

ACM DocEng 2014

Tutorial & Workshops Program

Sponsor:



Supporters:



TUESDAY, 16 SEPTEMBER, 2014

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. 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.

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



Welcome to Colorado!



PDF Tutorial

(Organizers: Matthew Hardy, Steven Bagley)

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 Break

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 Break

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!



DChanges 2014

Document Changes: Modeling, Detection, Storage and Visualization

(Organizers: Gioele Barabucci, Uwe M. Borghoff, Angelo Di Iorio, Sonja Maier, Ethan Munson)

<https://sites.google.com/site/dchanges14/>

09:00 Opening Session

09:30 Keynote

- **Jean-Yves Vion-Dury** Some Perspectives on Change Management: From Basic Primitive Tools to Context-Aware Applicative Components

10:30 Coffee Break

11:00 Session I Research Topics

- **Stephen Coakley, Jacob Mischka and Cheng Thao** Version-Aware Word Documents
- **Meenu Pandey, Ethan Munson and Cheng Thao** Progress Report on Version-Aware LibreOffice
- **Nigel Whitaker** Understanding Changes in N-Way-Merge: Use-Cases and User Interface Demonstrations
- **Dennis Hamilton** Tracked Changes: Navigating the Document-Format Anti-Pattern

12:30 Lunch

13:30 Session II & Roundtable Discussion Change Tracking

- **Svante Schubert and Patrick Durusau** Untethered, Asynchronous Change Tracking for File Formats (Work-In-Progress / Position Paper)
- **Dennis Hamilton** Progress Report: Protected Change Tracking (Work-In-Progress / Position Paper)
- **Nigel Whitaker** Change-Tracking and Related Work at W3C (Invited Talk)
- Discussion

15:00 Coffee Break

15:30 Session III & Roundtable Discussion Moving Forward with Implementations

- **Johannes Wilm and Daniel Frebel** Real-World Challenges to Collaborative Text Creation (Work-In-Progress / Position Paper)
- Discussion

16:30 Closing Session



Welcome to Colorado!



SemADoc

Semantic Analysis of Documents

(Organizers: Carlotta Domeniconi, Evangelos Milios)

<https://sites.google.com/site/semadoc2014>

09:00 Session I

- Introduction
- **Stefano Ferilli** Automatic Extraction of Semantic Resources from Text (Keynote)

10:30 Coffee Break

11:00 Session II

- **Greg Harris, Anand Panangadan and Viktor Prasanna** Interactive Query Refinement for Boolean Search
- **Eric O. Scott, Haleh Vafaie, Zela Gungordu, Charles E. Horowitz and Bradford C. Brown** Text Mining for Quality Control of Court Records

12:30 Lunch

13:30 Session III

- **Kyle Williams, Madian Khabza and C. Lee Giles** Scholarly Big Data: Information Extraction, Data Mining & Semantics (Keynote)

15:00 Coffee Break

15:30 Session IV

- **Fred Freitas** Graph-based Extractive Summarization at CIn-UFPE: Contributions and Results (Keynote)
- Discussion and Wrap-up



Welcome to Colorado!



DH-CASE II

Collaborative Annotations in Shared Environments: metadata, tools and techniques in the Digital Humanities

(Organizers: Patrick Schmitz, Laurie Pearce, Quinn Dombrowski)

<http://research-it.berkeley.edu/dhcase2014>

9:00 **Introductions**

9:30 **Session I**

- **Michael Piotrowski, Giovanni Colavizza, Florian Thiery and Kai-Christian Bruhn** The Labeling System: A New Approach to Overcome the Vocabulary Bottleneck
- **Lisa Snyder** VSim: Scholarly annotations in real-time 3D environments

10:30 **Coffee Break**

11:00 **Session II**

- **Quinn Dombrowski and Ronelle Alexander** Digital Humanities Development without Developers: Bulgarian Dialectology as Living Tradition
- **Kristen Schuster** Digital Editions of Antiquarian Texts: A Critical Assessment of Metadata and Digital Curation Practices

12:00 **Discussion** Common themes in annotation

12:30 **Lunch**

13:30 **Discussion** Models for reuse and reusability

15:00 **Coffee Break**

15:30 **Discussion** Identifying best practices



Welcome to Colorado!

