ACM DocEng 2014 Conference Program & Information

Sponsor:

Supporters:

DocEng 2014 is built on the tradition of past symposia previously held around the world including Florence (2013) Paris (2012), Mountain View (2011), Manchester (2010) and Munich (2009).

The ACM Symposium on Document Engineering provides an annual international forum for presentations and discussions on principles, tools and processes that improve our ability to create, manage and maintain documents. It is sponsored by ACM by means of the ACM SIGWEB Special Interest Group. All DocEng Proceedings are available through the ACM Digital Library.

The conference is located the Hewlett-Packard facilities in Fort Collins, Colorado (3404 E. Harmony Rd. Fort Collins CO 80528). On the 17th, for the conference Social Activity will occur at Pinball Jones in Old Town Fort Collins. The Conference Banquet will be held at the Marriott (conference hotel) on Horsetooth Avenue in Fort Collins. The detailed schedule and list of attendees is on the following pages.

Welcome to Colorado!

ACM DocEng 2014 Conference Schedule

TUESDAY, 16 SEPTEMBER, 2014

Tuesday is the day for the DocEng 2014 Workshops! The workshops will be held in four separate conference areas clustered around the Lobby of Building 3 of the Hewlett Packard site on 3404 E. Harmony Road, Fort Collins CO 80528 (see https://www.google.com/maps/@40.5264896,-105.0143516,17z). You turn into the HP plant off Harmony Road and Building 3 is more or less straight north with the lobby on the west side of Building 3.

The workshops themselves are widely varied in content, just like the DocEng main conference. There are strictly speaking three workshops and one tutorial. This year, each workshop/tutorial will take up the full day.

1. **DChanges 2014** - Document Changes: Modeling, Detection, Storage and Visualization (Organizers: Gioele Barabucci, Uwe M. Borghoff, Angelo Di Iorio, Sonja Maier, Ethan Munson)

Change tracking and detection, versioning and collaborative editing. This edition in particular is focused on interpretation, visualization, processing and exploitation of changes. The final program outline is on-line at https://sites.google.com/site/dchanges14/program.

2. **SemADoc** – Semantic Analysis of Documents (Organizers: Carlotta Domeniconi, Evangelos Milios)

Document content analysis and semantic enrichment to generate a layer of semantic description of documents that is useful for document management tasks, such as semantic information retrieval, conceptual organization and clustering of document collections for sense making, semantic expert profiling, and document recommender systems. The final program outline is on-line at https://sites.google.com/site/semadoc2014/6-workshop-programme.

3. **DH-CASE II** – Collaborative Annotations in Shared Environments: metadata, tools and techniques in the Digital Humanities (Organizers: Patrick Schmitz, Laurie Pearce, Quinn Dombrowski)

The tools and environments that support annotation, broadly defined, including modeling, authoring, analysis, publication and sharing. See http://research-it.berkeley.edu/dhcase2014.

4. **PDF Tutorial** (Organizers: Matthew Hardy and Steven Bagley)

The focus of this tutorial is to give attendees practical knowledge of how to create and handle PDFs that take advantage of the non-print features of PDF to provide rich access to the information within, using a variety of commercial and open-source tools. We will get under-the-hood of PDF and analyze the poor practices that cause PDFs to be inaccessible; see how to access the text and graphics within a PDF; and the features of PDF that can be used to make the information much more accessible. We will also discuss some of the new ISO standards that provide profiles for producing Accessible PDFs. See next page for more details.


Welcome to Colorado!

