 22
 Muromtsev, D. 43
- Nagi, K. 30
 Nantajeewarawat, E. 27, 52
 Ndengue, J. 42
 Neerincx, M. 42
 Nikula, U. 43
 Noia, T. 55
 Nonaka, A. 35
 Nova, N. 24
 Nozza, D. 37
- Obeso, M. 28
 Ohashi, T. 40
 Ohmori, T. 30
 Ok, H. 32, 43
 Oliveira, E. 55
 Oliveira, H. 25
 Olszewska, J. 57
 Ongena, F. 47
 Onori, M. 56
 Ouafouh, S. 31
 Ouamani, F. 52
 Onissi, H. 42
- Paiva, R. 25, 26
 Palacios, L. 48
 Panda, R. 26
 Pani, F. 54
 Paquet, E. 35

Parra, E.	21	Scaringella, L.	57	Tuovinen, L.	54
Peters, R.	42	Sedighi, M.	57	Turck, F.	47
Picariello, A.	55	Seljak, B.	29	Tzach, Y.	31
Pillet, M.	26	Shao, L.	50		
Pinardi, S.	21	Shapira, O.	29	Vasile, I.	40
Pipek, V.	21	Shintani, T.	30	Vázquez, H.	42
Pirinen, R.	40	Sierra, G.	22, 41	Venant, S.	47
Pismak, A.	51	Sigmanek, C.	49	Verstichel, S.	47
Pope, D.	30	Simon, L.	48	Vidal, M.	27
Popescu, O.	22	Soares, C.	29	Vihervaara, J.	23
Potolea, R.	40, 41, 51	Somani, A.	56	Viktor, H.	35
Pourabbas, E.	52	Sosamphan, P.	47	Vo, N.	22
Prabhakar, A.	49	Speicher, D.	21, 37	Vollenbroek, W.	27
Prast, M.	41, 44	Splunter, S.	57	Vries, S.	27
		Spruit, M.	35, 51, 54	Vujic, M.	27
Radchenko, I.	43	Sudolska, A.	43		
Raj, R.	38	Sulaiman, S.	43	Wang, F.	53
Ravarini, A.	24	Suzuki, Y.	35	Wang, J.	50
Rendón, A.	41	Syed, S.	35	Wang, W.	47
Reynaud, C.	48	Szaszák, G.	50	Watanabe, M.	40
Rinaldi, A.	55			Wikander, J.	56
Roatti, H.	55	Taglino, F.	52	Witschel, F.	44
Robbers, S.	43	Takeda, R.	40	Witschel, H.	53
Romero, O.	26	Tammet, T.	38	Wong, K.	35
Ronkainen, A.	44	Tan, H.	21	Wong, S.	38
		Tan, M.	55		
Sacco, M.	55	Tarasov, V.	21	Xie, S.	39
Saia, R.	38	Thi, Q.	23		
Saijo, M.	40, 48	Thiele, M.	26, 50	Yarkeev, A.	51
Saissi, Y.	41	Thönssen, B.	44	Yasui, Y.	35
Salah, A.	56	Tolan, E.	55	Yongchareon, S.	47
Saleh, M.	53	Toledo, C.	54		
Salonen, A.	40	Tomeo, P.	55	Zaiane, O.	35
Sanne, U.	53	Tomingas, K.	38	Zander, S.	23
Sannino, C.	48	Torgo, L.	36	Zanni-Merk, C.	23, 26, 48
Saoud, N.	52	Torres-Moreno, J.	22, 41	Zellou, A.	31, 41
Sartori, F.	21	Tsopa, E.	51	Zhao, Y.	37, 55
Sato, T.	48	Tsukagoshi, H.	40	Zurfluh, G.	28
Satta, V.	24	Tündik, M.	50		