

# DocEng 2013

September 10-13, Florence, Italy

## Conference Programme

### Tuesday, 10 September

The workshops DChanges 2013 and DH-CASE 2013 take up the full day. The workshop "Reimagining Digital Publishing for Technical Documents" takes place in the morning only. All participants with a workshop registration (standalone or additional) are free to attend any workshop, regardless of whether they have an accepted paper or not.

0930 - 1050	<b>Morning workshops</b> (DChanges 2013, DH-CASE 2013, Reimagining Digital Publishing for Technical Documents)
1050 - 1110	Break
1110 - 1250	Morning workshops (contd.)
1250 - 1400	Lunch
1400 - 1510	<b>Afternoon workshops</b> (DChanges 2013, DH-CASE 2013)
1510 - 1530	Break
1530 - 1730	Afternoon workshops (contd.)

### Wednesday, 11 September

0900 - 0915	<b>Introduction</b>
0915 - 1035	<b>ProDoc Doctoral Consortium</b> (Session Chair: Cerstin Mahlow)
1035 - 1055	Break
1055 - 1155	<b>Keynote</b> (Session Chair: Simone Marinai)  Symbolic Machine Learning Methods for Historical Document Processing <i>Floriana Esposito</i>
1155 - 1300	<b>Digital Humanities</b> (Session Chair: Fabio Vitali)  Revisiting a Summer Vacation: Digital Restoration and Typesetter Forensics <i>Steven Bagley, David Brailsford, Brian Kernighan</i>  Interacting with Digital Cultural Heritage Collections via Annotations: The CULTURA Approach <i>Maristella Agosti, Owen Conlan, Nicola Ferro, Cormac Hampson, Gary Munnelly</i>  Early Modern OCR Project (eMOP) at Texas A&M University: Using Aletheia to Train Tesseract <i>Katayoun Torabi, Jessica Durgan, Bryan Tarpley</i>
1300 - 1400	Lunch
1400 - 1530	<b>Dealing with Multiple Versions</b> (Session Chair: John Lumley)  Uncertain Version Control in Open Collaborative Editing of Tree-Structured Documents <i>M. Lamine Ba, Talel Abdessalem, Pierre Senellart</i>

LSEQ: An Adaptive Structure for Sequences in Distributed Collaborative Editing  
*Brice Nédelec, Pascal Molli, Achour Mostefaoui, Emmanuel Desmontils*

Introduction to the Universal Delta Model  
*Gioele Barabucci*

Version Aware LibreOffice Documents  
*Meenu Pandey, Ethan Munson*

1530 - 1550 Break

1550 - 1720 **Search and Sense Making** (Session Chair: Monica Landoni)

Interactive Text Document Clustering using Feature Labeling  
*Seyednaser Nourashrafeddin, Evangelos Milios, Dirk Arnold*

A Graph-based Topic Extraction Method Enabling Simple Interactive Customization  
*Ajitesh Srivastava, Axel J. Soto, Evangelos Milios*

Searching Online Book Documents and Analyzing Book Citations  
*Zhaohui Wu, Sujatha Das, Zhenhui Li, Prasenjit Mitra, C. Lee Giles*

Near Duplicate Detection in an Academic Digital Library  
*Kyle Williams, C. Lee Giles*

### Thursday, 12 September

0900 - 1000	<b>Keynote</b> (Session Chair: Kim Marriott)  Turning Documents into Evidence: Structuring Digitised Libraries to support Scholarship <i>Alastair Dunning</i>
1000 - 1030	<b>Architecture and Processes</b> (Session Chair: Ethan Munson)  Augmenting Digital Documents with Negotiation Capability <i>Jerzy Kaczorek, Bogdan Wiszniewski</i>  A Framework for Usage-based Document Reengineering <i>Madjid Sadallah, Benoît Encelle, Azze-Eddine Maredj, Yannick Prié</i>
1030 - 1050	Break
1050 - 1245	<b>Document Recognition and Analysis I</b> (Session Chair: Charles Nicholas)  An Adaptive Thresholding Algorithm Based on Edge Detection and Morphological Operations for Document Images <i>Renata Neves, Cleber Zanchettin, Carlos Mello</i>  Information Extraction Efficiency of Business Documents Captured with Smartphones and Tablets <i>Daniel Esser, Klemens Muthmann, Daniel Schuster</i>