Timing of Events on 16 September

Event	Location	Timing
Workshop/Tutorials Session 1	CVC, CTC, 3Lower-C5, 3Lower-C7	0900-1030
Coffee Break	Customer Common Area + Onsite Starbucks	1030-1100
Workshop/Tutorials Session 2	CVC, CTC, 3Lower-C5, 3Lower-C7	1100-1230
Lunch	Customer Common Area + Onsite Starbucks after 1300	1230-1330
Workshop/Tutorials Session 3	CVC, CTC, 3Lower-C5, 3Lower-C7	1330-1500
Coffee Break	Customer Common Area + Onsite Starbucks	1500-1530
Workshop/Tutorials Session 4	CVC, CTC, 3Lower-C5, 3Lower-C7	1530-1700

Workshop/Tutorial Locations are the Customer Visitor Center (CVC), the Customer Training Center (CTC), the Conference Room Building 3 Lower-C5 and the Conference Room 3 Lower-C7. All rooms have projectors and flip charts, etc. to encourage interactive discussion.

PDF Tutorial

09:00 **Welcome and Introduction**

09:15 **PDF Internals** This session looks at how a PDF is structured; how the content on the page is described from the bottom-up, and how those pages – and the resources they use -- are fitted together to form a PDF file.

10:30 Coffee

11:00 **Higher-level structures** How more abstract higher level structures have been built on top of the basic PDF structures to support accessibility, and navigation around the PDF.

12:30 Lunch

13:30 **Rogues Gallery** A look at a “rogues gallery” of bad PDFs in light of what we learnt in the morning and explaining why they cause problems.

14:00 **Creating and Manipulating PDF programmatically** A look at the various options for creating and manipulating PDFs programmatically. Both open-source and commercial options will be considered, and we’ll give tips to avoid creating PDFs that end up in the “rogues gallery”.

15:00 Coffee

15:30 **The Future** A look at where PDF is heading and what is new in PDF 2.0

16:15 Close

Welcome to Colorado!

WEDNESDAY, 17 SEPTEMBER, 2014

Wednesday is the opening day of the DocEng 2014 Conference. The conference will be held in the large conference room (Upper-C7 and Upper-C9) of Building 6 of the Hewlett Packard site on 3404 E. Harmony Road, Fort Collins CO 80528 (see https://www.google.com/maps/@40.5264896,-105.0143516,17z). You turn into the HP plant off Harmony Road, and Building 6 is about a quarter mile along the loop road to the right (east). The lobby (where you check in with HP security) is on the south side (center) of the Building 6+Annex complex.

Timing of Events on 17 September

Event	Location	Timing
Security Badging	Building 6 Lobby	0830-0930
Welcome and Introduction	6Upper-C7/C9	0930-0945
ProDoc Doctoral Consortium (Chair: Cerstin Mahlow)	6Upper-C7/C9	0945-1045
Coffee Break	6Upper-C7/C9	1045-1115
Keynote The Evolving Scholarly Record: New Uses and New Forms (Clifford Lynch, Coalition for Networked Information)	6Upper-C7/C9	1115-1215
Birds of a Feather Session: How It works (Chair: Patrick Schmitz)	6Upper-C7/C9	1215-1245
Lunch	Blue Spruce Room near Onsite Starbucks	1245-1345
Recap of the Workshop Sessions (Chair: Sonja Maier)	6Upper-C7/C9	1345-1415
Modeling and Representation (Chair: Steven Bagley)	6Upper-C7/C9	1415-1530
Coffee Break	6Upper-C7/C9	1530-1600
Document Analysis I (Chair: Helen Balinsky)	6Upper-C7/C9	1600-1730
CONFERENCE EVENT	Pinball Jones, Old Town Fort Collins (http://pinballjones.com/)	1900-2100


Welcome to Colorado!

