

CFP: 2nd Workshop on Analytics for Noisy Unstructured Text Data (AND) with SIGIR 2008

Source: <http://sci.tech-archive.net/Archive/sci.lang/2008-05/msg00010.html>

- *From:* "shourya.roy@xxxxxxxxx" <shourya.roy@xxxxxxxxx>
 - *Date:* Wed, 30 Apr 2008 22:08:16 -0700 (PDT)
-

Hi All.

We are organizing the 2nd workshop on Analytics for Noisy Unstructured Text Data (AND) in conjunction with SIGIR 2008 in Singapore on July 24, 2008. We request you to submit your relevant work to this workshop. We also request you to send this CFP to people whom you think would be interested in this workshop.

Thank you very much

=====

SIGIR-08 Workshop

2nd Workshop on Analytics for Noisy Unstructured Text Data
24 July 2008 , Singapore

<http://and2008workshop.googlepages.com/>

Call for Papers

Workshop Description and Objectives

Noise is an unavoidable fact of life. It can manifest itself at the earliest stages of processing in the form of degraded inputs that our systems must be prepared to handle. People are adept when it comes to pattern recognition tasks involving typeset or handwritten documents or recorded speech, machines less-so. From the perspective of downstream processes that take as their inputs the outputs of recognition systems, including document analysis and OCR, noise can be viewed as the errors made by earlier stages of processing, which are rarely perfect and sometimes quite brittle.

Noisy unstructured text data is also found in informal settings such as online chat, SMS, email, message board and newsgroup postings, blogs, wikis and web pages. In addition to the aforementioned recognition errors, such text may contain spelling errors, abbreviations, non-standard terminology, missing punctuation, misleading case information, as well as false starts, repetitions, and pause-filling sounds such as um and uh in the case of speech.

CFP: 2nd Workshop on Analytics for Noisy Unstructured Text Data (AND) with SIGIR 2008

By its very nature, noisy text warrants moving beyond traditional text analytics techniques. Noise introduces challenges that need special handling, either through new methods or improved versions of existing ones. We invite you to submit your own unique perspectives on this important topic.

Topics of Interest (not limited to)

Information Retrieval and Information Extraction on noisy texts
IR-related tasks (classification, clustering, genre recognition, document summarization, keyword search) on noisy texts
Formal models for noise, characterization and classification of noise
Treatment of noisy data in special application fields

- Historical Texts
- Multilingual Texts
- Blogs
- Chat logs/SMS
- Social Network Analysis
- Patent Search
- Optical Character Recognition
- Automated Speech Recognition
- Machine Translation

Data sets, benchmarks and evaluation techniques for analysis of noisy texts

Participation

We hope that the workshop will allow researchers working in areas related to unstructured data analytics, Natural Language Processing, Information Extraction, Information Retrieval, etc., to focus on the needs of users extracting useful information from noisy text. The target audience is a mixture of academia and industry researchers working with noisy text. We believe this work is of direct relevance to domains such as call centers, the world-wide web, and government organizations that need to analyze huge amounts of noisy data.

Important Dates

Paper Submission: May 16th, 2008
Notification of Acceptance: Jun 6th, 2008
Camera-Ready papers due: Jun 20th, 2008
Workshop at SIGIR 2008: Jul 24th, 2008

Submission Requirements

We invite papers up to 8 pages in length in the style specified at <http://and2008workshop.googlepages.com/submission> There will also be a Best Student Paper Award. Papers with a student as the primary author/presenter will be eligible for this award.

Publication

We are currently in negotiation with a leading publisher for the proceedings to be available onsite. We have also received tentative approval for a special issue of a journal for post-workshop publication of selected papers.

CFP: 2nd Workshop on Analytics for Noisy Unstructured Text Data (AND) with SIGIR 2008

Workshop Chairs

Daniel Lopresti

Lehigh University

Shourya Roy

IBM Research, India Research Lab

Klaus U Schulz

University of Munich

L. Venkata Subramaniam

IBM Research, India Research Lab

Workshop contacts

* L. V. Subramaniam lvsubram@xxxxxxxxxxx

* Shourya Roy rshourya@xxxxxxxxxxx

Please visit the workshop website

***** <http://and2008workshop.googlepages.com/> *****

for information about participation and submitting papers.

For general information, please visit the SIGIR website

***** <http://www.sigir2008.org/> *****

.