

# CLEFeHealth2012

## The CLEF 2012 Workshop on Cross-Language Evaluation of Methods, Applications, and Resources for eHealth Document Analysis

Hanna Suominen

ICTA, National ICT Australia and The Australian National University, Locked Bag 8001,  
2601 Canberra, ACT, Australia  
[hanna.suominen@nicta.com.au](mailto:hanna.suominen@nicta.com.au)

**Abstract.** CLEFeHealth2012 is the CLEF2012 workshop on cross-language evaluation of methods, applications, and resources for eHealth document analysis. Its focus is on written and spoken natural-language processing with the use scenario of people using ICT tools to analyse and integrate eHealth documents across languages, genres, and jargons. The significance of workshop lies in taking steps toward systematic evaluation and continuous improvement of eHealth ICT. The target audience of CLEFeHealth2012 is interested in syntactic, semantic, and pragmatic aspects of electronic health documents, their intelligent re-use, their automated analysis, and evaluation of the analysis methods. The potential participants represent academia, industry, healthcare, and government. The programme of the workshop consists of three invited talks on collaborative datasets, resources, tools, and infrastructure by Australian and US researchers; a panel of Austrian, Finnish, Irish, Swedish, and Swiss experts in eHealth; eleven paper presentations; and three workshops for roadmap development, student mentoring, and partnership building, respectively. CLEFeHealth2012 invites you to develop with us a roadmap towards this vision of using systematically evaluated ICT tools to analyse and integrate eHealth documents across languages, genres, and jargons. The workshop takes place in Rome, Italy on Tuesday 18 September, 2012.

**Keywords:** eHealth; Evaluation; Informatics; Natural Language Processing

### 1 Foreword

*CLEFeHealth2012* is the CLEF2012 workshop on cross-language evaluation of methods, applications, and resources for *eHealth* document analysis, as defined by the World Health Organization as the transfer of health resources and healthcare by information and communication technologies (ICT).<sup>1</sup> Its focus is on written and spoken natural-language processing with the use scenario of people using ICT tools to analyze and integrate eHealth documents across languages, genres, and jargons.

---

<sup>1</sup> <http://bit.ly/S3HhNj> (last accessed: 10 Aug 2012)

The *target audience* of CLEFeHealth2012 is interested in syntactic, semantic, and pragmatic aspects of electronic health documents, their intelligent re-use, their automated analysis, and evaluation of the analysis methods. The potential participants represent academia, industry, healthcare, and government and include the following task-relevant stakeholders:

- ICT developers (e.g., university departments, research and development (R&D) laboratories, and software houses),
- resource developers (e.g., standards, classifications, and ontologies), and
- information consumers (e.g., policy makers, healthcare service providers, healthcare professionals, health researchers, and patient associations).

The workshop is *linked* internationally with

- eHealth clusters and projects:
  - Australian eHealth Cluster,
  - HEXAnord, Health Text Analysis Network in the Nordic and Baltic Countries,
  - iDASH (Integrating Data for Analysis, Anonymization and Sharing,
  - IKITIK, Information and language technology for health information and communication consortium, and
  - Khresmoi (Medical Information Analysis and Retrieval),
- eHealth workshops and evaluation labs:
  - ImageCLEF 2003–2012,
  - iDASH Privacy Workshop, and
  - Louhi Workshop Series on Health Document Text Mining and Information Analysis, organised since 2008 jointly with the 3rd International Symposium on Semantic Mining (Finland, Int J MedInform 2009 78(12)), 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (USA, J Biomed Semantics 2011 2 (3)), and 13th Conference on Artificial Intelligence in Medicine (Slovenia).

The *significance* of workshop lies in taking steps toward systematic evaluation and continuous improvement of eHealth ICT. The goal of the workshop is to promote R&D in health-information access by

1. establishing a discussion forum for our target audience,
2. supporting collaborative R&D of mono- and multilingual methods, applications, and resources for eHealth document analysis by providing
  - mentoring for students who wish to submit a paper to the Louhi2013 workshop organised in Sydney, Australia in February 2013 together with the Australian Health Informatics Summer School, and
  - a WWW forum for actors to describe their work-in-progress, call new partners, and interact.
3. taking steps towards systematic, cross-language evaluation of methods, applications, and resources for eHealth document analysis by discussing
  - aspects of ethics, legislation, and intellectual property in eHealth data, resource, and ICT dissemination,
  - statistical and user-feedback based evaluation protocols, settings, methods, and measures, and

- ICT and laboratory infrastructure for evaluation, and
4. developing a roadmap, which
    - addresses collaborative domain-datasets, resources, tools, and infrastructure,
    - has matching data available across languages, genres, and jargons,
    - includes an infrastructure for benchmarking against and together with other eHealth-ICT components and resources.

The *programme* of the workshop consists of

1. three invited talks on collaborative datasets, resources, tools, and infrastructure by Australian and US researchers,
2. a panel of Austrian, Finnish, Irish, Swedish, and Swiss experts in eHealth,
3. eleven paper presentations, and
4. three workshops for roadmap development, student mentoring, and partnership building, respectively.

The accepted papers describe original work in Australia, Austria, Finland, Norway, and Sweden. They address

1. evaluation of mono- and multilingual methods, applications and resources for eHealth document analysis and
2. development of statistical and user-feedback based evaluation protocols, settings, methods, and measures for cross-language evaluation of methods, applications, and resources for eHealth document analysis.

Ten papers focus on meeting the needs of medical experts and laypeople in information access and ease of understanding the content via user-centred vocabularies, ontologies, abbreviation processing, language generation, search engines, and tools to support patient-clinician interaction. One paper describes an eHealth evaluation setting. The papers were refereed by three programme committee members.

CLEFeHealth2012 *invites* you to develop with us a roadmap towards this vision of using systematically evaluated ICT tools to analyse and integrate eHealth documents across languages, genres, and jargons. The workshop takes place in Rome, Italy on Tuesday 18 September, 2012. We are also preparing a special issue for Artif Intell Med. Hopefully these proceedings will feed the virtuous cycle of scientists proposing new research questions; data and ICT developers building settings to answer them; and healthcare practitioners and policymakers transferring the answers to clinical practice.

Hanna Suominen, Chair of CLEFeHealth2012

Canberra, ACT, Australia  
August 2012

## **2      Sponsors**