Modeling and Representation (1415-1530)

1415-1445	ActiveTimesheets: extending Web-based multimedia documents with dynamic modification and reuse features (Full Paper)	Diogo Martins and Maria Da Graça Pimentel
1445-1500	Automated CSS Optimization by Logical Reasoning (Short Paper)	Marti Bosch, Pierre Geneves and Nabil Layaida
1500-1515	FlexiFont: A Flexible System to Generate Personal Font Libraries (Application Note)	Wanqiong Pan, Zhouhui Lian, Rongju Sun, Yingmin Tang and Jianguo Xiao
1515-1530	Circular Coding with Interleaving Phase (Short Paper)	Robert Ulichney, Matthew Gaubatz and Steven Simske

Document Analysis I (1600-1730)

1600-1630	A New Sentence Similarity Assessment Measure based on a Three-Layer Sentence Representation (Full Paper)	Rafael Mello, Rafael Lins, Fred Freitas, Steven J. Simske, Bruno Avila, Rodolfo Ferreira and Marcelo Riss
1630-1700	Paper Stitching using Maximum Tolerant Seam under Local Distortions (Full Paper)	Wei Liu, Wei Fan, Jun Sun and Naoi Satoshi
1700-1715	Abstract Argumentation for Reading Order Detection (Short Paper)	Stefano Ferilli, Domenico Grieco, Domenico Redavid and Floriana Esposito
1715-1730	Generating Summary Documents from a Variable-Quality PDF Document Collection (Short Paper)	Jacob Hughes, David Brailsford, Steven Bagley and Clive Adams

Pinball Jones (1900-2100)

Dinner this night is up to you. To put you in the right part of town for a wide selection, Pinball Jones has been reserved from 7 PM to 9 PM for DocEng delegates. Right in the heart of Old Town (below ground just northeast of the corner of Mountain and Collage), Kim Jones and her crew have created a wonderful basement-level vintage and modern pinball arcade (Pinball Jones is on the left side of the basement location--there is another game site to the right which you can pay to play during the event). Each delegate can play to their heart's content along with a free drink (they have a bar) and then decide which of the dozens of nearby eateries to choose for dinner, as desired.

Welcome to Colorado!

THURSDAY, 18 SEPTEMBER, 2014

Thursday is the very heart of the DocEng 2014 Conference. Again, the conference will be centered on the large conference room (Upper-C7 and Upper-C9) of Building 6 of the Hewlett Packard site on 3404 E. Harmony Road, Fort Collins CO 80528 (see https://www.google.com/maps/@40.5264896,-105.0143516,17z).

Timing of Events on 18 September

Event	Location	Timing
Document Analysis II (Chair: Cheng Thao)	6Upper-C7/C9	0900-1045
Coffee Break	6Upper-C7/C9	1045-1115
Keynote Web-Intrinsic Interactive Documents (Tony Wiley, HP Exstream R&D)	6Upper-C7/C9	1115-1215
Birds of a Feather Session: The Results (Chair: Patrick Schmitz)	6Upper-C7/C9	1215-1245
Lunch	6Upper-C7/C9 and Onsite Starbucks	1245-1345
Collection, Systems and Management (Chair: Jean-Yves Vion-Dury)	6Upper-C7/C9	1345-1530
Coffee Break	6Upper-C7/C9	1530-1600
Applications I (Chair: Dick Bulterman)	6Upper-C7/C9	1600-1730
CONFERENCE EVENT Banquet (Drinks, Dinner, Awards)	Conference Hotel Fort Collins Marriott (http://www.marriott.com/hotels/hotel-information/travel/ftcco-fort-collins-marriott/)	1830-2130 (Host Bar starts at 1830, dinner at 1915)

Conference Banquet (1830-2130)

The conference banquet is at the Fort Collins Marriott, from 6:30 PM – 9:30 PM in the Pavilion (if weather is good) or the Ballroom Salon (if the weather is poor). Conference organizers will be acknowledged and the location of DocEng2015 will be revealed.


Welcome to Colorado!

