

Call for Submissions

Fifth DESI Workshop on Standards for Using Predictive Coding and Other Machine Learning Algorithms Rome, Italy June 14, 2013

<http://www.umiacs.umd.edu/~oard/desi5/>

A Post-Conference Workshop held in conjunction with the
2013 International Conference on Artificial Intelligence and Law

The DESI V workshop will be a platform for the consideration of best practices in using predictive coding and other forms of machine learning algorithms mainly in the context of civil discovery. Submissions are encouraged on emerging best practices with respect to the use of these advanced techniques. Among the open questions we may expect to address are:

- 1) How transparent can and should the process be in sharing seed sets or training sets of documents with opposing parties, including the sharing of privileged documents?
- 2) What differences if any exist between seed sets developed through random sampling versus other forms of judgmental sampling (including picking seed documents using keywords)?
- 3) How are non-relevant documents used to optimize machine learning algorithms and should they be subject to similar disclosure?
- 4) Are there ways in which predictive coding and machine learning methods can be tuned to find highly relevant (“hot”) documents in large collections?
- 5) To what extent is metadata important in tuning predictive coding software to find similarity in documents?
- 6) In light of past research at the TREC Legal Track and elsewhere, are there absolute targets for metrics in recall and precision that could serve as standards in every case, or are achieving certain metrics dependent on the relevant data set and legal context?
- 7) What kinds of best practice standards are needed to help improve mutual understanding of what was actually done, and to improve overall “search quality”?
- 8) How should predictive coding techniques be audited in connection with an entity submitting itself to an ISO 9001 quality measurement process?
- 9) To what extent can and should machine learning approaches be used in other phases of the litigation process, to assist in aspects of the process such as identification, preservation, and collection?
- 10) What are the applications of predictive coding and other forms of machine learning in related “compliance” areas, including regulatory, enforcement, and investigations?

The workshop discussion will be grounded in the results of the recently completed TREC Legal Track, especially where supervised learning methods have shown promising results in terms of being able to more cost-effectively demonstrate rates of recall and precision that approximate the best that could be obtained through other methods, including exhaustive manual review.

Participation is invited from all interested parties, including those with backgrounds in

- Archives and records management
- Artificial intelligence and law
- Cognitive science
- Content analytics
- Computational linguistics
- Digital forensics
- Legal sensemaking
- Human-computer interaction
- Human language technology
- Information retrieval
- Knowledge management
- Legal informatics
- Litigation support
- Natural language processing
- Standards
- Text mining and text classification

Submissions:

Two types of written contributions are invited:

- Brief position papers (typically on the order of 2 pages) describing individual interests, for inclusion (without review) on the DESI V web site and distribution to workshop participants. Brief descriptions of this type are particularly valuable when bringing together diverse research communities. Additionally, these papers can help with our selection of discussants, discussion leaders, and panelists. Position papers are requested by May 8, 2013. Participation in the workshop is open, so prior submission of position papers is strongly encouraged but not strictly required.
- Original papers (4 to 10 pages) describing research or practice. After peer review, accepted papers will be posted on the DESI V website and made available in hard-copy to workshop participants. Authors of accepted research papers will be invited to present their work either as an oral or a poster presentation. Research papers are due on May 1, 2013; decisions will be returned on May 15, 2013.

Submissions should be sent by email to Doug Oard (oard@umd.edu) with the subject line DESI V POSITION PAPER or DESI V RESEARCH PAPER. All submissions received will be acknowledged within 3 days.

Organizing Committee:

Jason R. Baron, National Archives and Records Administration, USA

Jack G. Conrad, Thomson Reuters, Switzerland

Dave Lewis, David D. Lewis Consulting, USA

Debra Logan, Gartner Research, UK

Douglas W. Oard, University of Maryland, USA

Fabrizio Sebastiani, Istituto di Scienza e Tecnologia dell'Informazione, Italy