

Web-Scale Near-Duplicate Search: Techniques and Applications

As bandwidth accessible to average users is increasing, audiovisual material has become the fastest growing data type on the Internet. The impressive growth of the social Web where users can exchange user-generated content contributes to the overwhelming number of multimedia files available. Among these huge volumes of data, there exist large numbers of near-duplicates and copies. File copies are easy to detect using hashes. Near-duplicates are based on the same original content, but have been edited and postprocessed, resulting in different files. Another type of near-duplicate relates to footage about the same event or scene. The detection of near-duplicates poses a challenge to multimedia content analysis, especially when speed, scale, and copied fragment length are pushed to operational levels. Near-duplicates carry both informative and redundant signals, for example, providing rich visual clues for indexing and summarizing videos from different sources. On the other hand, the excessive amount of near-duplicates makes browsing Web videos streamed over Internet an extremely time-consuming task. As a result, there is strong interest from industry, academia, and governmental agencies in Web-scale search, elimination, detection, and use of near-duplicates for various multimedia applications.

Topic

This special issue seeks innovative contributions dedicated to the theme of Web-scale near-duplicate search. Topics of interests include, but are not limited to, the following:

Techniques and algorithms

- near-duplicate and/or partial near-duplicate detection;
- cross-media search of near-duplicates;
- semantic-based detection of near-duplicates;
- framework and algorithm for real-time near-duplicate detection;
- semantic indexing and hashing techniques;
- similarity and perception learning;
- multimedia fingerprint extraction; and
- instance search, matching, and localization.

Applications

- search results ranking and diversification;
- novelty detection;
- topic detection, tracking, and threading;
- data-driven applications;
- Internet media management and service;
- Web-scale multimedia mining;
- Web-scale summarization and browsing of multimedia data; and
- multimedia archaeology mining.

Guest Editors

Chong-Wah Ngo, *City University of Hong Kong*, cwngo@cs.cityu.edu.hk

Changsheng Xu, *Chinese Academy of Sciences*, csxu@nlpr.ia.ac.cn

Wessel Kraaij, *TNO and Radboud University Nijmegen*, w.kraaij@cs.ru.nl

Abdulmotaleb El Saddik, *University of Ottawa, Ontario*, abed@mclab.uottawa.ca

Submission Procedures and Deadlines

Submit your paper at <https://mc.manuscriptcentral.com/cs-ieee>. When uploading your paper, please select the appropriate special issue title under the category "Manuscript Type." If you have any questions regarding the submission system, please contact Andy Morton at mm-ma@computer.org. All submissions will undergo a blind peer review by at least two expert reviewers to ensure a high standard of quality. Referees will consider originality, significance, technical soundness, clarity of exposition, and relevance to the special issue topics. All submissions must contain original, previously unpublished research or engineering work. Papers must stay within the following limits: 6,500 words maximum, 12 total combined figures and tables with each figure counting as 200 words toward the total word count, and 18 references.

To submit a paper to the July-September 2013 special issue, please observe the following deadlines:

29 June 2012: Full paper must be submitted using our online manuscript submission service and prepared according to the instructions for authors (please see the Author Resources page at <http://www.computer.org/multimedia/author.htm>).

28 September 2012: Authors notified of acceptance, rejection, or needed revisions.

30 November 2012: Revisions due

31 January 2013: Final versions due.