Document Analysis II (0900-1045)

0900-0930	Transforming Graph-based Sentence Representations to Alleviate Overfitting in Relation Extraction (Full Paper)	Rinaldo Lima, Jamilson Batista, Rafael Ferreira, Fred Freitas, Rafael Lins, Steven Simske and Marcelo Riss
0930-1000	Ruling Analysis and Classification of Torn Documents (Full Paper)	Markus Diem, Florian Kleber and Robert Sablatnig
1000-1030	On Automatic Text Segmentation (Full Paper)	Boris Dadachev, Alexander Balinsky and Helen Balinsky
1030-1045	P-GTM: Privacy-Preserving Google Tri-gram Method for Semantic Text Similarity (Short Paper)	Owen Davison, Abidalrahman Mohammad and Evangelos Milios

Collection, Systems and Management (1345-1530)

1345-1415	Fine-grained Change Detection in Structured Text Documents (Full Paper)	Hannes Dohrn and Dirk Riehle
1415-1445	Classifying and Ranking Search Engine Results as Potential Sources of Plagiarism (Full Paper)	Kyle Williams, Hung-Hsuan Chen and C. Lee Giles
1445-1515	An Ensemble Approach for Text Document Clustering using Wikipedia Concepts (Full Paper)	Seyednaser Nourashrafeddin, Evangelos Milios and Dirk Arnold
1515-1530	Image-Based Document Management: Aggregating Collections of Handwritten Forms (Short Paper)	John Barrus and Edward Schwartz

Applications I (1600-1730)

1600-1630	ARTIC: Metadata Extraction from Scientific Papers in PDF using Two-Layer CRF (Full Paper)	Alan Souza, Viviane Moreira and Carlos Heuser
1630-1645	Connecting Content and Annotations with LiveStroke (Short Paper)	Michael Gormish and John Barrus
1645-1700	Building digital project rooms for web meetings (Short Paper)	Laurent Denoue, Matthew Cooper, Andreas Girgensohn and Scott Carter
1700-1715	The Virtual Splitter: Refactoring Web Applications for the Multiscreen Environment (Short Paper)	Mira Sarkis, Cyril Concolato and Jean-Claude Dufourd
1715-1730	SimSeeX: A Similar Document Search Engine (Application Note)	Kyle Williams, Jian Wu and C. Lee Giles

Welcome to Colorado!

FRIDAY, 19 SEPTEMBER, 2014

Friday brings the DocEng 2014 Conference to a close. Once more, the conference will be centered on the large conference room (Upper-C7 and Upper-C9) of Building 6 of the Hewlett Packard site on 3404 E. Harmony Road, Fort Collins CO 80528 (see https://www.google.com/maps/@40.5264896,-105.0143516,17z).

Timing of Events on 19 September

Event	Location	Timing
Generation, Manipulation and Presentation (Chair: Evangelos Milios)	6Upper-C7/C9	0900-1045
Coffee Break	6Upper-C7/C9	1045-1115
Applications II (Chair: Michael Gormish)	6Upper-C7/C9	1115-1245
Lunch	6Upper-C7/C9 and Onsite Starbucks after 1300	1245-1345
Awards, Closing Notes and Farewell Message	6Upper-C7/C9	1345-1415

Generation, Manipulation and Presentation (0900-1045)

0900-0930	Pagination: It's what you say, not how long it takes to say it (Full Paper)	Joshua Hailpern, Niranjan Damera-Venkata and Marina Danilevsky
0930-1000	Extracting web content for personalized presentation (Full Paper)	Rodrigo Chamun, Daniele Pinheiro, Diego Jornada, João Oliveira and Isabel Manssour
1000-1030	Truncation: All the News that Fits We'll Print (Full Paper)	Joshua Hailpern, Niranjan Damera-Venkata and Marina Danilevsky
1030-1045	JAR Tool: Using Document Analysis for Improving the Throughput of High Performance Printing Environments (Short Paper)	Mariana Kolberg, Luiz Fernandes, Mateus Raeder and Carolina Fonseca


Welcome to Colorado!