## Conference Programme (contd.)

	Dominant Color Segmentation of Administrative Document Images by Hierarchical Clustering <i>Elodie Carel, Vincent Courboulay, Jean-Christophe Burie, Jean-Marc Ogier</i>		Using RDFS/OWL to ease Semantic Integration of Structured Documents <i>Jean-Yves Vion-Dury</i>
	Optical Font Recognition using Conditional Random Field <i>Aziza Satkhodzina, Ildus Ahmadullin, Jan Allebach</i>		Reviewing the TEI ODD System <i>Sebastian Rahtz; Lou Burnard</i>
	A Shape-Based Layout Descriptor for Classifying Spatial Relationships in Handwritten Math <i>Francisco Alvaro, Richard Zanibbi</i>		Assisted Editing in the Biomedical Domain: Motivation and Challenges <i>Fabio Rinaldi</i>
	Evaluating Glyph Binarizations Based on their Properties <i>Shira Faigenbaum, Arie Shaus, Barak Sober, Eli Turkel, Eli Piasezky</i>		Managing Content, Metadata and User-created Annotations in Web-Based Applications <i>Marián Šimko; Martin Franta; Martin Habdák; Petra Vrabecová</i>
	Document Noise Removal Using Sparse Representations Over Learned Dictionary <i>Thanh-Ha Do, Salvatore Tabbone, Oriol Ramos Terrades</i>	1015 - 1035	<b>BoF (Discussion Groups) Report</b> (Session Chair: Patrick Schmitz)
		1035 - 1055	Break
1245 - 1300	<b>BoF Session (Discussion Groups)</b> (Session Chair: Patrick Schmitz)	1055 - 1155	<b>Keynote</b> (Session Chair: Kim Marriott)
1300 - 1400	Lunch		Not a Hamburger. Or, How I Stopped Worrying and Learned to Love the PDF <i>Steve Pettifer</i>
1400 - 1530	<b>Document Layout and Presentation Generation I</b> (Session Chair: Steven Bagley)	1155 - 1250	<b>Multimedia I</b> (Session Chair: Dick Bulterman)
	Functional, Extensible, SVG-based Variable Documents <i>John Lumley</i>		Multimedia Authoring Based on Templates and Semi-automatic Generated Wizards <i>Roberto Gerson Azevedo, Rodrigo Costa Mesquita Santos, Eduardo Cruz Araújo, Luiz Fernando Gomes Soares, Carlos De Salles Soares Neto</i>
	Hierarchical Probabilistic Model for News Composition <i>Ildus Ahmadullin, Niranjan Damera-Venkata</i>		Content-based Copy and Paste from Video Documents <i>Laurent Denoue, Scott Carter, Matthew Cooper</i>
	Balancing Font Sizes for Flexibility in Automated Document Layout <i>Ricardo Piccoli, João Batista Oliveira</i>		MoViA: A Mobile Video Annotation Tool <i>Bruna Cunha, Olibário Neto, Maria da Graça Pimentel</i>
	Splitting Wide Tables Optimally <i>Mihai Bilauca, Patrick Healy</i>	1250 - 1305	<b>Poster Introductions</b> (Chair: Peter King)
1530 - 1550	Break	1305 - 1420	<b>Lunch/Posters and Demos</b>
1550 - 1705	<b>Document Recognition and Analysis II</b> (Session Chair: Steven Simske)	1420 - 1525	<b>Multimedia II</b> (Session Chair: Cécile Roisin)
	Visual Saliency and Terminology Extraction for Document Annotation <i>Benjamin Duthil, Mickaël Coustaty, Vincent Courboulay, Jean-Marc Ogier</i>		Multimedia Document Synchronization in a Distributed Social Context <i>Jack Jansen, Pablo Cesar, Dick Bulterman</i>
	Supervised Polarity Classification of Spanish Tweets Based on Linguistic Knowledge <i>David Vilares, Miguel Alonso, Carlos Gómez-Rodríguez</i>		NCL4WEB - Translating NCL Applications to HTML5 Web Pages <i>Esdras Silva, Joel dos Santos, Débora C. Muchaluat Saade</i>
	Hi-Fi HTML Rendering of Multi-format Documents in DoMinUS <i>Stefano Ferilli, Floriana Esposito, Domenico Redavid</i>	1525 - 1545	Go Beyond Boundaries of iTV Applications <i>Caio Cesar Viel, Erick Lazaro Melo, Maria da Graça Pimentel, Cesar Teixeira</i>
	PDFX: Fully-automated PDF-to-XML Conversion of Scientific Literature <i>Alexandru Constantin, Steve Pettifer, Andrei Voronkov</i>	1545 - 1600	Break
	Recognising Document Components in XML-based Academic Articles <i>Angelo Di Iorio, Silvio Peroni, Francesco Poggi, Fabio Vitali, David Shotton</i>	1600 - 1645	<b>Multimedia II (Cont)</b> (Session Chair: Cécile Roisin)
1705 - 1725	<b>Workshop Summary</b> (Chair: David Brailsford)		Interchanging and Preserving Presentation Recordings <i>Kai Michael Höver, Max Mühlhäuser</i>
			<b>Document Layout and Presentation Generation II</b> (Session Chair: Niranjan Damera-Venkata)
			No Need to Justify Your Choice: Pre-compiling Line Breaks to Improve eBook Readability <i>Alex Pinkney, Steven Bagley, David Brailsford</i>
			Reflowing and Annotating Scientific Papers on eBook Readers <i>Simone Marinai</i>
			Automatic Generation of Limited-depth Hyper-documents from Clinical Guidelines <i>Mark Truran, Jonathan Siddle, Gersende Georg, Marc Cavazza</i>
			<b>Closing Notes</b>
		1645 - 1700	
<hr/>			
<b>Friday, 13 September</b>			
0900 - 1015	<b>Metadata and Annotation</b> (Session Chair: Stefano Ferilli)		
	Improving Term Extraction by Utilizing User Annotations <i>Jozef Harinek; Marián Šimko</i>	1645 - 1700	