Applications II (1115-1245)

1115-1145	Humanist-centric tools for Big Data: Berkeley Prosopography Services (Full Paper)	Patrick Schmitz and Laurie Pearce
1145-1215	The Impact of Prior Knowledge on Searching in Software Documentation (Full Paper)	Klaas Andries de Graaf, Peng Liang, Antony Tang and Hans van Vliet
1215-1230	What Academics Want When Reading Digitally (Short Paper)	Juliane Franze, Kim Marriott and Michael Wybrow
1230-1245	A Platform for Language Independent Summarization (Short Paper)	Luciano Cabral, Rafael Lins, Rafael Mello, Fred Freitas, Bruno Avila, Steven Simske and Marcelo Riss

END OF DOC ENG 2014

ACM DocEng 2014 Conference Attendees

Steven Bagley (University of Nottingham)
Helen Balinsky (Hewlett-Packard)
John Barrus (Ricoh Innovation, Inc.)
Marti Bosch (Inria)
David Brailsford (University of Nottingham)
Dick Bulterman (FX Palo Alto Laboratory)
Luciano Cabral (Federal University of Pernambuco)
Scott Carter (FX Palo Alto Laboratory)
Rodrigo Chamun (PUCRS)
Stephen Coakley (University of Wisconsin-Whitewater)
Boris Dadachev (Cardiff University & HP Labs)
Niranjan Damera-Venkata (Hewlett-Packard Company)
Klaas Andries De Graaf (VU University Amsterdam)
Markus Diem (Vienna UT)
Hannes Dohrn (University of Erlangen-Nürnberg)
Quinn Dombrowski (UC Berkeley)
Jean-Claude Dufourd (Telecom ParisTech)
Stefano Ferilli (University of Bari)
Luiz Gustavo Fernandes (Pontifícia Universidade Católica do Rio Grande do Sul)
Juliane Franze (Fraunhofer ESK)
Fred Freitas (University of Pernambuco)

Welcome to Colorado!

Michael Gormish (Ricoh Innovations, Corp.)
Joshua Hailpern (Hewlett-Packard Laboratories)
Dennis Hamilton (independent)
Matthew Hardy (Adobe Systems)
Greg Harris (University of Southern California)
Tamir Hassan (Hewlett-Packard Laboratories)
Peter King (University of Manitoba)
Zhouhui Lian (Peking University)
Rinaldo Lima (Federal University of Pernambuco)
Thea Lindquist (University of Colorado Boulder Libraries)
Rafael Lins (Federal University of Pernambuco)
Wei Liu (Fujitsu Research and Development Center)
Clifford Lynch (CNI)
Cerstin Mahlow (University of Stuttgart)
Sonja Maier (Universität der Bundeswehr München)
Diogo Martins (Universidade Federal do ABC)
Patrick McLeod (University of North Texas)
Evangelos Milios (Dalhousie University)
Ethan Munson (UW-Milwaukee)
Charles Nicholas (UM-Baltimore County)
Seyednaser Nourashrafeddin (Dalhousie University)
Laurie Pearce (UC Berkeley)
Michael Piotrowski (Leibniz Institute of European History)
Sebastian Rönnau (Zalando SE)
Mira Sarkis (Telecom Paristech)
Patrick Schmitz (UC Berkeley)
Svante Schubert (Freelancer)
Kristen Schuster (University of Missouri)
Eric Scott (The MITRE Corporation)
Jacek Siciarek (Gdansk University of Technology)
Steven Simske (Hewlett-Packard Laboratories)
Alan Souza (UFRGS)
Margaret Sturgill (Hewlett-Packard Laboratories)
Cheng Thao (University of Wisconsin-Whitewater)
Robert Ulichney (Hewlett-Packard Laboratories)
Marie Vans (Hewlett-Packard Laboratories)
Jean-Yves Vion-Dury (Xerox Research)
Nigel Whitaker (DeltaXML Ltd.)
Anthony Wiley (HP Exstream)
Kyle Williams (The Pennsylvania State University)
Johannes Wilm (Fidus Writer)
Michael Wybrow (Monash University)

Welcome to Colorado!